UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

Bárbara Helohá Falcão Teixeira

Detecção automática de fronteiras prosódicas na fala espontânea

Belo Horizonte
2022

Bárbara Helohá Falcão Teixeira

Detecção automática de fronteiras prosódicas na fala espontânea

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutora em Estudos Linguísticos.

Área de Concentração: Linguística Teórica e Descritiva
Linha de Pesquisa: Estudos Linguísticos baseados em *corpora*
Orientador: Prof. Dr. Tommaso Raso
Coorientador: Prof. Dr. Plínio Almeida Barbosa

Belo Horizonte
2022

Teixeira, Bárbara Helohá Falcão.
T266d Detecção automática de fronteiras prosódicas na fala espontânea [manuscrito] / Bárbara Helohá Falcão Teixeira. – 2022.
1 recurso online (224 f. : il., tabs., color., p&b) : pdf.

Orientador: Tommaso Raso.

Corientador: Plínio Almeida Barbosa.

Área de concentração: Linguística Teórica e Descritiva.

Linha de Pesquisa: Estudos Linguísticos Baseados em Corpora.

Tese (doutorado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 175-187.

Apêndices: f. 188-190.

Anexos: f. 191-209.

Exigências do sistema: Adobe Acrobat Reader.

1. Fala – Teses. 2. Análise prosódica (Linguística) – Teses. 3. Linguística de corpus – Teses. I. Raso, Tommaso. II. Barbosa, Plínio Almeida. III. Universidade Federal de Minas Gerais. Faculdade de Letras. IV. Título.

CDD: 414

Ficha catalográfica elaborada pelo Bibliotecário Israel José da Silva – CRB/6-2128

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

FOLHA DE APROVAÇÃO

Detecção automática de fronteiras prosódicas na fala espontânea

BÁRBARA HELOHÁ FALCÃO TEIXEIRA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGUÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGUÍSTICOS, área de concentração LINGUÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Estudos Linguísticos Baseados em Corpora.

Aprovada em 28 de junho de 2022, pela banca constituída pelos membros:

Prof(a). Tommaso Raso - Orientador

UFMG

Prof(a). Plinio Almeida Barbosa - Coorientador

Unicamp

Prof(a). Heliana Ribeiro de Mello

UFMG

Prof(a). Sandra Madureira

PUC-SP

Prof(a). Joao Antonio de Moraes

UFRJ

Prof(a). Bruno Neves Rati de Melo Rocha

UFMG

Belo Horizonte, 28 de junho de 2022.

Documento assinado eletronicamente por **Heliana Ribeiro de Mello**, **Professora do Magistério**

Superior, em 29/06/2022, às 11:15, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Bruno Neves Rati de Melo Rocha**, **Professor do Magistério Superior**, em 29/06/2022, às 13:20, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Tommaso Raso**, **Professor do Magistério Superior**, em 29/06/2022, às 16:14, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Plinio Almeida Barbosa**, **Usuário Externo**, em 29/06/2022, às 16:45, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **João Antônio de Moraes**, **Usuário Externo**, em 30/06/2022, às 17:11, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Sandra Madureira Fontes**, **Usuária Externa**, em 12/07/2022, às 11:50, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1542076** e o código CRC **8772A980**.

AGRADECIMENTOS

Ao Prof. Tommaso Raso pela orientação, paciência e dedicação.
Ao Prof. Plínio A. Barbosa pela coorientação, paciência e incentivo.
À Banca Examinadora pela leitura e avaliação da tese.
À CAPES pela bolsa de estudo.
Aos colegas do LEEL pela colaboração em diversos momentos.
À Profa. Heliana Mello pelo incentivo desde a iniciação científica.
Aos meus pais Ana e Pedro pelo apoio e compreensão.
Ao Kiko (*in memorian*) pela companhia.
Ao Carlos por tudo!

RESUMO

A fala é segmentada em unidades entonacionais delimitadas por fronteiras prosódicas. Essa segmentação tem consequências importantes para a sintaxe, estrutura informacional e cognição. Este trabalho tem como objetivo investigar os parâmetros fonético-acústicos envolvidos na percepção das fronteiras prosódicas, e desenvolver modelos destinados à detecção automática das fronteiras prosódicas em fala espontânea monológica masculina de português brasileiro. Duas amostras foram segmentadas em unidades entonacionais por dois grupos de segmentadores treinados. As fronteiras percebidas pelos segmentadores foram marcadas como terminais ou não-terminais. Um *script* foi usado para extrair parâmetros fonético-acústicos ao longo do sinal de fala em uma janela à direita e à esquerda ao redor de cada fronteira de palavra fonológica. Os parâmetros extraídos compreendem medidas de (1) Taxa de articulação e elocução; (2) Duração normalizada do segmento; (3) Frequência fundamental; (4) Intensidade; (5) Pausa física. O *script* considera como fronteira prosódica as posições em que pelo menos 50% dos segmentadores indicaram uma fronteira do mesmo tipo. Os modelos de identificação das fronteiras prosódicas foram desenvolvidos a partir de um algoritmo de Análise Discriminante Linear e foram usadas diferentes estratégias de treinamento heurístico. Os modelos apresentaram problemas semelhantes em razão da importância dos parâmetros relacionados às pausas físicas. Foi realizado um teste de percepção de pausa por um grupo de anotadores. A duração de pausa física que pode ser percebida como pausa, evitando confusão com a fase oclusiva dos segmentos, foi analisada. Neste trabalho, a duração proposta é 100 ms. Adicionalmente, submodelos para identificar fronteiras com e sem pausa percebida foram desenvolvidos. O submodelo de identificação de fronteiras com pausa percebida apresentou resultados superiores aos resultados do submodelo de identificação de fronteiras sem pausa percebida.

Palavras-chave: fala espontânea; fronteiras prosódicas; detecção automática.

ABSTRACT

Speech is segmented into intonational units delimited by prosodic boundaries. This segmentation is claimed to have important consequences for syntax, information structure and cognition. This work aims to investigate the phonetic-acoustic parameters involved in the perception of prosodic boundaries, and to develop models for automatic detection of prosodic boundaries in Brazilian Portuguese male monological spontaneous speech. Two samples were segmented into intonational units by two groups of trained annotators. The boundaries perceived by the annotators were tagged as terminal or non-terminal. A script was used to extract phonetic-acoustic parameters along the speech signal in both a rightward and a leftward window around the boundary of each phonological word. The extracted parameters comprise measures of (1) Speech rate and rhythm; (2) Normalized duration; (3) Fundamental frequency; (4) Intensity; (5) Physical pause. The script considers as prosodic boundaries positions at which at least 50% of the annotators indicated a boundary of the same type. The models were developed from a Linear Discriminant Analysis algorithm and different heuristic training strategies were used. The models presented similar problems due to the importance of parameters related to physical pauses. A pause perception test was performed by a group of annotators. The duration of physical pause that can be perceived as a pause, avoiding confusion with the occlusive phase of the segments, was analyzed. In this work, the proposed duration is 100 ms. Additionally, submodels for identifying boundaries with and without a perceived pause were developed. The submodel for identifying boundaries with a perceived pause showed better results than the submodel for identifying boundaries without a perceived pause.

Keywords: spontaneous speech; prosodic boundaries; automatic detection.

LISTA DE TABELAS

Tabela 1: Pausas percebidas (Walker e Trimboli, 1982) 56
Tabela 2: Desempenho do modelo TB-nb 74
Tabela 3: Composição do modelo TB-nb 74
Tabela 4: Desempenho dos modelos NTB-1, NTB-2 e NTB-3 75
Tabela 5: Composição dos modelos NTB-1, NTB-2 e NTB-3 77
Tabela 6: Descrição dos dados 83
Tabela 7: Origem dos falantes 84
Tabela 8: Coeficiente Kappa das segmentações realizadas pelo grupo de 14 segmentadores 88
Tabela 9: Coeficiente Kappa das segmentações realizadas pelo grupo de 19 segmentadores 88
Tabela 10: Distribuição de posições contabilizadas a partir da janela de análise de 10 unidades 94
Tabela 11: Parâmetros fonético-acústicos extraídos automaticamente pelo *script* para Praat BreakDescriptor 95
Tabela 12: Distribuição de pausas físicas contabilizadas a partir da janela de análise de 10 unidades 96
Tabela 13: Distribuição de posições contabilizadas a partir da janela de análise de 7 unidades 97
Tabela 14: Distribuição de pausas físicas contabilizadas a partir da janela de análise de 7 unidades .. 97
Tabela 15: Distribuição de posições contabilizadas a partir da janela de análise de 3 unidades 98
Tabela 16: Distribuição de pausas físicas contabilizadas a partir da janela de análise de 3 unidades .. 98
Tabela 17: Desempenho do modelo TB-nb nas fases de desenvolvimento, validação e revalidação . 102
Tabela 18: Composição do modelo TB-nb 103
Tabela 19: Desempenho do modelo NTB-1 nas fases de desenvolvimento e revalidação 107
Tabela 20: Desempenho do modelo NTB-2 nas fases de desenvolvimento e revalidação 107
Tabela 21: Desempenho do modelo NTB-3 nas fases de desenvolvimento e revalidação 107
Tabela 22: Modelos NTB-1, NTB-2 e NTB-3 108
Tabela 23: Desempenho do modelo TB-nbfull na fase de desenvolvimento 112
Tabela 24: Composição do modelo TB-nbfull 112
Tabela 25: Desempenho do modelo NTB-nbfull na fase de desenvolvimento 115
Tabela 26: Composição do modelo NTB-nbfull 116
Tabela 27: Visão geral das amostras balanceadas 117
Tabela 28: Desempenho do modelo TB-b1 nas fases de desenvolvimento e validações 118
Tabela 29: Composição do modelo TB-b1 119
Tabela 30: Desempenho do modelo TB-b2 nas fases de desenvolvimento e validações 121
Tabela 31: Composição do modelo TB-b2 121
Tabela 32: Desempenho do modelo TB-bfull nas fases de desenvolvimento e validações 123
Tabela 33: Composição do modelo TB-bfull 123
Tabela 34: Desempenho do modelo NTB-b1 na fase de desenvolvimento 125
Tabela 35: Desempenho do modelo NTB-b1 nas fases de validação 126
Tabela 36: Composição do modelo NTB-b1 126
Tabela 37: Desempenho do modelo NTB-b2 na fase de desenvolvimento 128
Tabela 38: Desempenho do modelo NTB-b2 nas fases de validação 128
Tabela 39: Composição do modelo NTB-b2 128
Tabela 40: Desempenho do modelo NTB-bfull na fase de desenvolvimento 130
Tabela 41: Desempenho do modelo NTB-bfull nas fases de validação 131
Tabela 42: Composição do modelo NTB-bfull 131
Tabela 43: Desempenho do modelo TB-b7 nas fases de desenvolvimento e validação 135
Tabela 44: Composição do modelo TB-b7 135
Tabela 45: Desempenho do modelo NTB-b7 na fase de desenvolvimento 137
Tabela 46: Desempenho do modelo NTB-b7 nas fases de validação 137
Tabela 47: Composição do modelo NTB-b7 137

Tabela 48: Desempenho do modelo TB-b3 nas fases de desenvolvimento e validação 140
Tabela 49: Composição do modelo TB-b3..... 140
Tabela 50: Desempenho do modelo NTB-b3 na fase de desenvolvimento..... 142
Tabela 51: Desempenho do modelo NTB-b3 nas fases de validação 142
Tabela 52: Composição do modelo NTB-b3..... 142
Tabela 53: Resultados do Kappa de concordância entre a anotação de cada segmentador e a anotação fruto do acordo de pelo menos 50% entre segmentadores 145
Tabela 54: Distribuição de posições analisadas na tarefa de percepção de pausa..... 147
Tabela 55: Distribuição de posições analisadas na tarefa de percepção de pausa..... 147
Tabela 56: Duração das pausas físicas analisadas na tarefa de percepção de pausa 148
Tabela 57: Coeficiente Kappa relativo à percepção de pausa em todas as posições de fronteira prosódica 151
Tabela 58: Coeficiente Kappa relativo à percepção de pausa - exclusão das fronteiras que 14 anotadores concordam quanto à ausência de pausa percebida 151
Tabela 59: Coeficiente Kappa relativo à percepção de pausa - exclusão das fronteiras que 11 ou mais anotadores concordam quanto à ausência de pausa percebida 152
Tabela 60: Coeficiente Kappa relativo à percepção de pausa de cada anotador 152
Tabela 61: Coeficiente Kappa relativo à percepção de pausa nos grupos 1 e 2 153
Tabela 62: Pausas percebidas pelos anotadores 154
Tabela 63: Pausas físicas com duração entre 54 ms e 89 ms 155
Tabela 64: Pausas percebidas com duração entre 50 ms e 90 ms 155
Tabela 65: Taxa de articulação dos textos..... 157
Tabela 66: Fatores relevantes para a percepção de pausa em posição de TB com pausa física 159
Tabela 67: Fatores relevantes para a percepção de pausa em posição de NTB com pausa física 160
Tabela 68: Fatores relevantes para a percepção de pausa em posição de disfluência com pausa física 161
Tabela 69: Distribuição de posições usadas para desenvolver e validar os submodelos de identificação de fronteiras..... 167
Tabela 70: Desempenho do submodelo FRO-compausapercebida na fase de desenvolvimento..... 168
Tabela 71: Desempenho do submodelo FRO-compausapercebida na fase de validação..... 168
Tabela 72: Composição do submodelo FRO-compausapercebida..... 168
Tabela 73: Desempenho do submodelo POS-sempausapercebida na fase de desenvolvimento..... 171
Tabela 74: Desempenho do submodelo POS-sempausapercebida na fase de validação..... 171
Tabela 75: Composição do submodelo POS-sempausapercebida..... 171
Tabela 76: Desempenho do submodelo FRO-sempausapercebida na fase de desenvolvimento 172
Tabela 77: Desempenho do submodelo FRO-sempausapercebida na fase de validação 173
Tabela 78: Composição do modelo FRO-sempausapercebida..... 173
Tabela 79: Simulação do desempenho do submodelo FRO-compausapercebida 174
Tabela 80: Simulação do desempenho do submodelo FRO-compausapercebida 174
Tabela 81: Simulação do desempenho do submodelo FRO-sempausapercebida em posição de fronteira sem pausa percebida identificada corretamente pelo submodelo anterior 175
Tabela 82: Simulação do desempenho do submodelo FRO-sempausapercebida em posições sem fronteira marcadas indevidamente como fronteira sem pausa percebida pelo submodelo anterior 175
Tabela 83: Desempenho do submodelo GER-sempausapercebida na fase de desenvolvimento 179
Tabela 84: Desempenho do submodelo GER-sempausapercebida na fase de validação 179
Tabela 85: Composição do submodelo GER-sempausapercebida 180
Tabela 86: Simulação do desempenho do submodelo GER-sempausapercebida em posição sem fronteira, TB e NTB sem pausa percebida 181

LISTA DE FIGURAS

Figura 1: Output da ferramenta desenvolvida por Ananthakrishnan e Narayanan (2008) 64
Figura 2: *QR code* para acessar os áudios 82
Figura 3: Camadas de anotação da tarefa de percepção de fronteira 90
Figura 4: Funcionamento do BreakDescriptor 2.0 93

LISTA DE QUADROS

Quadro 1: Combinações de parâmetros fonético-acústicos associados às fronteiras em diferentes línguas 45
Quadro 2: Combinações de parâmetros sintáticos e fonético-acústicos associados às fronteiras em diferentes línguas 46
Quadro 3: Desempenho de algumas ferramentas de identificação automática de fronteiras prosódicas 66
Quadro 4: Instruções fornecidas aos segmentadores na tarefa de percepção de fronteiras da amostra I 86
Quadro 5: Instruções fornecidas aos segmentadores na tarefa de percepção de fronteiras das amostras II e extra 87
Quadro 6: Instruções fornecidas aos segmentadores na tarefa de percepção de pausa 150
Quadro 7: Diagrama da proposta de reconhecimento das fronteiras com e sem pausa percebida a partir da ordenação hierárquica de submodelos com base em quatro etapas interligadas 170
Quadro 8: Diagrama da proposta de reconhecimento das fronteiras com e sem pausa percebida a partir da ordenação hierárquica de submodelos com base em três etapas interligadas 178

LISTA DE ABREVIATURAS E SIGLAS

DES	Fase de desenvolvimento
F-AC	Variáveis fonético-acústicas
FRO-compausapercebida	Submodelo de identificação de fronteiras com pausa percebida
FRO-sempausapercebida	Submodelo de identificação de fronteiras sem pausa percebida
Full-3NTB	Amostra de desenvolvimento do modelo de identificação de fronteiras não-terminais desenvolvido a partir de parâmetros extraídos da janela composta por 3 unidades
Full-3TB	Amostra de desenvolvimento do modelo de identificação de fronteiras terminais desenvolvido a partir de parâmetros extraídos da janela composta por 3 unidades
Full-7NTB	Amostra de desenvolvimento do modelo de identificação de fronteiras não-terminais desenvolvido a partir de parâmetros extraídos da janela composta por 7 unidades
Full-7TB	Amostra de desenvolvimento do modelo de identificação de fronteiras terminais desenvolvido a partir de parâmetros extraídos da janela composta por 7 unidades
FULL-b (NTB)	Amostra balanceada em função do total de fronteiras não-terminais encontrado na amostra total sob a condição de janelas de análise compostas por 10 unidades
FULL-b (TB)	Amostra balanceada em função do total de fronteiras terminais encontrado na amostra total sob a condição de janelas de análise compostas por 10 unidades
Full-GERsempausapercebida	Amostra usada na fase de desenvolvimento do submodelo de identificação de posições sem fronteira, fronteiras terminais e não-terminais sem pausa percebida
GER-sempausapercebida	Submodelo de identificação de posições sem fronteira, fronteiras terminais e não-terminais sem pausa percebida
GLM	Modelo Linear Generalizado
HMM	Modelo oculto de Markov
I-b (NTB)	Amostra balanceada em função do total de fronteiras não-terminais encontrado na amostra I sob a condição de janelas de análise compostas por 10 unidades
I-b (TB)	Amostra balanceada em função do total de fronteiras terminais encontrado na amostra I sob a condição de janelas de análise compostas por 10 unidades
II-b (NTB)	Amostra balanceada em função do total de fronteiras não-terminais encontrado na amostra II sob a condição de janelas de análise compostas por 10 unidades
II-b (TB)	Amostra balanceada em função do total de fronteiras terminais encontrado na amostra II sob a condição de janelas de análise compostas por 10 unidades
LDA	Análise discriminante linear
LEEL	Laboratório de Estudos Empíricos e Experimentais da Linguagem
LEX	Variáveis lexicais
NN	Modelos em redes neurais
NTB-1	Modelo 1 de identificação de fronteiras não-terminais de Teixeira (2018)
NTB-2	Modelo 1 de identificação de fronteiras não-terminais de Teixeira (2018)

NTB-3	Modelo 3 de identificação de fronteiras não-terminais de Teixeira (2018)
NTB-b1	Modelo de identificação de fronteiras não-terminais desenvolvido a partir de categorias balanceadas em função do total de fronteiras não-terminais encontrado na amostra I sob a condição de janelas de análise compostas por 10 unidades
NTB-b2	Modelo de identificação de fronteiras não-terminais desenvolvido a partir de categorias balanceadas em função do total de fronteiras não-terminais encontrado na amostra II sob a condição de janelas de análise compostas por 10 unidades
NTB-b3	Modelo de identificação de fronteiras não-terminais desenvolvido a partir de parâmetros extraídos da janela composta por 3 unidades
NTB-b7	Modelo de identificação de fronteiras não-terminais desenvolvido a partir de parâmetros extraídos da janela composta por 7 unidades
NTB-bfull	Modelo de identificação de fronteiras não-terminais desenvolvido a partir de categorias balanceadas em função do total de fronteiras não-terminais encontrado na amostra total
NTB-nbfull	Modelo de identificação de fronteiras não-terminais desenvolvido a partir de todas as posições da amostra total
POS-sempausapercebida	Submodelo de identificação de fronteiras sem pausa percebida e posições sem nenhuma fronteira
REV	Fase de revalidação
RF	Floresta aleatória
SIM	Fase de simulação
SIN	Variáveis sintáticas
SVM	Máquinas de vetores de suporte
TB-b1	Modelo de identificação de fronteiras terminais desenvolvido a partir de categorias balanceadas em função do total de fronteiras terminais encontrado na amostra I sob a condição de janelas de análise compostas por 10 unidades
TB-b2	Modelo de identificação de fronteiras terminais desenvolvido a partir de categorias balanceadas em função do total de fronteiras terminais encontrado na amostra II sob a condição de janelas de análise compostas por 10 unidades
TB-b3	Modelo de identificação de fronteiras terminais desenvolvido a partir de parâmetros extraídos da janela composta por 3 unidades
TB-b7	Modelo de identificação de fronteiras terminais desenvolvido a partir de parâmetros extraídos da janela composta por 7 unidades
TB-bfull	Modelo de identificação de fronteiras terminais desenvolvido a partir de categorias balanceadas em função do total de fronteiras terminais encontrado na amostra total sob a condição de janelas de análise compostas por 10 unidades
TB-nb	Modelo de identificação de fronteiras terminais de Teixeira (2018) sob a condição de janelas de análise compostas por 10 unidades
TB-nbfull	Modelo de identificação de fronteiras terminais desenvolvido a partir de todas as posições da amostra total
UFMG	Universidade Federal de Minas Gerais
VAL	Fase de validação
V-V	Unidades Vogal-Vogal

LISTA DE SÍMBOLOS

*	Início de turno de fala
[/]	Retratação
/	Fronteira prosódica não-terminal
//	Fronteira prosódica terminal
+	Interrupção
XXX:	Identificação do falante

Sumário

1. Introdução 19

1.1. Justificativa 19

1.2. Objetivos 23

1.3. Organização do texto 24

2. Revisão de literatura 25

2.1. Estudos sobre a segmentação da fala em unidades entonacionais em diversas línguas 25

2.2. Diferentes objetos de estudo: Unidades entonacionais ou fronteiras prosódicas 26

2.2.1. Abordagens com foco na análise das unidades entonacionais 26

2.2.2. Abordagens com foco na análise das fronteiras prosódicas 29

2.2.2.1. Tipos de fronteiras prosódicas propostos em diferentes abordagens 30

2.2.2.2. Desafios dos estudos cujo foco é a análise das fronteiras prosódicas 35

2.3 Grau de confiabilidade estatística na anotação de fronteiras prosódicas 35

2.4. Parâmetros comumente associados à percepção de fronteiras prosódicas 41

2.5. Combinações de parâmetros fonético-acústicos associados à percepção das fronteiras prosódicas 42

2.6. A pausa e a sua importância para as fronteiras prosódicas 49

2.6.1. Noção de pausa proposta na literatura 49

2.6.2. O papel da pausa física na percepção das fronteiras prosódicas 58

2.6.3. A importância da pausa percebida para o desenvolvimento da ferramenta de identificação automática de fronteira prosódica 59

2.7. Ferramentas de segmentação automática da fala 60

2.8. Ponto de partida 69

2.9. Resumo do capítulo 79

3. Panorama geral da metodologia empregada 80

3.1. Dados 81

3.2. Tratamento dos dados 85

3.3. Princípios de funcionamento do *script* BreakDescriptor 90

3.4. Resumo do capítulo 99

4. Identificação automática das fronteiras prosódicas 99

4.1. Modelos de identificação de fronteiras desenvolvidos a partir de parâmetros extraídos de janelas de análise compostas por 10 unidades 102
4.1.1. Revalidação de modelos anteriores de identificação de fronteira 102
4.1.1.1. Revalidação do modelo anterior de identificação de fronteiras terminais 102
4.1.1.2. Revalidação dos modelos anteriores de identificação de fronteiras não-terminais 105
4.1.2. Modelos de identificação de fronteiras desenvolvidos a partir de todas as posições disponíveis 112
4.1.2.1. Modelo de identificação de fronteiras terminais 112
4.1.2.2. Modelo de identificação de fronteiras não-terminais 115
4.1.3. Modelos de identificação de fronteiras desenvolvidos a partir de categorias balanceadas entre si 117
4.1.3.1. Modelos de identificação de fronteiras terminais 118
a. Modelo desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras terminais da amostra I 118
b. Modelo desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras terminais da amostra II 121
c. Modelo desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras terminais da amostra total 123
4.1.3.2. Modelos de identificação de fronteiras não-terminais 125
a. Modelo desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras não-terminais da amostra I 125
b. Modelo desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras não-terminais da amostra II 128
c. Modelo desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras não-terminais da amostra total 130
4.2. Modelos de identificação de fronteiras desenvolvidos a partir de parâmetros extraídos da janela de análise de 7 unidades 133
4.2.1. Modelo de identificação de fronteiras terminais 134
4.2.2. Modelo de identificação de fronteiras não-terminais 137
4.3. Modelos de identificação de fronteiras desenvolvidos a partir de parâmetros extraídos da janela de análise de 3 unidades 139
4.3.1. Modelo de identificação de fronteiras terminais 140
4.3.2. Modelo de identificação de fronteiras não-terminais 142
4.4. Resumo do capítulo 144

5. Procedimento de investigação da percepção de pausa 144

5.1. Dados 145
5.2. Anotadores 148
5.3. Tarefa de percepção de pausas 149
5.4. Acordos relativos à percepção de pausa em posição de fronteira prosódica 150
5.5. Análise da duração mínima de pausa percebida 153
5.6. Análise de fatores potencialmente responsáveis pela percepção de pausa 156
5.6.1. Fatores potencialmente responsáveis pela percepção de pausa em posição de fronteira terminal com pausa física 159
5.6.2. Fatores potencialmente responsáveis pela percepção de pausa em posição de fronteira não-terminal com pausa física 160
5.6.3. Fatores potencialmente responsáveis pela percepção de pausa em posição de disfluência com pausa física 161
5.6.4. Parâmetros potencialmente responsáveis pela percepção de pausa em posição de fronteira com pausa física 162
5.7. Resumo do capítulo 163

6. Identificação automática de fronteiras com e sem pausa percebida 164
6.1. Dados 165
6.2. Submodelo de identificação de fronteiras com pausa percebida 167
6.3. Ordenação hierárquica de submodelos com base em quatro etapas interligadas 169
6.4. Ordenação hierárquica de submodelos com base em três etapas interligadas 177
6.5. Resumo do capítulo 182

7. Considerações finais 182

REFERÊNCIAS 192

APÊNDICES 205

ANEXOS 208

1. Introdução

Este trabalho visa desenvolver modelos e submodelos que serão destinados à construção de uma ferramenta automática de detecção de fronteiras prosódicas em dados de fala espontânea. Este trabalho tem dois propósitos principais. O primeiro é contribuir ao melhor entendimento teórico acerca das pistas que veiculam a percepção de fronteiras de unidades entonacionais. Por isso, é importante esclarecer que o foco deste trabalho não está no estudo das unidades em si, mas sim nas pistas prosódicas utilizadas para marcar suas fronteiras. Do ponto de vista prático, este trabalho visa possibilitar o desenvolvimento de uma ferramenta computacional que auxilie a compilação de *corpora* de fala espontânea de português do Brasil (doravante, PB), tornando o processo de segmentação da fala mais rápido, poupando-se simultaneamente tempo e esforços humanos, o que pode ser visto como uma contribuição para a linguística de *corpus* em geral.

A ferramenta deve ser construída de modo que o processo de identificação automática de fronteira prosódica seja baseado na percepção humana de fronteira prosódica em conjunto com uma série de parâmetros fonético-acústicos extraídos automaticamente nas posições de fronteira e nas posições adjacentes às fronteiras. Primeiramente, são apresentadas as justificativas e os objetivos deste trabalho. Em seguida, são apresentadas a revisão da literatura da área, a metodologia empregada e as análises realizadas. Por fim, são discutidos os resultados obtidos.

1.1. Justificativa

A importância da segmentação da fala em unidades entonacionais está diretamente relacionada à mensagem que a sequência de fala transmite. Tanto na fala, quanto na escrita, uma sequência linguística composta por uma série de palavras só pode ser adequadamente interpretada se as relações linguísticas estabelecidas entre as palavras são devidamente compreendidas. No caso da fala, a compreensão adequada das relações linguísticas depende da segmentação da sequência linguística em unidades entonacionais.

A principal função das unidades entonacionais é captar adequadamente o domínio das relações linguísticas em textos falados, sinalizando as relações semânticas e sintáticas entre as palavras. Portanto, a segmentação do fluxo da fala em unidades entonacionais marcadas por fronteiras prosódicas é o primeiro passo que deve ser feito para a interpretação de uma

sequência linguística. O conceito de unidade entonacional pode ser estudado por meio de duas perspectivas relacionadas entre si. Por um lado, as unidades entonacionais englobam trechos de fala que possuem uma função comunicativa. Por outro, as unidades entonacionais pressupõem o estabelecimento de suas fronteiras. De fato, as unidades entonacionais podem ser vistas como agrupamentos de palavras separados por fronteiras prosódicas que se realizam no sinal sonoro através de determinadas configurações de parâmetros fonético-acústicos. As configurações de parâmetros não são uniformes e nem aleatórias.

A importância das unidades entonacionais e respectivas fronteiras pode ser facilmente observada no exemplo 1 descrito abaixo.

Exemplo 1 – Na casa de Carlos ninguém sabe nada.
i. Na casa, de Carlos ninguém sabe nada.
ii. Na casa de Carlos, ninguém sabe nada.

No exemplo 1, a sequência é segmentada em duas unidades diferentes que afetam diretamente a interpretação semântico-sintática da sequência. Com a opção (i), *de Carlos* é um argumento do verbo *sabe*. Com a opção (ii) de segmentação, *de Carlos* é um modificador de *casa*. A diferença de interpretação sintática e semântica evidenciada através das duas possibilidades de segmentação do exemplo 1 é devida exclusivamente à segmentação da sequência. Assim, se quisermos compreender a mensagem que uma sequência realmente transmite, é necessário observar a segmentação das unidades entonacionais da sequência.

O exemplo 2[1], descrito abaixo, também mostra a importância das unidades entonacionais e respectivas fronteiras.

Exemplo 2 - João vai pro Rio até amanhã.
i. João vai pro Rio até amanhã.
ii. João? Vai pro Rio até amanhã?
iii. João! Vai pro Rio! Até amanhã.
iv. João? Vai pro Rio até amanhã.
v. João vai pro Rio? Até amanhã.

No exemplo 2 descrito acima, a estrutura pode ser segmentada em uma única unidade, duas ou três unidades, mas, naturalmente, a estrutura também pode ser segmentada de maneira diferente. Independentemente das ilocuções que podem ser veiculadas nas unidades em que o trecho foi segmentado, as fronteiras prosódicas das unidades entonacionais fornecem informações morfossintáticas cruciais para a interpretação dos textos falados. Com a opção (i),

[1] Exemplo retirado de Izre'el et al. (2020).

por exemplo, *João* se configura como sujeito do verbo *vai*, que por sua vez será terceira pessoa do presente indicativo. Em (ii) e em (iii) *João* não pode ser considerado o sujeito do verbo *vai*, mas sim um enunciado autônomo e *vai* deve ser analisado como segunda pessoa do imperativo. *Até amanhã* pode ser considerado um adjunto em (i), em (ii) e em (iv), mas em (iii) e (v) constitui por si só um enunciado totalmente autônomo cujo significado é muito diferente do significado dado pelos adjuntos.

A importância das unidades entonacionais e respectivas fronteiras é bastante evidente e pode ser observada através de simples trechos de fala espontânea. Vejamos abaixo o exemplo (3).

Exemplo 3 - C-ORAL-BRASIL I (bpubdl01)[2]
*PAU: não tá dando a altura daquele que a Isa marcou lá né

No caso da fala, para compreender adequadamente o exemplo 3, é necessário ouvi-lo. Muitos pesquisadores têm concentrado seus estudos em uma única questão: Que tipo de informação as fronteiras fornecem, ou seja, qual é seu valor funcional? A pergunta pode ser respondida por diversos autores da área. A literatura mostra que as fronteiras prosódicas podem resolver casos de ambiguidade local no discurso falado (Kjelgaard e Speer, 1999; Speer, Kjelgaard e Dobroth, 1996; Warren, Grabe e Nolan, 1995). As fronteiras prosódicas também podem desempenhar um papel na sinalização da presença de dependências de longa distância (Kraljic e Brennan, 2005; Schafer, Speer e Warren, 2005; Snedeker e Trueswell, 2003). Vários estudos evidenciam que as fronteiras prosódicas são utilizadas pelo falante para guiar o ouvinte na reconstrução da segmentação pretendida e na compreensão adequada da mensagem (Swerts, 1997; Carlson, 2009; Watson e Gibson, 2004; Frazier et al., 2006). Em resumo, as fronteiras

[2] O exemplo foi retirado do corpus C-ORAL-BRASIL (Mello e Raso, 2012); Todos os exemplos de áudio podem ser acessados através do seguinte *QR Code*:

Os arquivos de áudio também estão disponíveis no seguinte link:
https://drive.google.com/drive/folders/1DyBrgqM5ji-nZMkksrIFyyZvb5jjHHZc?usp=sharing

prosódicas podem ser vistas como índices linguísticos decisivos para depreender o significado nos níveis ilocucionário, semântico e sintático da comunicação falada à medida que captam adequadamente quantas ilocuções estão presentes em uma sequência, o domínio das relações linguísticas na fala e fornecem um critério de segmentação da fala em unidades relevantes do ponto de vista comunicativo da linguagem.

A segmentação em unidades entonacionais depende de fatores de natureza diversificada. Tais fatores podem ser de natureza fisiológica, cognitiva, informacional e sintática. Os diversos fatores que explicam a segmentação da fala em unidades entonacionais não se excluem, mas integram-se e mostram que um ou outro elemento pode determinar e explicar as posições de estabelecimento de fronteira prosódica.

No que diz respeito às razões fisiológicas, essa segmentação é importante para o falante produzir grupos respiratórios adequados à capacidade de expiração do pulmão humano. Neste caso, a segmentação das unidades entonacionais ocorreria nos pontos de inspiração (Stetson, 1928). Porém, essa relação não é necessária, porque um único grupo respiratório pode conter várias unidades entonacionais e várias fronteiras prosódicas. Apesar da relação entre inspiração do ar e fronteira de unidade entonacional não ser direta, é indiscutível que a respiração é um fator que pode determinar as posições de estabelecimento de fronteira de unidade entonacional. Do ponto de vista cognitivo, a segmentação é importante, porque ela é uma necessidade do processamento linguístico devido ao limite da capacidade de memória de trabalho do ser humano (Cowan, 1998). Sendo assim, as fronteiras de unidade entonacional determinam grupos com extensão cognitivamente confortável para o processamento linguístico. Contudo, isso ainda não explica o porquê de termos unidades muito curtas e outras bastante longas.

No nível linguístico, o estabelecimento de fronteira prosódica de unidade entonacional pode ser explicado por razões pragmáticas e sintáticas. No nível pragmático, a fronteira prosódica de unidade entonacional delimita unidades pragmáticas com valor informacional atribuído pelos falantes (Cresti, 2000). No nível sintático, as unidades entonacionais marcadas por fronteiras prosódicas correspondem a unidades de natureza sintática e, portanto, fronteira prosódica e fronteira sintática estariam fortemente correlacionadas (Nespor e Vogel, 1986; Selkirk, 2005). Porém, as fronteiras prosódicas nem sempre estão correlacionadas às fronteiras sintáticas. Para os casos em que não há essa correlação, os trabalhos que mapeiam a correlação entre fronteira sintática e fronteira prosódica sugerem que o ouvinte utiliza outras pistas para reconstruir as relações sintáticas da sequência de fala.

Este trabalho é motivado pela necessidade de compreender melhor a percepção de fronteiras prosódicas terminais e não-terminais. Neste trabalho, são investigadas as

combinações de parâmetros fonético-acústicos que podem estar envolvidas na percepção auditiva das fronteiras. Além disso, este trabalho analisa uma proposta de refinamento interno às macrocategorias de fronteiras prosódicas percebidas.

Um aspecto estritamente metodológico que permite o desenvolvimento deste trabalho é a utilização de um conjunto de dados representativos da fala espontânea extraído dos *corpora* C-ORAL-BRASIL I e II (Mello e Raso, 2012; Mello, Raso e Ferrari, em preparação), pois a maioria dos trabalhos sobre o tema utiliza dados de fala gravados em laboratório ou dados retirados de noticiários radiofônicos (Xu, 2010; Cruttenden, 1997; Ostendorf et al., 1995; Wightman e Ostendorf, 1994; Ross e Ostendorf, 1996).

Outra razão que motiva este trabalho é a possibilidade de oferecer modelos descritivos que poderão funcionar como preditores automáticos da realização de fronteiras prosódicas em português brasileiro, já que ainda não existe uma ferramenta com essa funcionalidade aplicável a dados dessa língua. Ainda sobre esse ponto, os modelos poderão ser extremamente úteis para a compilação de novos *corpora* de fala espontânea, porque pouparão simultaneamente tempo e trabalho humano, mesmo que seja necessário realizar uma revisão manual após a segmentação gerada automaticamente.

1.2. Objetivos

Com o desenvolvimento deste trabalho, objetiva-se:

i. Buscar e analisar as combinações de parâmetros fonético-acústicos envolvidos na percepção de fronteiras terminais e não-terminais em dados de fala espontânea;
ii. Buscar um refinamento interno às macrocategorias de fronteiras terminais e não-terminais em correspondência com a percepção humana;
iii. Contribuir com a melhor compreensão da noção de “pausa”;
iv. Desenvolver submodelos de classificação estatística adequados para predizer automaticamente a realização de diferentes subtipos de fronteiras terminais e não-terminais em dados de fala espontânea;
v. Contribuir à busca de um processo mais rápido de compilação de *corpora* orais segmentados em unidades entonacionais.

1.3. Organização do texto

Esta tese é composta por sete capítulos, considerando a introdução apresentada neste momento. Com exceção dos capítulos de introdução e considerações finais (capítulos 1 e 7), no final de cada capítulo, é apresentado um resumo dos principais pontos explorados ao longo do texto. O resumo é apresentado com o intuito de sintetizar e relembrar os conteúdos abordados.

O capítulo 2 apresenta uma revisão de literatura da área. A revisão de literatura é abrangente e discute sobre as principais abordagens teóricas para estudo das unidades entonacionais e fronteiras prosódicas, o grau de confiabilidade estatística na anotação de fronteiras prosódicas, os parâmetros fonético-acústicos associados às fronteiras e as suas combinações. Adicionalmente, o capítulo 2 apresenta uma revisão de literatura detalhada sobre o fenômeno da "pausa" e a sua importância para a percepção e identificação automática das fronteiras. No final do capítulo 2, também são apresentadas ferramentas e modelos desenvolvidos com o objetivo de identificar fronteiras prosódicas automaticamente.

O capítulo 3 apresenta um panorama geral sobre a metodologia utilizada, mas, neste capítulo, não são detalhadas as análises estatísticas realizadas. No capítulo 3, são apresentados apenas os dados usados, o tratamento recebido pelos dados, o *script* usado para extrair os parâmetros fonético-acústicos e os parâmetros extraídos automaticamente através deste *script*.

As análises realizadas neste trabalho são apresentadas detalhadamente ao longo dos capítulos 4, 5 e 6. O capítulo 4 apresenta as estratégias de desenvolvimento e validação de diversos modelos de identificação automática de fronteiras terminais e não-terminais. No capítulo 4, também é discutido o desempenho dos modelos, observando as posições identificadas corretamente bem como as limitações apresentadas pelos modelos. O capítulo 5 apresenta uma investigação acerca da percepção auditiva de pausa por um grupo de anotadores. No capítulo 5, é analisada a duração mínima de pausa percebida pelos ouvintes e também são analisados outros elementos que podem influenciar a percepção da pausa. O capítulo 6 é destinado à descrição e à análise de duas propostas diferentes para reconhecer automaticamente fronteiras com e sem pausa percebida.

No capítulo 7, são retomados os objetivos e os resultados deste trabalho. Além disso, o capítulo 7 apresenta o que é necessário para implementar a ferramenta de identificação automática de fronteiras prosódicas terminais e não-terminais.

2. Revisão de literatura

A revisão de literatura é dividida em oito subseções. A seção 2.1 apresenta uma série de estudos sobre a segmentação da fala em unidades entonacionais em diversas línguas. A seção 2.2 apresenta as principais abordagens pelas quais as unidades entonacionais e fronteiras prosódicas podem ser analisadas. A seção 2.3 aborda a questão do grau de confiabilidade estatística na anotação de fronteiras prosódicas. A seção 2.4 apresenta um panorama geral sobre os parâmetros fonético-acústicos comumente associados às fronteiras prosódicas na literatura da área. Na seção 2.5, as combinações de parâmetros comumente associados às fronteiras prosódicas são discutidas. Na seção 2.6, é apresentada uma revisão de literatura a respeito da importância da "pausa" para as fronteiras prosódicas. Na seção 2.7, são apresentadas ferramentas computacionais já desenvolvidas com a finalidade de segmentar a fala automaticamente. Por fim, na seção 2.8, é apresentado um estudo sobre a percepção de fronteiras prosódicas, que pode ser visto como o ponto de partida deste trabalho.

2.1. Estudos sobre a segmentação da fala em unidades entonacionais em diversas línguas

Trabalhos sobre a segmentação da fala em unidades entonacionais são encontrados pelo menos desde a década de 70. Inicialmente, observa-se que os trabalhos a respeito do tema, em sua maioria, utilizavam dados de fala produzidos em laboratório a partir de leituras, eliciações e tarefas pré-definidas pelo pesquisador, como por exemplo, tarefas de *map task*[3]. A partir dos anos 90, com o desenvolvimento tecnológico de ferramentas para análise e extração de dados acústicos, os trabalhos sobre o tema ganham um maior impulso.

Uma revisão da literatura da área mostra que há estudos orientados para essa questão em diversas línguas, como inglês (White et al., 2010; Krivokapić, 2007; Mo et al., 2008; Campbell, 1993; Swerts e Geluykens, 1994; Schuetze-Coburn et al., 1991; Brierley e Atwell, 2007; Ananthakrishnan e Narayanan, 2008; Mo, 2008; Ostendorf e Veilleux, 1994; Shriberg et al., 2000; Takeuchi et al., 2007; Wheeldon e Lahiri, 1997; Harris et al., 1981), sueco (Carlson et al., 2005), alemão (Fuchs et al., 2010; Strom, 2006; Kompe et al., 1997), holandês (Swerts, 1997; Swerts e Geluykens, 1994; Buhmann et al, 2002; Swerts et al., 1994; Cutler et al., 1997),

[3] A tarefa de *map task* é usada para gravar dados de fala. Geralmente, os sujeitos analisam um mapa que representa uma cidade. No mapa, um ou alguns pontos de referência são marcados. Solicita-se ao sujeito para descrever rotas para se locomover entre os pontos marcados. Trata-se, portanto, de fala eliciada.

mandarim (Tseng e Chang, 2008; Fon et al., 2011; Xie, 2008; Tseng et al., 2005; Tseng e Fu, 2005), húngaro (Szaszák e Beke, 2012), espanhol (Nebot, 2013), hebraico (Izre'el, 2011; Amir et al., 2004) e francês (Vaissière, 1983).

A segmentação da fala em unidades entonacionais tem sido consistentemente explorada no âmbito de compilação de grandes *corpora* orais de português brasileiro, inglês, italiano, francês, espanhol, hebraico, português europeu, holandês etc. Para o inglês, podemos citar o Santa Barbara Corpus of Spoken American English (Du Bois et al., 2005), o *corpus* AixMARSEC (Auran et al., 2004), o Hong Kong Corpus of Spoken English (Cheng et al., 2005) e o Corpus News da Rádio da Universidade de Boston (Ostendorf et al., 1995); para o português europeu, italiano, francês e espanhol, foram publicados os *corpora* da família C-ORAL-ROM (Cresti e Moneglia, 2005); para o português brasileiro, o *corpus* C-ORAL-BRASIL (Mello e Raso, 2012); para o hebraico, podemos citar o Corpus of Spoken Israeli Hebrew (Izre'el et al., 2009); para o holandês, temos o Spoken Dutch Corpus (Oostdijk, 2000); para línguas diversas, podemos citar o *corpus* of Spoken AfroAsiatic Languages (Mettouchi e Chanard, 2010).

2.2. Diferentes objetos de estudo: Unidades entonacionais ou fronteiras prosódicas

Os trabalhos sobre a segmentação da fala em unidades entonacioanais podem ser diferenciados quanto ao tipo de análise desenvolvida. Alguns autores focam seu interesse no estudo das unidades entonacionais, outros, por outro lado, focam no estudo das fronteiras prosódicas em si. Naturalmente, as duas abordagens se complementam e alguns estudos analisam as unidades entonacionais e respectivas fronteiras de forma integrada em uma única perspectiva. Independente do foco escolhido, é amplamente reconhecido que as unidades entonacionais e respectivas fronteiras compreendem um fenômeno percebido pelos ouvintes. Este trabalho tem como foco o estudo das fronteiras, mas apresenta as duas abordagens nas seções 2.2.1 e 2.2.2.

2.2.1. Abordagens com foco na análise das unidades entonacionais

Muitos termos são utilizados na literatura da área para se referir à unidade entonacional. A unidade entonacional também é chamada de grupo tonal (Palmer, 1922; Halliday 1967/1985), unidade tonal (Crystal, 1969/1975, Brazil, Coulthard e Johns, 1980; Brazil 1985), unidade

rítmica (Pike, 1945), frase entoacional[4] (Pierrehumbert, 1980; Nespor e Vogel, 1986), dentre outras, apesar de nem sempre sua definição coincidir completamente. Quanto à terminologia, é importante observar que as unidades em discussão não são delimitadas exclusivamente pela melodia, mas sim por uma variedade de parâmetros prosódicos. Dentre eles estão parâmetros de natureza duracional, intensidade, qualidade de voz etc. Por isso, um termo mais adequado seria unidade prosódica (Szczepek Reed, 2010). Neste trabalho, embora o foco não seja a análise da unidade entonacional, optou-se pelo uso do termo unidade entonacional, porque é um termo bem estabelecido na literatura da área.

Uma definição que a literatura apresenta para unidade entonacional é um grupo de palavras delimitado por meio de fronteiras prosódicas relevantes perceptualmente e veiculadas fisicamente no sinal acústico através de um contorno entonacional coerente diferente dos contornos precedentes e subsequentes (Chafe, 1994; Du Bois et al., 1992). Naturalmente, a unidade entonacional pode ser composta por uma única palavra ou uma série de palavras. Essa definição é amplamente utilizada em trabalhos sobre o tema, porém, apresenta alguns problemas.

Primeiramente, há um problema de circularidade. Por um lado, a literatura da área argumenta que a unidade entonacional é estabelecida quando a fronteira prosódica é marcada. Por outro, a literatura também argumenta que as fronteiras são definidas em relação às unidades que eles dividem. Assim, para capturar as propriedades da unidade entonacional, é necessário fazer referência à fronteira e, para capturar as propriedades da fronteira prosódica, é necessário fazer referência à unidade entonacional, impossibilitando uma definição autônoma.

Além disso, embora a fronteira prosódica seja facilmente percebida em alguns casos, não há uma definição clara sobre o que se configura como um contorno entonacional coerente diferente dos contornos precedentes e subsequentes. Seria apenas uma mudança de direção de frequência fundamental (f0)? Ou a mudança precisa ser de determinada grandeza mínima, média ou máxima? De fato, aspectos diferentes da entonação contribuem na percepção de fronteira prosódica, no entanto, os diferentes aspectos de f0 não são discutidos. Por fim, a definição em discussão faz referência apenas à entonação, porém, várias fronteiras percebidas não podem ser atribuídas apenas a mudanças evidentes no contorno entonacional de f0, mas sim a outros parâmetros fonético-acústicos.

[4] Frequentemente, a linguística brasileira usa o termo frase entoacional. Em português, outro termo utilizado é sintagma entoacional (Galves, 2001).

Apesar da variação quanto ao termo convencionalmente adotado em diferentes tradições literárias, o conceito de unidade composta por uma ou mais palavras[5] que é separada por fronteiras prosódicas é comum a todos trabalhos. Desta forma, em todos os trabalhos sobre o tema, a segmentação das unidades entonacionais ou unidades equivalentes pressupõe necessariamente a presença de pelo menos uma fronteira prosódica que indica os limites da unidade em questão. Neste caso, o que diferencia uma perspectiva da outra é como a fronteira é identificada e definida. As fronteiras podem ser definidas segundo princípios perceptuais e/ou teóricos. Assim, as fronteiras podem ser definidas dando um peso maior ou menor à percepção e/ou à teoria.

Do ponto de vista funcional, a unidade entonacional pode ser estudada e definida linguisticamente com base em perspectivas diferentes. As principais são a perspectiva sintática (Selting, 2000; Schegloff, 1998, Selkirk, 1984; Selkirk, 2005), informacional (Cresti, 2000; Halford, 1994 e 1996; Mukherjee, 2001) e cognitiva (Croft, 1995/ 2007; Kibrik e Podlesskaya, 2009). Neste caso, os trabalhos argumentam que há um paralelismo entre a dimensão prosódica relativa às unidades entonacionais e as dimensões sintática, informacional ou cognitiva modeladas em termos de unidades.

Em perspectivas sintáticas, pressupõe-se um mapeamento entre os constituintes sintáticos e as unidades entonacionais. Em perspectivas informacionais, argumenta-se que as unidades entonacionais se correlacionam substancialmente com unidades informacionais. Em perspectivas cognitivas, pressupõe-se um mapeamento entre unidades entonacionais e unidades cognitivas que delimitam a quantidade de informação com a qual falantes e ouvintes podem lidar de uma vez só. Por fim, a capacidade de expiração do pulmão humano (Stetson, 1928) e o limite da capacidade de memória de trabalho do ser humano (Cowan, 1998) contribuem na segmentação de unidade entonacional.

Para os fins deste trabalho, não é relevante discutir sobre as características de cada perspectiva linguística pela qual as unidades entonacionais podem ser estudadas. Basta compreender que as diferentes propostas não são sempre excludentes entre si, mas sim indicar que as unidades entonacionais podem ser explicadas e motivadas por razões diversas, que podem se integrar, em níveis diferentes do processo comunicativo. Assim, fatores linguísticos diversificados podem estar envolvidos no fenômeno linguístico das unidades entonacionais.

Nos textos de fala, frequentemente, há desacordos quanto ao tipo e à posição da fronteira prosódica. Sendo assim, para estes casos específicos, os estudos diversos sobre a unidade

[5] Em casos excepcionais, apenas sílabas.

entonacional têm como prática recorrente remeter-se a fatores diversos pressupostos dentro das próprias teorias para, então, identificar a posição da fronteira e o tipo de fronteira. No entanto, este tipo de procedimento é problemático metodologicamente, pois as unidades entonacionais são identificadas à luz de critérios teóricos diversos, o que implica condicionar a identificação da unidade entonacional a elementos teóricos. Para uma discussão completa sobre as questões que permeiam as abordagens cujo foco é a unidade entonacional, veja-se Ford (2004), Szczepek Reed (2010) e Stelma e Cameron (2007).

2.2.2. Abordagens com foco na análise das fronteiras prosódicas

Diversas abordagens cujo foco principal é o estudo sistemático das fronteiras prosódicas são apresentadas na literatura. Em geral, nestes trabalhos, o pressuposto básico relatado é identificar as fronteiras à luz da percepção auditiva. Desta forma, os critérios de identificação, caracterização e distinção das fronteiras são majoritariamente perceptuais e isto implica que a identificação da fronteira prosódica é feita a partir de um processo que envolve a sensação auditiva que é provocada pelo sinal sonoro nos ouvintes.

Há um certo grau de variação quanto às fronteiras percebidas, e são evidenciados alguns problemas a depender da proposta. Porém, é indiscutível que a percepção das fronteiras prosódicas é crucial para a comunicação falada, pois captam o domínio das relações linguísticas em textos de fala. As fronteiras prosódicas propostas em diversas abordagens teóricas são apresentadas ao longo da seção 2.2.2.1, os problemas de algumas propostas também são abordados na seção 2.2.2.1.

Alguns trabalhos que investigam as questões retratadas nos pontos i e ii são Collier et al. (1993), Cruttenden (1997) e Ostendorf, e Price (1992), porém, foram avaliados dados de fala lida, fala produzida em laboratório ou fala radiofônica. Outros estudos também discutem sobre os principais parâmetros fonético-acústicos associados à percepção das fronteiras prosódicas, mas analisaram dados de fala espontânea (Amir, Silber-Varod e Izre'el, 2004; Barth-Weingarten, 2016; Oliveira, 2000). Alguns trabalhos que investigam os sinais neuronais associados à percepção das fronteiras são Drury et al. (2016), Glushko et al. (2016); Nickels et al. (2013), Hwang e Steinhauer (2011), Pauker et al. (2011), Steinhauser (2003), Steinhauer e Friederici (2001).

2.2.2.1. Tipos de fronteiras prosódicas propostos em diferentes abordagens

As tradições funcionalista e formalista são amplamente discutidas nos estudos linguísticos. As diferenças entre essas duas abordagens residem no ponto de partida da análise e na direção de condução das análises. A tradição funcionalista parte das funções linguísticas e é seguida por análises sobre a estruturação das unidades. A tradição formalista parte da identificação e estruturação de unidades, e, então, são identificadas as funções comunicativas veiculadas pelas unidades. A tradição gerativa está principalmente relacionada a esta segunda abordagem. Naturalmente, as diferenças têm consequências na análise das fronteiras prosódicas, mas as diferenças encontradas no que diz respeito às fronteiras são devidas ao conceito de unidade entonacional adotado nas tradições funcionalista e formalista. Enquanto na tradição funcionalista, a unidade entonacional é vista funcionalmente, na tradição formalista, a unidade entonacional é vista como um domínio estrutural. Em relação às fronteiras prosódicas, em ambas as tradições, a percepção das fronteiras é observada. Na tradição formalista, geralmente, as fronteiras prosódicas são discutidas com base em diferentes graus/níveis de saliência perceptual. Na tradição funcionalista, as fronteiras são discutidas com base em diferentes valores comunicativos veiculados perceptualmente pelas fronteiras.

Alguns exemplos de trabalhos nos quais as fronteiras são vistas como um fenômeno perceptual organizado por meio da gradação de diferentes níveis de força perceptiva limitado a um determinado número de categorias são Byrd e Saltzman (2003), Ladd (1988), Silverman et al. (1992), Simon e Christodoulides (2016), Reichel e Mády (2014), Wightman et al. (1992), Krivokapić, (2007), dentre outros. Alguns autores simplesmente distinguem entre dois níveis de fronteiras prosódicas (fronteiras fortes *versus* fracas; Simon e Christodoulides, 2016; Reichel e Mády, 2014). Outros acreditam que é possível individualizar mais de dois níveis de força perceptiva, podendo chegar até sete níveis (Jun e Fougeron, 2000; Mertens, 1993; Rossi, 1999; Wightman et al., 1992; Barbosa, 2006; Barbosa, 1994; Krivokapić, 2007; Nespor e Vogel, 1986; Selkirk, 1995).

Um exemplo de trabalho muito conhecido que trata as fronteiras prosódicas como um fenômeno perceptual organizado através de diferentes níveis de saliência perceptual compreende o sistema de anotação entonacional de base fonológica ToBI (*Tone and Break Indices*; Silverman et al., 1992). O ToBI foi desenvolvido com objetivo de auxiliar a transcrição de dados e o desenvolvimento de modelos computacionais de síntese e reconhecimento de fala. O sistema de notação ToBI se baseia no sistema de tons proposto por Pirrehumbert (1980) e no sistema de índices de fronteiras (quebras, na terminologia do ToBI) entre palavras adjacentes

proposto por Price et al. (1991). Originalmente, a notação foi idealizada para descrever a entonação do inglês norte americano. Posteriormente, com o intuito de descrever outras línguas, foram desenvolvidos vários trabalhos que apresentam adaptações ao sistema original, como por exemplo, português europeu (Frota, 2014), espanhol (Sosa, 1999), grego (Arvaniti e Baltazani, 2000), alemão (Grice et al., 1996; Reyelt et al., 1996) e japonês (Venditti, 1999). A notação original ToBI utiliza diferentes rótulos distribuídos em 4 camadas (em inglês, *tiers*) ordenados da seguinte forma: 1) Marcação de tons; 2) Transcrição ortográfica; 3) Índices de fronteira para marcar os constituintes prosódicos (numerados de 0 a 4 para marcar os agrupamentos prosódicos); 4) Observações importantes para compreensão das camadas 1, 2 e 3.

No caso do ToBI especificamente, os índices de fronteira não podem ser comparados com a distinção entre fronteiras terminais e não-terminais, mesmo se aceitarmos que existem diferentes tipos de fronteiras terminais e não-terminais. O que distingue as quebras previstas no ToBI é a saliência perceptual da fronteira. As fronteiras numeradas com o número 2 compreendem fronteiras nas quais é possível perceber a fronteira, mas não há uma clara marcação de f0. Quando há uma clara marcação de f0, mas a percepção de fronteira não é verificada, também é anotado o índice de fronteira de valor 2. As mais fortes, numeradas com 3 e 4, se correlacionam respectivamente com a *intermediate phrase* (ip), que é delimitada por um tom de fronteira, e *intonational phrase* (IP), que é delimitada por um tom de fronteira e alongamento silábico final (Pierrehumbert, 1980).

Ainda em relação ao ToBI, é possível observar que fatores perceptuais e fatores formais coparticipam na atribuição das fronteiras prosódicas, pois a anotação de algumas fronteiras pode ser motivada por associações não perceptuais predeterminadas. Nas diretrizes que orientam as anotações do ToBI, é estabelecido que algumas fronteiras devem ser anotadas, caso determinados tons de fronteira (*boundary tone*; Ladd, 1996; Pierrehumbert, 1980) sejam visíveis através de *softwares* de análise de fala, mesmo se a percepção de fronteira não for verificada pelo segmentador. Este procedimento é questionável, porque, na prática, a percepção das fronteiras muitas vezes tem um papel reduzido. Isso é verificado quando a percepção de índice de fronteira (quebras, na terminologia do ToBI) não ocorre, mas os segmentadores selecionam o índice ‘3’, caso um acento frasal (*phrase accent*) for anotado na camada tonal, e ‘4’, caso um tom de fronteira for anotado. O grau de confiabilidade estatística na anotação de fronteiras a partir do sistema notacional ToBI será abordado na seção 2.3. Para uma revisão geral sobre os problemas do ToBI, veja-se Wightman (2002).

Entre os trabalhos que estabelecem uma gradação de diferentes níveis de saliência perceptual das fronteiras, é possível verificar algumas diferenças. Barth-Weingarten (2016)

relata que, às vezes, uma única posição pode ser classificada perceptualmente como posição de fronteira, mas também como posição sem nenhuma fronteira. Em alguns casos, de acordo com autora, também é difícil determinar a posição exata de percepção de fronteira. Por esse motivo, a autora sugere que o sistema de anotação precisa lidar de forma satisfatória com os casos com dúvidas quanto à percepção de fronteira e localização da fronteira.

A solução apresentada por Barth-Weingarten para resolver os casos com alguma dúvida quanto à percepção das fronteiras depende da situação. Para os casos de dúvida quanto ao estabelecimento da fronteira, a autora os denomina como "possíveis fronteiras." Para os casos de dúvida quanto à posição exata da fronteira, a autora os denomina como "áreas fronteiriças" e indica também a abrangência da área fronteiriça. O sistema de anotação proposto por Barth-Weingarten compreende, então, os níveis gradualmente organizados da seguinte forma: nenhuma fronteira, área fronteiriça, possível fronteira e fronteira[6]. A diferença entre Barth-Weingarten e demais estudos que estabelecem diferentes níveis de saliência perceptual das fronteiras é a inclusão dos níveis que tratam das possíveis fronteiras e áreas fronteiriças.

A categoria área fronteiriça não é comum a todos os trabalhos sobre as fronteiras prosódicas, mas ainda assim, a noção de área fronteiriça pode de alguma forma ser expandida para qualquer estudo sobre a percepção de fronteiras prosódicas. De fato, a maioria das posições de percepção de fronteira coincide com o final de palavra fonológica. Porém, as posições dos parâmetros que veiculam a percepção das fronteiras podem se localizar em posições que não necessariamente estão em final de palavra fonológica, o que nos permite inferir que existe uma área/ região na qual são estabelecidos os parâmetros que veiculam a percepção das fronteiras prosódicas.

Em outros trabalhos, as fronteiras prosódicas são vistas como um fenômeno perceptual *continuum* que varia gradualmente (Barnwell, 2011; Peters, Kohler e Wesener, 2005). Neste caso, argumenta-se que as fronteiras são dispostas em um *continuum* que varia gradualmente de percepção de ausência de fronteira até a percepção de presença de fronteira. Também é relatado que o *continuum* varia gradualmente de percepção clara de nenhuma fronteira à percepção clara de fronteira. Outro termo frequentemente usado para tratar da organização das fronteiras prosódicas com base em um *continuum* é o termo gradiente e palavras derivadas de gradiente. Geralmente, nestes trabalhos, o foco recai nos diferentes níveis de fronteiras prosódicas que podem ser percebidos pelo sistema auditivo humano, mas também há um grande enfoque nos parâmetros fonético-acústicos envolvidos na percepção dos níveis de fronteira

[6] Nos termos de Barth-Weingarten, nenhuma cesura, área cesural, possível cesura e cesura.

prosódica. Para autores vinculados ao *continuum* que varia de nenhuma fronteira até fronteira, a vantagem é capturar, de forma mais satisfatória, a percepção humana da fala, porque as fronteiras são tratadas com maior granularidade. Neste sentido, este tipo de sistema de anotação permitiria elaborar uma imagem mais completa sobre a estruturação prosódico-fonética da fala e saliência perceptual das fronteiras.

Embora seja interessante tratar as fronteiras prosódicas como um fenômeno perceptual com comportamento *continuum*, há a possibilidade de sempre encontrar algum grau de fronteira, por mínimo que seja. Por isso, seria extremamente difícil estabelecer qualquer consideração funcional sobre as fronteiras, pois o número de fronteiras prosódicas que podem ser produzidas e percebidas seria infinito. Também seria muito difícil capturar os parâmetros fonético-acústicos que diferem as posições de nenhuma fronteira das posições de fronteira.

De fato, quanto à percepção de fronteiras prosódicas organizadas através da gradação de diferentes níveis de saliência perceptual, algumas observações/questões são destinadas a um ou outro trabalho em razão das especificidades de cada trabalho. De modo geral, a principal questão a ser observada é a clara discordância sobre os possíveis níveis de saliência perceptual pelos quais as fronteiras podem ser produzidas e percebidas. Os diferentes níveis de força perceptiva das fronteiras variam consideravelmente entre os trabalhos de modo que alguns trabalhos distinguem entre dois níveis, outros trabalham com sete níveis. Porém, há dúvidas sobre a capacidade humana de distinguir perceptualmente tantos níveis de forma confiável em virtude dos limites da memória de trabalho (Cowan, 1998/ 2016).

Para autores ligados à tradição funcionalista, as fronteiras prosódicas se organizam e se diferenciam por meio de categorias perceptuais distintas, mas sem nenhuma gradação entre as categorias. Neste caso, argumenta-se que as fronteiras prosódicas veiculam a percepção de diferentes valores comunicativos. Trata-se, portanto, de uma distinção qualitativa perceptual. Também são verificadas variações entre as diversas propostas vinculadas à tradição funcionalista, mas ainda assim, a variação é menor do que a variação relatada nos trabalhos ligados à tradição formalista. Muitos autores distinguem entre fronteiras que transmitem a percepção de conclusão prosódica e linguística e fronteiras que transmitem a percepção da continuidade (Pike, 1945; Schegloff, 1998; Szczepek Reed, 2004; Swerts, 1994; Campbell, 1993; Swerts et al., 1994; Degand e Simon, 2009). Em geral, os trabalhos desta vertente teórica estabelecem que as fronteiras que veiculam a percepção de continuidade são chamadas de fronteiras não-terminais (não-conclusivas), já as fronteiras que veiculam a percepção de finalização são chamadas de fronteiras terminais (finais/conclusivas).

A grande vantagem de adotar um sistema de anotação que distingue fronteiras terminais

de não-terminais é compreender aspectos prosódicos que contribuem diretamente na interpretação da mensagem do ponto de vista funcional, como já foi visto na seção 1.1. Além disso, falantes, devidamente treinados, facilmente conseguem diferenciar perceptualmente fronteiras terminais de não-terminais (Moneglia et al., 2005; Mello et al., 2012;). Uma maior discussão sobre o grau de confiabilidade estatística em tarefas de anotação de fronteiras terminais e não-terminais é apresentada na seção 2.3.

Alguns trabalhos da área optam por um sistema de anotação sem nenhuma distinção perceptual entre as fronteiras percebidas, informando apenas as posições de percepção de fronteira prosódica (Pijper e Sanderman, 1994; Watson e Gibson, 2004; Mo et al., 2008; Park, 2002; Croft, 1995; Maschler, 2009; Auer, 2010). O objetivo principal, neste caso, é concentrar o estudo nas fronteiras prosódicas, porque são elas que os ouvintes monitoram naturalmente em uma situação de interação falada**.** Assim, as fronteiras percebidas não são diferenciadas quanto aos níveis de saliência perceptual, mas também não são diferenciadas quanto aos diferentes valores comunicativos veiculados perceptualmente.

Independentemente das particularidades de cada trabalho que opta pela anotação das fronteiras sem diferenciá-las perceptualmente, relata-se apenas a oposição entre presença *versus* ausência de fronteira percebida. Para este tipo de anotação, a grande vantagem é a possibilidade de obter um alto índice de acordo e confiabilidade estatística na anotação de fronteiras prosódicas percebidas, pois as categorias em análise são apenas duas. A desvantagem é tratar de forma única fronteiras que possivelmente transmitem informações diversificadas nos textos de fala.

O trabalho aqui apresentado visa investigar a percepção de fronteiras terminais e não-terminais com o intuito de possibilitar a construção de uma ferramenta computacional capaz de informar ao mesmo tempo as posições e a natureza perceptual de fronteiras prosódicas em dados de fala espontânea em PB. Este trabalho não deve ser visto como um trabalho ligado a uma teoria linguística especificamente, mas sim como um trabalho que parte da hipótese de uma distinção funcional entre as fronteiras terminais e não-terminais. Assim sendo, o processo de identificação das fronteiras prosódicas não é condicionado por uma teoria específica e não gera um efeito de circularidade à medida em que o critério básico de identificação das fronteiras é estritamente perceptual. A análise é guiada por uma abordagem cujo foco principal é compreender os parâmetros fonético-acústicos e os pesos dos parâmetros que podem ser relevantes no processo de identificação perceptual das fronteiras em discussão. Portanto, a ideia geral é apresentar e discutir os parâmetros envolvidos no fenômeno de percepção de fronteiras terminais e não-terminais.

2.2.2.2. Desafios dos estudos cujo foco é a análise das fronteiras prosódicas

As pesquisas cujo objetivo principal é investigar a questão da identificação perceptual das fronteiras prosódicas enfrentam uma série de dificuldades relacionadas à metodologia. Podemos citar o problema da validade ecológica da fala produzida em laboratório, o problema da qualidade acústica e comparabilidade dos dados de fala espontânea, a dificuldade de identificar os parâmetros, a dificuldade de identificar a combinação dos parâmetros envolvidos na percepção das fronteiras, o problema do tratamento estatístico e computacional (entre outros).

Os trabalhos também enfrentam uma outra dificuldade inerente à natureza de um fenômeno de base perceptual, o que torna o processo de desenvolvimento de ferramentas de identificação automática de fronteiras prosódicas muito complexo. A percepção de fronteira não é absoluta entre os falantes. Em alguns casos, a percepção das fronteiras não é clara. Em outros, há dúvidas quanto à localização da fronteira percebida. Também é possível observar que nem sempre há acordo quanto ao tipo de fronteira prosódica percebida. Isso tudo ocorre porque a percepção das fronteiras não constitui uma decisão categórica (Bolinger, 1972; Birkner, 2006; Barth-Weingarten, 2016), embora possa ser estabelecida uma certa diferença entre tipologias de fronteiras, conforme mostra a seção 2.2.2.1.

Enquanto algumas fronteiras são muito salientes perceptualmente e são percebidas por (quase) todos os segmentadores, outras têm um acordo bem menor quanto à sua percepção, à sua localização e à sua natureza. Diversos elementos podem influenciar a percepção de fronteira e o grau de confiabilidade estatística das anotações de fronteiras percebidas; a seção 2.3. aborda esta questão detalhadamente. Soma-se à lista o desafio de investigar as combinações de parâmetros fonético-acústicos associados à percepção das fronteiras prosódicas, pois as combinações variam entre línguas e, provavelmente, variam de acordo com os diferentes estilos de fala e diferentes diastratias (gênero, idade, nível cultural e características individuais). Desta forma, descrever de forma precisa os parâmetros fonético-acústicos associados à percepção das fronteiras é uma atividade com alto nível de complexidade.

2.3 Grau de confiabilidade estatística na anotação de fronteiras prosódicas

O grau de confiabilidade estatística na anotação de fronteiras prosódicas geralmente é mensurado por meio de diferentes paradigmas experimentais e técnicas estatísticas. O

procedimento mais comum é um determinado número de segmentadores anotar as posições de percepção de fronteira prosódica e, depois, submeter os dados a um teste estatístico para medir a concordância entre avaliadores. Entre as diferentes técnicas estatísticas que podem ser usadas neste contexto, o índice de concordância é uma medida utilizada para analisar o acordo intraobservadores e entre-observadores na atribuição de categorias. No caso das fronteiras prosódicas, o objetivo geral é simplesmente mensurar estatisticamente o quanto a percepção de fronteiras converge entre os segmentadores. Alguns trabalhos optam por mensurar o grau de confiabilidade estatística na anotação de diferentes tipos de fronteiras prosódicas percebidas, outros optam por avaliar a anotação de fronteiras percebidas sem realizar nenhuma distinção entre as fronteiras. As duas possibilidades de avaliação de anotação são apresentadas.

O coeficiente Kappa tem sido frequentemente utilizado para avaliar a concordância entre anotadores em tarefas de etiquetagem prosódica (Syrdal e McGory, 2000), e, em particular, a confiabilidade do acordo intraobservadores, que significa a confiabilidade do mesmo segmentador (Carletta, 1996). Dentre as versões propostas na literatura, a versão Fleiss do Kappa (Fleiss, 1971) fornece uma medida de concordância geral sobre um número fixo de anotadores no caso de classificação categórica ao contrário da versão Cohen do Kappa (Cohen, 1960), que fornece uma medida de concordância entre pares. O coeficiente Kappa, nas duas versões, é considerado uma medida mais robusta para avaliação de concordância, porque, em seu cálculo, as concordâncias que poderiam ser obtidas aleatoriamente são desconsideradas. O coeficiente Kappa pode assumir qualquer valor entre 0 e 1 de modo que, quanto maior o coeficiente de Kappa, maior é o índice de acordo entre segmentadores.

A segmentação perceptual dos *corpora* da coleção C-ORAL, por exemplo, foi feita por segmentadores treinados para identificar perceptualmente fronteiras terminais e não-terminais. O grau de confiabilidade dessas segmentações foi definido pelo acordo entre os segmentadores com base no coeficiente Kappa (Fleiss, 1971). O processo de transcrição e segmentação do C-ORAL-BRASIL passou por diversas etapas, que incluem a formação dos segmentadores, a realização de revisões e a realização dos testes estatísticos que validaram o acordo entre os segmentadores. O acordo geral para a segmentação é de 0,87 e 0,78 para fronteiras terminais e não-terminais respectivamente (Mello et al., 2012). Para mais detalhes sobre os critérios de segmentação prosódica adotada no C-ORAL-BRASIL, veja-se Mello et al., 2012. Os resultados do acordo entre segmentadores do *corpus* C-ORAL-ROM são descritos em Moneglia et al. (2005). O acordo entre segmentadores do C-ORAL-ROM, mesmo com procedimentos metodológicos diferente dos procedimentos adotados no C-ORAL-BRASIL,

garantem também um excelente grau de confiabilidade estatística na anotação destes dois diferentes tipos de fronteiras perceptualmente relevantes.

Em francês, um experimento avaliou o grau de confiabilidade estatística na tarefa de anotação perceptual de dois níveis de fronteiras prosódicas (Lacheret el al., 2010). O nível 1 compreende fronteiras prosódicas percebidas com pausa[7], independentemente da duração da pausa e função da fronteira. O nível 2 compreende fronteiras prosódicas que são percebidas como fronteiras sem nenhuma pausa. A tarefa foi realizada por 15 segmentadores *experts*. Os dados utilizados compreendem uma entrevista de 63 segundos. Os anotadores não tinham acesso aos parâmetros fonético-acústicos de f0, intensidade etc. As fronteiras foram anotadas com base na percepção na percepção apenas. O grau de confiabilidade das anotações foi definido pelo coeficiente Kappa (Fleiss, 1971). Os resultados mostram uma concordância moderada para anotação de fronteiras percebidas com o nível 2 (coeficiente Kappa de 0,56) e alta concordância para as fronteiras anotadas com o nível 1 (coeficiente Kappa de 0,86).

O coeficiente Kendall's W. (Kendall, 1948; em inglês, *Kendall coefficient of concordance W*.) é utilizado para avaliar a concordância entre observadores e intraobservadores para as situações em que as etiquetas atribuídas pelos observadores se organizam em uma escala ordinal. Para o caso das fronteiras prosódicas, o coeficiente em questão é especialmente útil para os trabalhos que diferenciam as fronteiras prosódicas com base em diferentes níveis/graus de força perceptiva ordenados sob uma determinada escala ordinal. Em Pijper e Sanderman (1994), o grau de confiabilidade estatística de anotação de fronteiras prosódicas percebidas foi avaliado a partir do coeficiente de Kendall's W. Os resultados obtidos mostram que segmentadores treinados têm uma percepção muito confiável sobre a presença de fronteira prosódica, mas não têm uma percepção muito confiável quando são anotadas fronteiras prosódicas de diferentes níveis de força perceptiva. Em virtude destes resultados, esses pesquisadores optam por um sistema de anotação mais simples que mostra apenas se a fronteira está ou não presente, em vez de anotar fronteiras prosódicas de diferentes níveis perceptivos (Pijper e Sanderman, 1994; Watson e Gibson, 2004).

Com o sistema notacional ToBI, o grau de confiabilidade estatística das anotações de fronteiras percebidas foi avaliado principalmente em dados de fala lida, fala produzida em ambiente de laboratório e entrevistas informais. O grau de confiabilidade em dados de fala

[7] Frequentemente, o termo "pausa" é usado para se referir às: i) interrupções do fluxo de fala observadas no sinal sonoro; ii) interrupções do fluxo de fala efetivamente percebidas pelos ouvintes como pausa. Neste trabalho, o termo "pausa física" é usado para se referir à interrupção do fluxo de fala observada no sinal sonoro. O termo "pausa percebida" é usado para se referir às interrupções do fluxo de fala efetivamente percebidas pelos ouvintes como pausa. No caso do trabalho de Lacheret el al. (2010), a "pausa" em questão diz respeito à pausa percebida.

espontânea ainda é pouco analisado na literatura. A métrica de avaliação do grau de confiabilidade estatística das anotações previstas no sistema notacional ToBI é apresentada em Silverman et al. (1992) e Pitrelli et al. (1994). De modo geral, o grau de confiabilidade estatística é medido através do número de pares de transcritores concordando com a etiqueta marcada. Em um grupo de quatro segmentadores, chamados de “A”, “B”, “C” e “D”, por exemplo, é possível estabelecer seis diferentes pares: “AB”, “AC”, “AD”, “BC”, “BD” e “CD” ou vice-versa “BA”, “CA”, “DA”, “CB”, “DB” e “DC”. Caso três dos quatro segmentadores, por exemplo os segmentadores “A”, “B” e “C”, concordem com uma etiqueta qualquer, a concordância dessa etiqueta é relatada como concordância de 50%, porque, entre os seis pares possíveis, apenas três pares (“AB”, “AC” e “CB”) concordam na etiqueta atribuída.

Um trabalho (Pitrelli et al., 1994) realizado a partir de fala lida (Paul e Baker, 1992), fala proveniente de interações entre homem e máquina (Hirschman et al., 1993) e fala eliciada (Gross et al., 1993) avaliou a concordância entre 26 segmentadores com pouca ou muita experiência com o sistema notacional ToBI. Neste trabalho, foram avaliados todos os índices previstos no ToBI para representar as fronteiras prosódicas. A concordância obtida corresponde a 66,6%. Em coreano (Jun et al., 2000), um estudo realizado a partir de entrevistas, dados de fala provenientes da mídia televisiva e fala lida também avaliou a concordância quanto à anotação dos índices de fronteiras previstos no ToBI. São avaliados segmentadores treinados (*experts*) com maior experiência e segmentadores familiarizados com o sistema de notação ToBI. A concordância entre os segmentadores treinados corresponde a 65,5%, considerando os diferentes níveis de fronteiras previstos no ToBI. Neste mesmo estudo, a concordância entre segmentadores treinados em conjunto com segmentadores familiarizados com o ToBI corresponde a 58,5% nas mesmas condições. Nos dois trabalhos, as instruções fornecidas aos segmentadores para marcar as fronteiras prosódicas são aquelas apresentadas no manual do ToBI. Porém, as diretrizes do ToBI tendem a reduzir o papel da percepção das fronteiras, pois a anotação dos índices de fronteira é influenciada pelas demais camadas de modo que índices de fronteiras podem ser anotados, mesmo que a percepção da fronteira não seja efetivamente verificada pelos segmentadores.

Em uma tarefa de percepção de fronteira, um total de 25 ouvintes, dentre os quais 90% são falantes nativos de francês e 10% não nativos, foram solicitados a marcar uma linha vertical entre as palavras que perceberam fronteiras prosódicas em dados de francês. Os ouvintes não possuíam treinamento avançado em percepção de fronteira prosódica. Os áudios são provenientes de um debate transmitido por mídia televisiva e de uma fala gravada após tarefas pontuais dadas pelos pesquisadores. Os dois áudios foram recortados em diversos trechos com

duração variando de 13 a 53 segundos. O grau de confiabilidade das anotações foi definido pelo teste estatístico Kappa na versão proposta por Brennan e Prediger (1981). O Kappa foi calculado para cada trecho recortado, os resultados mostram que os valores de Kappa variam de 0,75 a 0,88 com uma média de 0,83.

Um estudo realizado em inglês (Cole et al., 2010) incluiu 97 sujeitos distribuídos em quatro grupos. Os sujeitos não haviam sido treinados anteriormente, mas eram falantes nativos de inglês americano. Em Cole et al. (2010), solicitou-se aos sujeitos que marcassem palavras proeminentes ou fronteiras prosódicas entre grupos de palavras em uma amostra de fala espontânea de conversação. Enquanto ouviam os áudios, eles acompanhavam uma transcrição impressa, ortográfica, mas não pontuada. Os ouvintes indicaram a percepção de proeminência e fronteira sublinhando as palavras percebidas como proeminente e marcando uma barra entre duas palavras nas posições de percepção de fronteira. Para cada texto usado no estudo, foi calculado o coeficiente Kappa na versão Cohen para mensurar a concordância entre segmentadores. Os resultados mostram que o Kappa varia de 0,24 a 0,850, a depender do texto. A média do coeficiente Kappa é 0,58.

Em geral, a literatura mostra que não há um consenso geral sobre o grau de confiabilidade estatística na anotação de fronteiras prosódicas percebidas. Os resultados apresentados na literatura diferem-se bastante. Para alguns autores, a anotação de determinadas fronteiras prosódicas percebidas não é completamente confiável (Pijper e Sanderman, 1994; Lacheret el al., 2010). Para outros, no entanto, há um alto índice de concordância entre segmentadores treinados para identificar perceptualmente duas fronteiras prosódicas (Mello et al., 2012; Moneglia et al.; 2005). No caso do sistema notacional ToBI, o grau de concordância apresentado é em alguma medida questionável, porque, frequentemente, a atribuição dos índices de fronteira é motivada por associações não necessariamente perceptuais, conforme foi explicado na seção 2.2.2.1.

Naturalmente, surge uma discussão em torno dos fatores que poderiam influenciar a percepção de fronteira prosódica, influenciando também o grau de confiabilidade estatística da anotação das fronteiras. Primeiramente, a percepção de fronteira e, consequentemente, o grau de concordância entre segmentadores são afetados pelo tipo de fala analisada. Normalmente, as falas lidas e produzidas em laboratório (*speech lab*) permitem a identificação perceptual das fronteiras com maior facilidade, enquanto a fala espontânea pode levar a uma concordância menor. Na fala espontânea, a tipologia interacional também pode influenciar no grau de concordância entre segmentadores. Os turnos de fala dos diálogos costumam ser curtos e muito interativos, o que facilita a marcação das fronteiras prosódicas, porque, ao final de cada turno,

sempre há uma fronteira prosódica. As conversações também são bastante interativas. Por um lado, assim como nos diálogos, a interatividade facilita a anotação de fronteiras percebidas. Por outro, a interatividade pode influenciar o grau de acordo entre segmentadores, porque, geralmente, há muitas sobreposições, o que torna o processo de percepção e anotação das fronteiras mais difícil. Obviamente, outros aspectos da fala, como a velocidade da fala ou características idiossincráticas, podem tornar o processo de identificação perceptual da fronteira mais difícil e afetar o grau de confiabilidade estatística das anotações de fronteiras percebidas.

Fatores não relacionados à fala em si também podem afetar o grau de confiabilidade estatística das anotações de fronteiras percebidas. Um deles pode ser o grau de atenção que o segmentador dedicou à tarefa de anotação das fronteiras percebidas. Certamente, o tempo e a qualidade da capacitação destinada ao aprendizado do processo de percepção de fronteiras prosódicas influenciam. A regularidade com a qual os segmentadores realizam segmentações de trechos de fala também podem interferir. Por fim, a individualidade do sistema de percepção do ser humano também pode ter algum efeito no grau de confiabilidade da anotação de fronteiras percebidas, porque alguns segmentadores podem ser mais sensíveis a determinados aspectos, percebendo certas fronteiras com maior facilidade, enquanto outros não.

A forma como a fronteira prosódica é vista nos diversos sistemas de anotação também pode influenciar no grau de confiabilidade estatística das anotações de fronteiras percebidas. Geralmente, fronteiras diferenciadas quanto à função comunicativa veiculada perceptualmente tendem a ter um maior acordo do que fronteiras discutidas com base em níveis de saliência perceptual. Em outras palavras, o acordo para a anotação de fronteiras terminais *versus* fronteiras não-terminais tende a ser melhor (isto é, mais alto) do que o acordo para a anotação de fronteiras de diferentes níveis de saliência perceptual. Além disso, existe a possibilidade do maior acordo quanto à anotação de fronteiras terminais e não-terminais ser obtido, porque os valores perceptuais de terminalidade e continuidade abrangem apenas duas categorias, o que, em tese, simplificaria o processo de percepção em razão de uma menor quantidade de operações necessárias para categorizar e diferenciar perceptualmente as fronteiras prosódicas.

Naturalmente, os parâmetros fonético-acústicos têm um papel fundamental na percepção ou não percepção das fronteiras prosódicas. É importante esclarecer que as fronteiras prosódicas não são objetos concretos sempre produzidos da mesma forma, mas sim instâncias parcialmente estáveis que são realizadas por meio da variação combinatória dos parâmetros fonético-acústicos que, aliado aos fatores que afetam a percepção, explicariam o eventual desacordo entre segmentadores. Se uma determinada fronteira pode ser realizada por meio de

uma combinação de parâmetros que pode sofrer variações, podemos inferir também que certas combinações são mais frequentes do que outras.

A priori, a frequência de uso das combinações de parâmetros poderia influenciar no grau de confiabilidade estatística das anotações de fronteiras percebidas, porque algumas combinações seriam mais prototípicas do que outras. Assim, as combinações mais frequentes seriam percebidas com maior facilidade e corresponderiam às fronteiras percebidas por um maior número de ouvintes (causando, então, um maior acordo entre segmentadores). As combinações pouco frequentes tornariam o processo de percepção de fronteira mais complexo e configuram-se como potenciais fronteiras a serem percebidas por um menor número de ouvintes (causando, assim, um menor acordo entre segmentadores).

Neste trabalho, o grau de confiabilidade estatística da anotação da percepção de fronteiras terminais e não-terminais foi avaliado através do coeficiente Kappa proposto por Fleiss (1971). O grau de confiabilidade estatística da anotação da percepção de diferentes subtipos de fronteiras terminais e não-terminais foi avaliado da mesma forma.

2.4. Parâmetros comumente associados à percepção de fronteiras prosódicas

A literatura da área mostra que diversos estudos buscam entender os parâmetros fonético-acústicos associados à percepção das fronteiras prosódicas. De modo geral, os parâmetros fonético-acústicos básicos associados à percepção das fronteiras são:

1. Pausa física (Martin, 1973; Mo e Cole, 2010; Shriberg, Stolcke, Hakkani-Tür e Tür, 2000; Swerts, 1997; Tseng e Chang, 2008; Tyler, 2013; Krivokapić, 2007; Swerts et al., 1992; Carlson et al., 2005; Campbell, 1993; Swerts e Geluykens, 1994; Schuetze-Coburn et al., 1991; Brierley e Atwell, 2007; Tseng e Fu, 2005; Couper-Kuhlen, 2006)
2. Alongamento em posições pré-fronteiriça e fronteiriça (Barbosa, 2008; Fon, Johnson e Chen, 2011; Fuchs, Krivokapić, e Jannedy, 2010; Hofhuis, Gussenhoven e Rietveld, 1995; Mo e Cole, 2010; Silber-Varod, 2011 e 2013; Tyler, 2013; Wightman, Shattuck-Hufnagel, Ostendorf, e Price, 1992);
3. Aumento da taxa de articulação em posições posteriores à fronteira (Amir, Silber-Varod e Izre'el, 2004; Tyler, 2013);

4. Reset de f0 (Hermes, 2006; Thorsen, 1985/1986; Du Bois, Cumming e Schuetz-Coburn, 1992; Couper-Kuhlen, 1986; Schuetze-Coburn, Shapley e Weber, 1991; Couper-Kuhlen, 2001);
5. Mudança abrupta na curva de f0 (Cruttenden, 1997);
6. Mudança de intensidade após a fronteira (Mo, 2008; Swerts, Collier e Terken, 1994; Tseng e Fu, 2005);
7. *Creaky voice* e outras qualidades de voz não modais (Dilley, Shattuck-Hufnagel, e Ostendorf, 1996; Gordon e Ladefoged, 2001; Hanson, Stevens, Kuo, Chen e Slifkaet, 2001; Zellers e Post, 2010).

Além disso, fatores semânticos, pragmáticos, sintáticos e interacionais contribuem na percepção auditiva de fronteira prosódica (Selting, 2000; Schegloff, 1998, Selkirk, 1984; Selkirk, 2005; Cresti, 2000; Halford, 1994 e 1996; Mukherjee, 2001; Croft, 1995/ 2007; Kibrik e Podlesskaya, 2009; Ono e Thompson, 1995; Park, 2002). Neste trabalho, são avaliados os parâmetros fonético-acústicos comumente associados à percepção das fronteiras e parâmetros extras relacionados aos parâmetros tradicionalmente estudados.

2.5. Combinações de parâmetros fonético-acústicos associados à percepção das fronteiras prosódicas

Evidentemente, a literatura da área dispõe de uma boa compreensão sobre os parâmetros fonético-acústicos envolvidos na percepção de fronteira prosódica, quando observados individualmente. Até o momento, não está disponível uma abordagem teórico-metodológica que os integre em um único modelo coeso, pois o modo como os parâmetros se combinam em posição de fronteira percebida ainda não é bem compreendido. Com exceção da pausa física, a presença de um dos parâmetros tradicionalmente citados na literatura da área, ou mesmo mais de um, não veicula automaticamente a percepção de fronteira prosódica. É importante considerar que os parâmetros fonético-acústicos envolvidos na percepção de fronteira não são variáveis categóricas, mas sim variáveis contínuas. Assim, não há conhecimento sobre o quanto é necessário de cada parâmetro para que uma fronteira seja percebida pelos ouvintes. A presença de um dos parâmetros tradicionalmente citados na literatura da área, ou mesmo mais de um, também não parece ser suficiente para diferenciar as fronteiras percebidas entre si.

Em um estudo sobre a produção e a percepção de fronteiras prosódicas em hebraico (Amir et al., 2004), são avaliados quatro parâmetros fonético-acústicos: i) aumento da taxa de elocução em posições posteriores à fronteira; ii) alongamento em posições anteriores à fronteira; iii) reset de f0; iv) pausa física. São avaliadas 47 fronteiras prosódicas nas quais o acordo entre os quatro segmentadores quanto à percepção das fronteiras é total. Os resultados são: i) 11 fronteiras (24%) são marcadas por todos os quatro parâmetros avaliados; ii) 9 fronteiras (19%) são marcadas por três parâmetros; iii) 17 fronteiras (36%) são marcadas por dois parâmetros; iv) 8 fronteiras (17%) são marcadas por apenas um parâmetro; v) duas fronteiras (4%) não são marcadas por nenhum dos parâmetros avaliados. Apesar do trabalho de Amir et al. não avaliar todos os parâmetros frequentemente citados na literatura da área tal como foi apresentado na seção 2.4, observa-se que a percepção auditiva de fronteira prosódica é dada por uma combinação de parâmetros que pode sofrer variações inclusive entre fronteiras com acordo total quanto à percepção.

Collier et al. (1993) analisaram diferentes combinações de parâmetros fonético-acústicos associados às fronteiras prosódicas percebidas. No trabalho em discussão, foram analisados enunciados lidos. Em uma tarefa de percepção, foi solicitado que 19 sujeitos[8] indicassem, com base em uma escala de 10 pontos, o quão forte eles perceberam fronteiras prosódicas entre as palavras dos enunciados. A média das dezenove anotações em cada posição foi usada como referência para medir a força perceptual das fronteiras. Quanto maior a média, maior seria a força perceptual daquela fronteira. A análise fonético-acústica concentrou-se em verificar a duração da pausa física, a mudança no contorno entonacional de f0 e a declinação de f0 em cada posição percebida como fronteira entre as palavras.

Os resultados de Collier et al. (1993) mostram que algumas fronteiras percebidas se associam aos três parâmetros analisados. Outras, no entanto, se associam a um ou dois dos parâmetros analisados. De modo geral, os autores indicam que as fronteiras prosódicas com maiores valores de força perceptual estão associadas a uma maior quantidade de parâmetros. Aumentos acentuados na curva de f0 e pausas físicas de maior duração se associam principalmente às fronteiras com maior força perceptual. Collier et al. (1993) apresentam outras considerações importantes a respeito da combinação de parâmetros associados à percepção das fronteiras. As principais observações são que: 1) a pausa física é acompanhada por uma mudança no contorno entonacional de f0 em 80% das fronteiras percebidas; 2) as pausas físicas

[8] O trabalho não informou se os sujeitos que realizaram a tarefa são treinados ou não.

que não são acompanhados por uma mudança no contorno entonacional de f0 geralmente são menores do que 100 ms.

Em português brasileiro, o trabalho de Oliveira (2000) investiga parâmetros fonético-acústicos que se associam a fronteiras relevantes perceptualmente em narrativas espontâneas. Neste trabalho, as narrativas analisadas foram extraídas de "entrevistas espontâneas". As narrativas foram gravadas em um laboratório preparado para realizar gravações com qualidade adequada para análise acústica. O procedimento adotado consistiu em solicitar aos participantes voluntários que "falassem livremente" sobre um dos temas indicados em uma lista elaborada previamente. Os participantes não foram instruídos a contar histórias. De acordo com Oliveira (2000), as narrativas selecionadas para o estudo surgiram espontaneamente ao longo das gravações, embora a espontaneidade factual dos dados seja questionável. As fronteiras prosódicas foram marcadas perceptualmente por dois segmentadores treinados e os casos de desacordo entre os dois segmentadores foram decididos pelo autor da tese. No trabalho em questão, foi analisada a ocorrência simultânea de determinados parâmetros fonético-acústicos em posição de fronteira prosódica. Os parâmetros analisados foram a pausa física, o reset de f0, o tom da fronteira, a variação no tom e a taxa elocução.

Os resultados de Oliveira (2000) mostram que a presença de um dos parâmetros analisados não constitui evidência suficiente para uma fronteira prosódica ser percebida, isto é, a presença de um dos parâmetros não veicula necessariamente a percepção de fronteira prosódica. Geralmente, o único parâmetro que sempre veicula a percepção de fronteira é a pausa física. No trabalho em discussão, as disfluências não são consideradas fronteiras prosódicas e, por este motivo, determinadas pausas físicas não necessariamente veiculam a percepção de fronteira. Quanto à ocorrência simultânea de determinados parâmetros em posição de fronteira prosódica percebida, foi observado que alguns parâmetros coocorrem sistematicamente uns com os outros. Nas posições de fronteira percebida, a pausa física está frequentemente associada aos maiores valores de reset de f0 e aos tons de fronteira mais baixos.

A seguir, o quadro 1 apresenta outros trabalhos que investigam a combinação de parâmetros fonético-acústicos associados às fronteiras prosódicas em diferentes línguas.

Quadro 1: Combinações de parâmetros fonético-acústicos associados às fronteiras em diferentes línguas

Autor / Ano	Amostra / Tamanho da amostra	Variáveis analisadas	Língua	Combinação de parâmetros associados às fronteiras
Dominguéz, Farrús e Wanner / 2016	Fala espontânea / 5 minutos	F-AC	Espanhol Inglês	F0 Intensidade Duração
Apopei e Paduraru / 2015	Fala lida e fala eliciada / 10 minutos	F-AC	Romeno	F0 Intensidade Duração Pausa física
Beke, Szaszák e Váradi / 2013	Fala espontânea / 35 minutos	F-AC	Húngaro	F0 Intensidade
Rosenberg / 2009	Fala espontânea / 67 minutos	F-AC	Inglês	F0 Intensidade Duração
Hirschberg e Nakatani / 1998	Fala lida e fala eliciada	F-AC	Inglês	F0 Intensidade

Todos os trabalhos citados no quadro 1 tinham como objetivo desenvolver uma ferramenta de detecção automática de fronteiras prosódicas. O desempenho das ferramentas é apresentado posteriormente na seção 2.7. Os trabalhos partiram de uma segmentação realizada por segmentadores, desenvolveram as ferramentas de detecção automática e compararam a segmentação dos segmentadores com a segmentação obtida através das ferramentas.

O problema dos trabalhos em discussão reside no fato de que as fronteiras tomadas como referência para avaliar o desempenho das ferramentas não foram atribuídas com base na percepção auditiva de ouvintes. As fronteiras marcadas pelos segmentadores foram marcadas com base em critérios estritamente formais. Uma posição foi considerada uma fronteira prosódica, caso pelo menos um determinado parâmetro fonético-acústico tenha sido verificado no sinal sonoro. Com uma ou outra variação, os parâmetros considerados pelos segmentadores para anotar uma fronteira prosódica foram a presença de pausa física no sinal sonoro (a duração mínima não foi mencionada), o aumento repentino na curva entonacional de f0 e a queda acentuada de intensidade.

A impressão geral passada pelos trabalhos citados no quadro 1 é que eles limitaram o estudo à investigação da combinação de parâmetros com maior capacidade de reconhecer automaticamente as fronteiras marcadas pelos segmentadores, de acordo com os critérios formais estabelecidos. Isso significa que as fronteiras não foram analisadas do ponto de vista da percepção. Assim, a análise das combinações de parâmetros fonético-acústicos é vista simplesmente como um recurso/ instrumento que auxilia no desenvolvimento de ferramentas com excelente capacidade de identificar automaticamente as fronteiras abordadas nos trabalhos em questão.

Apesar dos trabalhos apresentados no quadro 1 não analisarem as fronteiras efetivamente percebidas pelos ouvintes, os autores acreditam que os parâmetros associados às fronteiras prosódicas estão marcados exclusivamente no sinal sonoro. Porém, ainda assim, há discussões quanto à combinação dos parâmetros. Os parâmetros de f0 e intensidade são unânimes, independentemente da língua. Os parâmetros de duração são importantes para parte dos autores. A pausa física é observada por apenas um dos trabalhos.

A seguir, o quadro 2 apresenta mais alguns trabalhos que abordam a combinação de parâmetros associados às fronteiras prosódicas em diferentes línguas.

Quadro 2: Combinações de parâmetros sintáticos e fonético-acústicos associados às fronteiras em diferentes línguas

Autor / Ano	**Amostra / Tamanho da amostra**	**Variáveis analisadas**	**Língua**	**Combinação de parâmetros associados às fronteiras**
Klejch, Bell e Renals / 2016	Mídia televisa / 15 horas	SIN F-AC	Inglês	F0 Intensidade Duração Sintaxe Pausa física
Margolis, Ostendorf e Livescu / 2010	Fala radiofônica e fala telefônica / 2 horas	SIN F-AC	Inglês	F0 Intensidade Duração Sintaxe
Heldner e Megyesi / 2003	Entrevista (Fala radiofônica) / 25 minutos	SIN F-AC	Sueco	Duração Pausa física Sintaxe

Nos trabalhos apresentados no quadro 2, foram desenvolvidas ferramentas de identificação automática de fronteira com alta *performance*. O desempenho destas ferramentas é apresentado detalhadamente na seção 2.7. As fronteiras consideradas como referência para estimar o poder de desempenho das ferramentas também foram atribuídas pelos segmentadores com base em critérios formais. A diferença é o acréscimo de critérios sintáticos. Naturalmente, permanece o problema de atribuir as fronteiras com base em critérios estritamente formais, isto é, não avaliar a percepção das fronteiras. Independente disso, é evidente a discussão sobre a combinação de parâmetros associados às fronteiras. Entre os trabalhos relatados no quadro 2, parâmetros duracionais e sintáticos são unânimes, independentemente da língua. Parâmetros de pausa física, f0 e intensidade são importantes para parte dos trabalhados relatados.

A combinação de parâmetros associados às fronteiras prosódicas é objeto de ampla discussão na literatura da área. A discussão ocorre entre aqueles que avaliam fronteiras prosódicas atribuídas com base em critérios formais, mas também entre aqueles que avaliam fronteiras prosódicas atribuídas com base na percepção auditiva de ouvintes. A percepção

humana compreende um sistema global, que observa de forma integral todos os parâmetros fonético-acústicos codificados no sinal sonoro. Desta forma, todos os parâmetros fonético-acústicos contribuem para uma dada sensação auditiva.

Experimentos mostram que vários parâmetros fonético-acústicos se associam à percepção de um determinado fenômeno perceptual, mas o peso de certos parâmetros é maior do que o peso de outros no processo de percepção do fenômeno (Boersma, 2005; Cho e McQueen, 2006; Escudero, 2005; Gottfried e Beddor, 1988; Harnsberger, 2001; McGuire, 2007). Numerosos estudos sugerem que o peso relativo dos parâmetros essenciais para a percepção de fenômenos de fala varia entre línguas (Escudero, Benders e Lipski, 2009; Kluender, Lotto, Holt e Bloedel, 1998; Holt, Lotto e Kluender, 2001; Jusczyk, 1993). Assim, as diferenças quanto aos pesos dos parâmetros bem como as diferentes combinações de parâmetros podem ser devidas às diferentes línguas, à modalidade de fala (se fala espontânea, lida, eliciada etc.) e ao tipo de fronteira analisada. Possivelmente, o peso dos parâmetros que contribuem para a percepção de fenômenos perceptuais varia entre diferentes casos de um único fenômeno e também de acordo com os diferentes estilos de fala e diastratias (gênero, idade, nível cultural, características individuais etc).

Atualmente, poucos trabalhos da literatura da área investigam o peso relativo dos parâmetros envolvidos na percepção das fronteiras. Outro ponto que precisa ser esclarecido na literatura diz respeito à posição exata de estabelecimento dos parâmetros fonético-acústicos que se associam à percepção das fronteiras prosódicas. As posições exatas das fronteiras prosódicas são atribuídas fonologicamente pelos falantes. Geralmente, os segmentadores consideram que as posições de percepção das fronteiras prosódicas coincidem com o final de uma palavra fonológica. Os parâmetros que levam à percepção da fronteira prosódica, por sua vez, podem ocorrer em uma região ou posição que não coincide exatamente com o fim de palavra fonológica.

As questões relativas às posições de estabelecimento dos parâmetros são diversas: os parâmetros se estabelecem na posição exata em que os segmentadores atribuem a percepção de fronteira? Os parâmetros se estabelecem um pouco antes ou um pouco depois da posição em que os segmentadores atribuem a percepção de fronteira? Todos os parâmetros envolvidos são estabelecidos em uma única posição ou se distribuem em posições distintas, a depender do parâmetro? A posição em que os parâmetros se estabelecem pode variar um pouco? Possivelmente, os parâmetros são estabelecidos em posições próximas à posição em que os segmentadores atribuem a percepção de fronteira e na própria posição em que os segmentadores

atribuem a percepção de fronteira. Porém, não está claro quantas e quais posições precisam ser observadas para compreender os parâmetros essenciais para a percepção das fronteiras.

É importante discutir sobre os métodos usados para calcular os parâmetros fonético-acústicos. As medidas de intensidade, por exemplo, são muito influenciadas pela posição que o microfone se posiciona em relação ao sujeito gravado. Sendo assim, é importante que as medidas de intensidade sejam medidas de intensidade relativa. A análise consistente de fenômenos duracionais precisa ser feita por meio de um procedimento de normalização dos dados para minimizar os efeitos da duração intrínseca dos fones (um segmento pode durar mais simplesmente porque ele é, em média, intrinsecamente mais longo). No entanto, em ambos os casos, os procedimentos adequados para cálculo dos parâmetros de intensidade e duração nem sempre são feitos, o que pode afetar a conclusão final das análises. A rigor, seria interessante normalizar também a frequência fundamental para evitar distorções nos valores de f0 causadas pelas características das pregas vocais de cada falante (Gussenhoven, 2004), porém, a normalização de f0 também não costuma ser feita.

Em alguns trabalhos, os tons altos (com curva de frequência fundamental ascendente) são frequentemente associados às fronteiras que veiculam a percepção de continuidade. Os tons baixos, por sua vez, são associados às fronteiras que veiculam a percepção de terminalidade (Brown, Currie e Kenworthy, 1980; Blaauw, 1995; Swerts e Geluykens, 1994; Swerts, Geluykens e Terken, 1992; Van Donzel, 1999; Wichmann, House e Rietveld, 1997). Se observamos a fronteira que veicula a percepção de continuidade (NTB), muitas vezes podemos visualizar o aumento de frequência fundamental (f0), isto é, o tom alto. Contudo, há diversos casos de fronteiras que sinalizam perceptualmente o valor de continuidade e não são marcadas por esse aumento de f0. Desta forma, fronteiras classificadas perceptualmente como fronteiras do mesmo tipo se associam em alguma medida a parâmetros diferentes.

De modo geral, a literatura da área é útil e representa um grande avanço no que diz respeito aos parâmetros envolvidos na percepção das fronteiras. Porém, todas as questões apresentadas ao longo desta seção sugerem que ainda não há conhecimento sólido sobre como a composição/ combinação de parâmetros fonético-acústicos é formada para orientar a produção e a percepção de diferentes tipos de fronteiras. A literatura não é abrangente a ponto de entendermos o que pesa mais ao percebermos fronteiras prosódicas, pois inclusive os trabalhos que tratam da combinação de parâmetros não costumam analisar o peso de cada parâmetro envolvido na percepção das fronteiras (Auer, 2010). Por fim, a literatura sugere que é importante ampliar a análise, incluindo outros parâmetros além dos parâmetros

tradicionalmente estudados, pois algumas fronteiras não são marcadas por nenhum dos parâmetros tradicionalmente estudados.

Neste trabalho, é analisada a combinação de parâmetros fonético-acústicos associados à percepção das fronteiras terminais e não-terminais. Também é analisado o peso dos parâmetros fonético-acústicos envolvidos na percepção das fronteiras em discussão. A localização dos parâmetros fonético-acústicos é avaliada por meio da extração dos parâmetros em diferentes posições ao longo do sinal sonoro. Todas as análises citadas foram feitas com base em diferentes modelos e submodelos destinados à identificação automática das fronteiras percebidas.

2.6. A pausa e a sua importância para as fronteiras prosódicas

Nesta seção, são apresentadas diversas propostas de definição de "pausa" na literatura da área, é discutido o papel da pausa na percepção das fronteiras e no desenvolvimento da ferramenta de identificação automática de fronteiras prosódicas e, principalmente, se apresentam os procedimentos adotados para chegar à definição da duração mínima de interrupção do fluxo de fala que possa ao mesmo tempo ser percebida como pausa e implementada em um sistema de reconhecimento automático, minimizando a maior quantidade de erros. De fato, uma duração longa demais causaria a não detecção de muitas interrupções do fluxo de fala que são percebidas como pausa pelos ouvintes, enquanto uma duração curta demais identificaria como pausa diversas fases de oclusão.

2.6.1. Noção de pausa proposta na literatura

Na literatura, o termo "pausa" é usado para se referir ao silêncio entre dois trechos de fala, isto é, a interrupção da vocalização observada e mensurada acusticamente no sinal sonoro. O termo é usado também para se referir à percepção auditiva de interrupção do fluxo de fala. O termo em questão também é usado para se referir a sons que sinalizam a tomada de tempo para buscar itens lexicais, manter ou tomar o turno de fala. Neste caso, os autores adicionam um adjetivo ao termo e o fenômeno é chamado de "pausa preenchida". Alguns autores utilizam o termo "pausa" para se referir a pausas virtuais, que não necessariamente correspondem a uma interrupção do fluxo de fala, mas sinalizam uma pausa "cognitiva".

Sempre que uma oclusiva é produzida, é verificada uma porção de silêncio no sinal sonoro. O silêncio relativo às oclusões não é considerado uma pausa em nenhuma das acepções de "pausa" citadas no parágrafo anterior. Na literatura da área, geralmente, assume-se que os silêncios devidos às oclusões podem ser facilmente diferenciados das pausas verificadas e percebidas no sinal sonoro ao estabelecer um determinado limite de duração. A discussão em torno da pausa é, de fato, extensa e intensa. A discussão gira em torno dos diversos fenômenos que são designados com o termo "pausa", mas também há uma ampla discussão a respeito do limite usado para diferenciar o silêncio das oclusões das pausas observadas no sinal sonoro e percebidas pelo sistema auditivo humano.

Os fenômenos diversos que remetem ao termo "pausa" não são equivalentes e alguns autores apresentam claramente os seus conceitos de "pausa". Nespor e Vogel (1986), por exemplo, consideram as pausas perceptuais e as pausas reais. A pausa perceptual refere-se ao que é efetivamente percebido como pausa, podendo corresponder a uma variedade de fenômenos, como a mudança de *pitch*, a mudança de duração e a cessão da fonação. A pausa real refere-se à porção de silêncio no sinal sonoro, que não é devida às oclusões dos segmentos. As noções de pausa propostas em Zellner (1994) se assemelham muito com as noções propostas por Nespor e Vogel (1986). Zellner (1994) propõe o uso do termo "pausa física" para se referir ao silêncio entre dois trechos de fala e "pausa percebida" para se referir à interrupção do fluxo de fala identificável perceptualmente pelos ouvintes. Zellner (1994) sugere que as pausas físicas percebidas como interrupção de fluxo de fala com maior facilidade apresentam duração entre 200 ms e 250 ms.

Neste trabalho, o silêncio entre dois trechos de fala é visto como um fenômeno acústico que ocorre no sinal sonoro. A pausa é vista como um fenômeno perceptual capturado pelo sistema auditivo a partir da avaliação do sinal sonoro. Como já foi dito, para evitar ambiguidades, o termo "pausa percebida" é usado para se referir aos silêncios efetivamente percebidos pelos ouvintes como interrupção do fluxo de fala. O termo "pausa física" é usado para se referir ao fenômeno fonético de interrupção do fluxo de fala. Obviamente, os períodos de silêncio referentes à oclusão das consoantes oclusivas não são tratados como pausa física e nem como pausa perceptual.

Em alguns trabalhos, embora estejam implícitas as noções de pausa física e percebida, nem sempre são usados termos diferentes para tratar dos fenômenos em discussão. Com o objetivo de evitar confusão entre os diferentes fenômenos designados com o termo "pausa", neste trabalho, os termos "pausa física" e "pausa percebida" também são usados ao discutir sobre trabalhos que não adotam essa terminologia, mas observam claramente a distinção entre

os fenômenos físico e perceptual. Adicionalmente, o termo “pausa física” é usado para fazer referência a trabalhos que avaliam a pausa apenas do ponto de vista da interrupção do fluxo de fala no sinal sonoro.

Em alguns trabalhos sobre o tema, as noções de pausa física e pausa percebida nem sempre são consideradas. Frequentemente, é utilizado o termo “pausa” de forma genérica. Geralmente, o fenômeno que é chamado genericamente de “pausa” compreende o que é chamado de pausa física neste trabalho. As “pausas” relatadas não são atribuídas com base em alguma consideração perceptual, mas sim com base em uma decisão a respeito da duração mínima de pausa física necessária para constituir a “pausa” abordada. O motivo de escolher uma ou outra duração não é sempre explicado. Em alguns casos, a duração mínima é escolhida por razões relacionadas à metodologia empregada.

O trabalho de Siegman e Pope (1966) é um exemplo de trabalho que escolhe uma duração para a “pausa”, mas não explica a razão da escolha. O trabalho em questão não tem como foco investigar as pausas físicas e perceptuais. A percepção de pausa sequer é mencionada no trabalho. Os autores estabeleceram uma duração mínima, porque o foco do trabalho era analisar os níveis de fluência verbal medidos com base nas diferentes taxas de elocução e articulação. Portanto, foi necessário estabelecer o que era a “pausa” para que os parâmetros de interesse fossem calculados excluindo ou incluindo a duração das pausas consideradas. A duração escolhida por Siegman e Pope (1966) era de no mínimo 3 segundos de duração.

Hargreaves e Starkweather (1959) desenvolveram um circuito eletrônico de segmentação da fala em unidades. O circuito proposto funcionava delimitando uma unidade a cada pausa física de determinada duração. Foram testadas duas durações diferentes; as durações eram 500 e 1000 ms. Os autores utilizaram apenas o termo “pausa” de modo genérico, mas a percepção de interrupção de fluxo de fala foi avaliada por meio de um experimento. O objetivo do experimento era verificar se os sujeitos perceberiam como pausa as pausas físicas de pelo menos 500 ou 1000 ms. A tarefa dos sujeitos consistia em indicar as posições de percepção de pausa por meio do acionamento de um botão logo após o momento que perceberam a pausa. A avaliação perceptual dos sujeitos não foi considerada confiável por se tratar de uma tarefa que o tempo de reação influenciava os resultados do experimento. Em virtude disso, por segurança, os autores optaram por segmentar a fala a cada 1000 ms de pausa física.

Em um trabalho sobre disfluências em textos de fala (Lay e Paivio, 1969), foram avaliadas “pausas” e pausas preenchidas. As “pausas” retratadas no trabalho em discussão são as pausas físicas. As pausas físicas analisadas foram apenas aquelas que apresentaram duração igual ou superior a 1000 ms, porque o dispositivo de medir a duração apresentou *delay* ao medir

pausas físicas com duração inferior a 1000 ms. Portanto, as pausas efetivamente percebidas não foram avaliadas em nenhum momento. Foram analisados textos de fala eliciada. Alguns dos elementos avaliados foram: i) a taxa de pausa preenchida por palavra (número de pausa preenchida por palavra); ii) a taxa de pausa física por palavra (número de pausa física por palavra); iii) a taxa de repetições (calculada dividindo o número de palavras repetidas pelo número total de palavras); iv) a taxa de pausas físicas por segundo (calculada dividindo a duração das pausas físicas pela duração total do áudio); iv) a taxa de pausas preenchidas por segundo (duração das pausas preenchidas dividida pela duração total do áudio) etc.

Levin e Silverman (1965) é outro exemplo de trabalho dedicado ao estudo das disfluências. O foco era compreender elementos que poderiam estar associados às disfluências. Um dos elementos avaliados foi chamado genericamente de "pausa". A percepção de interrupção do fluxo de fala não foi considerada no estudo. As "pausas" consideradas no trabalho em discussão são as pausas físicas com duração acima de 1000 ms. Esta duração mínima foi escolhida, porque o dispositivo de medir a duração não foi capaz de medir com acurácia pausas físicas com duração abaixo de 1000 ms. Portanto, trata-se de um estudo em que a definição da duração mínima foi motivada por uma dificuldade metodológica da época.

O limiar de percepção de pausa, de acordo com Barik (1969), é entre 500 e 600 ms. No trabalho em questão, a duração mínima foi atribuída com base em um experimento de percepção auditiva de pausa em textos de fala lida e "histórias contadas livremente". Em razão da metodologia empregada, a duração mínima proposta pelo autor pode ser questionada. Barik decidiu eliminar dos dados ruídos "estranhos" eventualmente gravados. A eliminação dos ruídos foi feita em uma fita duplicada dos áudios (cópia dos áudios originais). O procedimento de eliminação dos ruídos não foi muito detalhado no trabalho. O trabalho apenas informou que os ruídos foram eliminados com base no acionamento de um botão ao ouvir os ruídos na versão copiada dos áudios. O sujeito que indicou os ruídos não foi especificado. Apesar da operação de eliminação de ruídos não ter sido apresentada claramente, há um risco de os trechos eliminados não corresponderem exatamente aos períodos de ruído no sinal sonoro, pois é gasto um certo tempo para acionar o botão após a percepção do ruído ocorrer. Uma consequência evidente da exclusão dos ruídos é a edição dos áudios. Não foi mencionado o quanto os áudios foram editados de modo geral. O principal problema do trabalho é, de fato, a edição dos áudios, pois, no experimento de percepção de pausa, os sujeitos avaliaram uma versão editada dos áudios, o que pode comprometer a integridade dos áudios e os resultados do experimento.

Goldman-Eisler (1958) investigou as funções sintáticas associadas às pausas físicas. A percepção auditiva de pausa não foi considerada no trabalho. A autora adotou como referência

a duração mínima de 250 ms de pausa física. No trabalho em questão, foi realizado um experimento para avaliar as funções sintáticas das pausas físicas de pelo menos 250 ms. A autora explica que optou pela duração mínima de 250 ms, porque as pausas físicas que apresentam duração inferior a 250 ms fazem parte da articulação das consoantes oclusivas. Em geral, o raciocínio de Goldman-Eisler resultou em uma visão padrão na qual as "pausas físicas mais curtas exprimem apenas questões articulatórias." Porém, a literatura mostrou que a duração dos períodos de silêncio referente às fases de oclusão varia (Lisker, 1957). Campione e Véronis (2002) mostraram que silêncios de aproximadamente 60 ms nem sempre fazem parte dos segmentos oclusivos (Campione e Véronis, 2002). Na literatura, também foi reconhecido que a maioria das pausas físicas com durações entre 130 ms a 250 ms não é devida a questões articulatórias apenas (Hieke, Kowal e O'Connell, 1983).

O trabalho de Sabin et al. (1979) tinha como objetivo analisar características de textos de fala lida e eliciada. Alguns elementos avaliados eram as pausas físicas de pelo menos 270 ms e as taxas de articulação e elocução dos textos. A impressão geral é que a duração mínima de pausa física foi atribuída apenas com intuito de determinar o que deveria ser considerado ao calcular as taxas de elocução e articulação. O trabalho em questão até forneceu uma justifica para o limiar de duração escolhido, porém, se limitou a dizer que a escolha da duração mínima seguiu parcialmente a proposta de Goldman-Eisler (1958). Não foi explicado por qual motivo foi adotado 270 ms ao invés de 250 ms tal como foi proposto por Goldman-Eisler. Independente disso, ao declarar a adoção parcial da proposta de Goldman-Eisler, as questões direcionadas à referência citada são automaticamente aplicáveis ao trabalho de Sabin et al. também.

O trabalho de Kowal e O'Connell (1987) tinha como objetivo analisar a organização temporal de textos narrativos monológicos. Os sujeitos assistiram a um filme e depois recontaram (narraram) a história do filme com as suas próprias palavras. Neste trabalho, um dos elementos avaliados foi chamado genericamente de "pausa". A "pausa" abordada no trabalho compreendia as interrupções de vocalização com duração mínima de 230 ms. A percepção auditiva de pausa não foi considerada em nenhum momento. Assim, a noção de distinção entre as pausas física e percebida não foi considerada. Além de o trabalho não fazer nenhuma menção à percepção, não foi explicado por qual motivo a duração mínima de 230 ms de pausa física foi escolhida. A organização dos textos narrativos orais foi avaliada através do cálculo de alguns parâmetros fonéticos tais como as taxas de elocução e articulação. Mais uma vez, a impressão geral é que a duração mínima foi escolhida apenas com intuito de determinar o que deveria ser considerado ao calcular as taxas de elocução e articulação.

Em Boomer (1965), a ocorrência de pausas físicas e preenchidas foi examinada quanto à sua localização nas orações. Neste trabalho, foram analisadas entrevistas sobre temas variados e a percepção auditiva da pausa não foi considerada em nenhum momento. As pausas físicas consideradas são aquelas que apresentam pelo menos 200 ms. Não foi explicado por qual motivo a duração mínima de 200 ms de pausa física foi escolhida. Levin et al. (1967) analisam disfluências produzidas por crianças em textos de fala eliciada. Um dos elementos analisados ao redor das disfluências foram as pausas físicas de pelo menos 80 ms. A percepção de pausa não foi abordada e não foi explicado por qual motivo a duração mínima de 80 ms de pausa física foi escolhida. Em Henderson et al. (1966), foram avaliados textos de fala lida e entrevistas. O trabalho tinha como objetivo analisar padrões de distribuição de pausas físicas na fala lida e nas entrevistas. Foram consideradas apenas as pausas físicas de pelo menos 100 ms. A percepção auditiva de pausa não foi considerada e não foi apresentada uma explicação a respeito do valor mínimo de pausa física.

Em Agnello (1963), são analisados dados de fala lida. Dezoito sujeitos gravaram a leitura dos textos. Em seguida, os dezoito sujeitos realizaram a tarefa de percepção de pausa nas gravações, reportando as posições em que perceberam pausas. A tarefa consistia em inserir na transcrição dos trechos vírgulas para representar as posições de percepção de pausa. No trabalho, assume-se que as pausas físicas menores do que 150 ms não são percebidas como pausa pelos ouvintes. Os resultados apresentados em Agnello (1963) mostram que a duração mínima de pausa física necessária para a percepção auditiva de pausa é cerca de 190 ms.

Em Ruder (1973), são analisadas sentenças de fala lida. Neste trabalho, o foco era avaliar a percepção de pausa em posição de determinadas palavras que se repetiam nas sentenças. No total, quatro sujeitos gravaram a leitura de 5 sentenças. A tarefa de percepção de pausa foi realizada por 30 sujeitos dentre os quais 26 homens e 4 mulheres com idade entre 21 e 35 anos. A percepção de pausa foi avaliada por meio de um dispositivo eletrônico de ajuste manual. A tarefa dos sujeitos consistia em acionar um botão, se percebessem pausa após as palavras que se repetiam nas sentenças, e arrastar este botão até a posição em que julgavam ser o final da pausa. O tempo para realizar os ajustes manuais era livre. A análise estatística consistia em mensurar a percepção de pausa a partir dos ajustes manuais realizados pelos sujeitos. Foi realizado um procedimento estatístico para converter a extensão dos ajustes manuais em milissegundos. Em seguida, foi calculada a média dos ajustes manuais que representavam as pausas percebidas. Os resultados mostraram que a duração média de pausa física necessária para a percepção da pausa é cerca de 95 ms.

Em Butcher (1981), alguns sujeitos gravaram a leitura de um texto. Em seguida, os mesmos participantes recontaram o texto lido com as suas próprias palavras. A tarefa de percepção de pausa foi feita por 5 sujeitos. Os sujeitos analisaram o texto lido e a versão eliciada correspondente. Os resultados indicaram que, raramente, as pausas físicas com duração entre 100 ms e 200 ms foram percebidas como pausa. As pausas físicas com duração acima de 200 ms, por sua vez, foram percebidas como pausa com maior facilidade. Não é dito claramente que a duração mínima de pausa física necessária para a percepção de pausa é 200 ms. Da forma como os resultados são expostos, entende-se apenas que as pausas físicas são percebidas como pausa com maior ou menor facilidade a depender da duração da pausa física.

O trabalho de Butcher (1981) vai além e investiga se a duração dos segmentos anteriores à pausa física contribui na percepção da pausa. Os resultados mostraram que os segmentos anteriores às pausas percebidas costumam ser alongados, mas a duração do alongamento não se correlaciona proporcionalmente com a duração da pausa. A autora defende que o alongamento é o elemento que explica a percepção de pausa em posições sem pausa física. Porém, não há evidências suficientes para afirmar que a ausência de percepção de pausa em uma dada posição com pausa física é causada pela falta de alongamento nos segmentos anteriores. Adicionalmente, os resultados mostraram que algumas pausas identificadas perceptualmente não correspondem a nenhuma pausa física no sinal sonoro.

Walker e Trimboli (1982) analisaram 5 minutos de um diálogo entre duas mulheres. Os dados compreendem 30 pausas físicas com duração variando de 100 ms a 1000 ms. Na transcrição, cada “ponto de transição no diálogo”, excluindo as sobreposições, foi marcado por um quadrado. Os pontos de transição no diálogo aos quais Walker e Trimboli (1982) se referem parecem ser as posições de mudança de turno de fala. A tarefa de percepção de pausa consistia em acrescentar a letra S nas posições designadas pelo quadrado, caso os sujeitos percebessem pausa. Vinte estudantes do primeiro ano do curso de psicologia realizaram a tarefa de percepção de pausas. As pausas físicas foram divididas em seis diferentes intervalos de duração. A duração das pausas físicas variava de 100 ms até 1000 ms. Para cada intervalo estabelecido, os resultados foram calculados de modo que as pausas identificadas perceptualmente fossem aquelas em que pelo menos 50% dos sujeitos marcaram a percepção de pausa, isto é, pelo menos 10 estudantes. Os intervalos e a porcentagem de pausa percebida em cada intervalo são apresentados na tabela 1.

Tabela 1: Pausas percebidas (Walker e Trimboli, 1982)

Intervalo de duração de pausa física	Porcentagem de pausa percebida
100 ms a 199 ms	42%
200 ms a 299 ms	52%
300 ms a 399 ms	69%
400 ms a 499 ms	77%
500 ms a 599 ms	88%
600 ms a 1000 ms	93%

Em Duez (1985), são analisados textos de fala chamados genericamente de “discursos políticos”, “entrevistas políticas” e “entrevistas casuais”. No experimento, 10 sujeitos com aproximadamente 35 anos sem nenhum tipo de distúrbio auditivo anotaram as pausas percebidas. A autora divide as pausas percebidas da seguinte forma:

i) Pausas com baixa saliência perceptual: Entre 1 e 3 sujeitos percebem pausa, a duração da pausa é abaixo de 400 ms;
ii) Pausas com saliência perceptual intermediária: Entre 4 e 8 sujeitos percebem pausa, a duração da pausa varia de 400 ms a 600 ms;
iii) Pausas com saliência perceptual alta: Entre 9 ou 10 sujeitos percebem pausa, duração da pausa é acima de 600 ms;

Dois casos interessantes no que diz respeito à percepção de pausa são encontrados nos experimentos realizados em Duez (1985). O primeiro caso compreende uma pausa física de 290 ms, que não é percebida como pausa por nenhum sujeito; o segundo caso compreende uma pausa física de 260 ms, que é percebida como pausa pelos 10 sujeitos que fizeram o experimento. Os dois casos em discussão podem ser vistos como evidências de que outros fatores, além da duração da pausa física, contribuem na percepção auditiva de pausa, pois pausas físicas parecidas no que diz respeito à duração podem ser percebidas por um maior ou menor número de sujeitos. O trabalho de Duez é um pouco mais amplo e busca avaliar o quanto a duração dos segmentos anteriores à pausa física influencia a percepção auditiva de pausa. Assim como em Butcher (1981), o alongamento da vogal anterior à pausa física é um elemento que auxilia na percepção auditiva de pausa. No trabalho em discussão, os resultados sugerem que a influência do alongamento aumenta à medida que a duração da pausa física diminui.

Um trabalho mais recente (Heldner, 2011) tem como foco investigar a percepção de pausa e a percepção de mudança de turno de fala. Os dados compreendem 10 minutos de um diálogo entre mulheres e 10 minutos de um diálogo entre homens. Neste trabalho, os dados de fala foram recortados em trechos menores. A duração dos trechos varia de 0,6 s a 6,5 s com média de 2,6 s e desvio-padrão de 1,3 s. No total, são analisados 70 estímulos. Dentre eles, 35 compreendem pausas físicas no sinal sonoro e 35 compreendem sobreposições de fala. A

duração das pausas físicas e sobreposições, em ambos os casos, varia de 5 ms a 400 ms. Todos os 70 estímulos analisados são posições em final de turno. A tarefa de percepção de pausa foi feita por 25 sujeitos dentre os quais 9 mulheres e 16 homens com idade entre 32 e 75 anos.

Na tarefa de percepção realizada em Heldner (2011), os sujeitos precisavam indicar a sua percepção de pausa e de sobreposição. Assim, a princípio, para cada posição, os sujeitos indicavam a sua percepção de pausa e, em seguida, indicavam a sua percepção de sobreposição. Os sujeitos não tinham acesso ao espectrograma, cada estímulo foi executado apenas uma vez. Neste momento, é apresentada apenas a análise estatística relativa à percepção de pausa. Um procedimento de regressão logística binária com seleção de variáveis com base na razão de verossimilhança foi usado para estimar a probabilidade de percepção de pausa dada a duração da pausa física no sinal sonoro. Evidentemente, os resultados mostraram que o número de pausas identificadas perceptualmente está correlacionado com a duração da pausa física. A duração mínima de pausa percebida é aproximadamente 120 ms, mas, de acordo com autor, o valor pode variar bastante entre diferentes sujeitos falantes.

Outro trabalho que avalia a percepção de pausa é Martin (1970). Alguns sujeitos receberam imagens para observar. Em seguida, os sujeitos tinham como tarefa descrever oralmente as imagens recebidas. Os áudios analisados, portanto, compreendem textos de fala semiespontânea, em que todos eram induzidos a realizar uma tarefa específica. Para mais detalhes sobre os áudios, veja-se Martin e Strange (1968). Dois sujeitos, diferentes daqueles que gravaram os textos, realizaram a marcação das pausas percebidas. Os resultados mostram que a grande parte das pausas identificada perceptualmente corresponde pausas físicas no sinal sonoro; a duração das pausas físicas identificada perceptualmente varia de 50 ms a 4970 ms. Algumas pausas identificadas perceptualmente também não correspondem a nenhum instante de pausa física no sinal sonoro, mostrando que o fenômeno perceptual se difere do fenômeno físico, apesar de um ajudar o outro. Alguns casos de pausa física com duração entre 50 ms e 110 ms não são percebidos como pausa. Para Martin, a percepção de pausa em posição de pausa física com duração entre 50 ms a 110 ms depende de pistas linguísticas adicionais, que esclarecem para o ouvinte se as pausas físicas entre 50 e 110 ms devem ou não serem interpretadas como pausa.

Com objetivos diferentes dos demais autores citados ao longo desta seção, Belz e Trouvain (2019) também analisam a percepção de pausa. Neste trabalho, o foco principal é verificar se as posições percebidas como pausa sempre compreendem posições de pausa física. Os resultados mostraram que algumas pausas percebidas pelos sujeitos são acompanhadas por eventos sutis de atividade respiratória e articulatória sem vibração das cordas vocais. Os itens

mais frequentes são ruídos respiratórios, especialmente inalação oral, e sons presumivelmente atribuídos à atividade glótica. Deste modo, os resultados encontrados corroboram a tese de que as pausas percebidas pelos ouvintes podem ser posições nas quais não é observada a completa ausência de vocalização.

Em geral, a maioria dos trabalhos sobre o tema analisa dados de fala lida ou fala eliciada. Alguns trabalhos analisam apenas as pausas físicas. Outros trabalhos, por sua vez, analisam as pausas percebidas. Evidentemente, aqueles que analisam o fenômeno com base em experimentos e testes de percepção tendem a apresentar trabalhos mais consistentes, porque fazem considerações verdadeiramente perceptuais.

Entre os trabalhos que analisam apenas as pausas físicas, não há um amplo consenso a respeito da duração mínima para constituir uma pausa física e evitar confusão com os períodos de silêncio relativos à fase oclusiva dos segmentos. Também não há um amplo consenso a respeito da duração mínima necessária para a pausa física ser percebida como pausa pelos ouvintes e evitar confusão com os períodos de silêncio relativos à fase oclusiva dos segmentos. Portanto, a duração mínima é em si objeto de discussão.

Alguns trabalhos são peculiares quanto à análise da percepção auditiva de pausa. Estes trabalhos avaliam a percepção auditiva de pausa, estabelecem a duração mínima de pausa percebida a partir de algum experimento perceptual, mas assumem de antemão que o sistema auditivo humano não é capaz de perceber como pausa pausas físicas de determinada duração. Desta forma, certas pausas físicas sequer são avaliadas no experimento de percepção de pausa em razão do pressuposto citado.

Este trabalho analisa a percepção de pausa em dados de fala espontânea em PB, contribuindo nas discussões da literatura da área. Na maioria dos trabalhos sobre a percepção de pausa, o único elemento investigado é a duração da pausa física observada no sinal sonoro, isto é, poucos trabalhos investigam a contribuição de outros elementos na percepção auditiva de pausa. Neste trabalho, é investigado se outros elementos, além da duração da pausa física, contribuem na percepção auditiva de pausa. Os demais elementos investigados são: o tipo de fronteira prosódica associado à pausa física; a taxa de articulação dos textos; alguns parâmetros segmentais e suprassegmentais em posições anteriores e posteriores à pausa física. Todos os elementos investigados são apresentados detalhadamente no capítulo 5.

2.6.2. O papel da pausa física na percepção das fronteiras prosódicas

Como já foi dito, neste trabalho, o termo “pausa física” é usado para se referir ao fenômeno fonético mensurável e verificado no sinal sonoro, enquanto o termo “pausa percebida” é usado para se referir ao fenômeno perceptual graças ao qual o ouvinte tem a percepção auditiva de interrupção do fluxo de fala. Dentre todos os parâmetros, a pausa física é seguramente o mais citado na literatura da área como parâmetro associado à percepção de fronteiras prosódicas, independentemente da abordagem teórica (Tyler, 2013; Krivokapić, 2007; Swerts et al., 1992; Swerts, 1997; Carlson et al., 2005; Campbell, 1993; Swerts e Geluykens, 1994; Schuetze-Coburn et al., 1991; Brierley e Atwell, 2007; Ananthakrishnan e Narayanan, 2008; Ostendorf e Veilleux, 1994; Tseng e Fu, 2005; Mo e Cole, 2010; Couper-Kuhlen, 2006).

A pausa física é de fato uma condição suficiente para veicular a percepção de fronteira prosódica. Porém, atualmente, não há muito conhecimento sobre como ela se relaciona aos diferentes tipos de fronteira prosódica. Na literatura da área, os trabalhos que analisam este tópico ainda são incipientes. Dentre os poucos trabalhos disponíveis, é reconhecido que a pausa física sozinha não é capaz de indicar o tipo de fronteira prosódica (Raso et al., 2015; Shriberg, 2005; Tseng e Chang, 2008). É importante lembrar também que a pausa física não é estabelecida diante de qualquer fronteira, pois, na fala espontânea, muitas fronteiras não se correlacionam com este parâmetro (Teixeira, 2018; Raso et al., 2020). Raso et al. (2015) também mostram que não é possível determinar se uma fronteira é terminal ou não-terminal orientando-se por uma duração específica de pausa física, porque esta varia muito em posição de qualquer tipo de fronteira. Desta forma, a pausa física não pode ser vista como um forte correlato da natureza funcional da fronteira.

2.6.3. A importância da pausa percebida para o desenvolvimento da ferramenta de identificação automática de fronteira prosódica

Do ponto de vista da implementação de uma ferramenta de identificação automática de fronteiras, a pausa percebida é um recurso simples de ser implementado, porque requer apenas que seja informado a duração mínima de pausa física necessária para constituir uma pausa percebida e, consequentemente, estabelecer uma fronteira também. No entanto, como já foi dito, determinar esta duração não é simples e a pausa física por si só não é capaz de distinguir os diferentes tipos de fronteiras prosódicas. Estudos anteriores (Raso et al., 2020; Teixeira, 2018) mostram que modelos desenvolvidos com a finalidade de identificar automaticamente

fronteiras terminais e não-terminais em dados de fala espontânea de PB frequentemente tendem a superestimar a relevância da pausa física como recurso de identificação das fronteiras. A superestimação é clara nos modelos de identificação de TB e NTB, pois, em ambos os modelos, a grande parte das fronteiras devidamente identificada é acompanhada por pausa física e não parece existir uma correlação suficientemente forte entre a duração da pausa física e a natureza da fronteira. A superestimação da pausa física nos modelos TB e NTB também é evidenciada por meio da produção de muitos falsos alarmes em posições de fronteira com pausa física. A depender do modelo, os falsos alarmes podem ser posições de TB com pausa física que são indevidamente marcadas como NTB e vice-versa.

Tendo em vista os resultados expostos em Raso et al. (2020), Teixeira (2018) e Raso et al. (2015), é necessário desenvolver estratégias para solucionar ou, pelo menos, minimizar os efeitos indesejados causados pelos parâmetros de pausa física. Neste contexto de valorização excessiva dos parâmetros de pausa física, é fundamental refletir se a anotação de pausas percebidas em posição de TB e NTB seria útil no desenvolvimento da ferramenta de identificação automática de fronteira prosódica. Como a percepção de interrupção de fluxo de fala se trata de um parâmetro altamente perceptível por anotadores humanos (treinados ou não), isso pode ser útil. A hipótese subjacente é que os anotadores seriam capazes de anotar quando percebem uma pausa em posição de fronteira prosódica.

Então, a percepção auditiva de pausa constitui um elemento candidato para separar os dados nas seguintes subcategorias: fronteira terminal com pausa percebida; fronteira terminal sem pausa percebida; fronteira não-terminal com pausa percebida; fronteira não-terminal sem pausa. Com a subdivisão perceptual das fronteiras em fronteiras com e sem pausa percebida, a proposta é desenvolver submodelos destinados à identificação das subcategorias de fronteiras relevantes perceptualmente para os ouvintes. O resultado esperado do processo de subdivisão perceptual das fronteiras e desenvolvimento de submodelos é eliminar ou, pelo menos, reduzir a supervalorização dos parâmetros de pausa física no processo de reconhecimento automático das fronteiras.

2.7. Ferramentas de segmentação automática da fala

Uma revisão de literatura da área mostra que já foram desenvolvidas algumas ferramentas computacionais para segmentar o sinal acústico em unidades menores que a palavra. Os *scripts* para Praat BeatExtractor (Barbosa, 2004) e VowelDetector (Arantes, 2021)

segmentam automaticamente o sinal acústico em *onsets* de vogais consecutivos, isto é, em unidades V-Vs. A segmentação em unidades V-Vs é útil para analisar posteriormente evolução da duração ao longo dos enunciados falados. Para a segmentação fonética da fala, também há outras ferramentas disponíveis (Svendsen e Soong, 1987; Van Hemert, 1991; Selmini e Violaro, 2008). Uma segmentação no nível fonético é útil para diversos propósitos, em especial para a extração de dados relativos a parâmetros como duração, frequência fundamental e intensidade nos segmentos. No entanto, este tipo de segmentação não parece ser muito útil para determinar as posições das fronteiras prosódicas.

A ferramenta EasyAlign (Goldman, 2011) para Praat é uma ferramenta de segmentação automática da fala de fácil utilização. Com esta ferramenta, é possível segmentar o fluxo da fala em fones, sílabas e palavras a partir de poucas etapas manuais. O resultado é uma anotação de vários níveis dentro de um TextGrid composto por camadas fonéticas, silábicas, lexicais e de enunciados (informados previamente manualmente). Contudo, a ferramenta EasyAlign apresenta alguns problemas. Primeiramente, o EasyAlign utiliza obrigatoriamente como input uma transcrição de base ortográfica dos enunciados. Dada a sua natureza, uma transcrição ortográfica não é capaz de representar totalmente o que foi dito oralmente pelos falantes, o que acaba sendo um problema, já que são perdidas informações fonéticas no TextGrid gerado pela ferramenta. No caso da fala espontânea, a utilização da ferramenta é muito limitada, pois são obtidos vários erros no TextGrid, especialmente nas camadas de anotação fonética e silábica. Por fim, a ferramenta EasyAlign não indica as posições de fronteira prosódica.

As primeiras ferramentas destinadas à segmentação da fala em unidades maiores que a palavra costumavam fornecer uma "pontuação" de base sintática aos textos de fala, isto é, identificavam limites de sentenças sintáticas em textos falados (Cutler, Dahan e Donselaar, 1997). Esse procedimento identifica as fronteiras por meio de critérios estritamente sintáticos, o que pode ser visto como um problema muito sério nas segmentações obtidas, porque, frequentemente, não há uma relação biunívoca entre fronteiras sintáticas e prosódicas.

Com o desenvolvimento tecnológico de ferramentas para análise e extração de dados acústicos, desenvolveram-se ferramentas mais adequadas para a detecção automática de fronteiras prosódicas. O sistema de detecção automática de fenômenos prosódicos descrito em Wightman e Ostendorf (1994) localiza fronteiras de constituintes prosódicos, anotadas de acordo com o índice de fronteiras concebido no sistema notacional ToBI. Para o desenvolvimento do sistema, foram usados dados, previamente anotados por segmentadores, extraídos dos *corpora* Boston University Radio News (Ostendorf et al., 1995) e "ambiguous sentences" (Price et al., 1991). A precisão geral do algoritmo, que funciona por meio da

extração e classificação estatística de parâmetros prosódicos, é de 55% para identificar a localização e o índice de fronteira em dados de fala radiofônica. Nos dados retirados do *corpus* de sentenças ambíguas, a precisão geral é de 67%.

Seguindo um caminho parecido ao de Wightman e Ostendorf (1994), o sistema de automatização proposto por Ross e Ostendorf (1996) detecta fronteiras de nível 4, identificadas de acordo com o índice de ruptura proposto pelo sistema notacional ToBI. As ferramentas desenvolvidas por Wightman e Ostendorf (1994) e Ross e Ostendorf (1996) se diferenciam em termos de abrangência de fronteiras identificadas pelas ferramentas. Enquanto Wightman e Ostendorf (1994) desenvolveram modelos para identificar a localização e o tipo de fronteira, Ross e Ostendorf (1996) desenvolveram um modelo para identificar exclusivamente a localização da fronteira anotada com o índice 4. Os resultados obtidos neste trabalho indicam que a ferramenta identifica fronteiras de índice 4 com a acurácia geral de 66,9%.

A ferramenta de detecção automática de fronteiras proposta por Ni et al. (2012) detecta com alta acurácia fronteiras, também anotadas de acordo com o sistema ToBI. Porém, ao construir a ferramenta, os diferentes índices de fronteira previstos no ToBI não são analisados. Deste modo, os índices de fronteira são tratados de modo geral, sem diferenciar os índices 2, 3 e 4. O modelo utiliza simultaneamente dois métodos de classificação distintos, variáveis de previsão acústicas, lexicais e sintáticas para prever a realização das fronteiras em inglês e mandarim. Os dados foram extraídos dos *corpora* Boston University Radio Speech (Ostendorf et al., 1995) e Annotated Speech Corpus of Chinese Discourse (Li et al., 2000). Originalmente, o *corpus* Boston University Radio Speech possui transcrições ortográficas, alinhamento fonético, etiquetagem em classes de palavras e anotação prosódica, usando o sistema ToBI para inglês norte americano (Silverman et al., 1992). O Annotated Speech Corpus of Chinese Discourse foi anotado usando o C-ToBI- Chinese Prosodic Labeling System (Li et al., 2000). Para os fins deste trabalho, são expostos apenas os resultados obtidos pelo algoritmo utilizando as variáveis de previsão fonético-acústicas. Em mandarim, com 30 medidas fonético-acústicas, o modelo alcança 85% de precisão para detectar fronteiras anotadas conforme as diretrizes do ToBI. Em inglês, com 24 medidas, o modelo alcança 82% de precisão.

Em um trabalho que descreve uma ferramenta computacional destinada à detecção de fronteiras prosódicas em inglês (Ananthakrishnan e Narayanan, 2005), são apresentados resultados expressivos a respeito da segmentação automática da fala em unidades. A ferramenta proposta foi desenvolvida com dados de fala retirados do Boston University Radio Speech Corpus (Ostendorf et al., 1995), também chamado de *corpus* de BURNC. Os dados provenientes do *corpus* BURNC estavam anotados de acordo com os índices de fronteira

previstos no sistema de notação ToBI. Porém, os autores que desenvolveram a ferramenta optaram por não avaliar os diferentes índices, considerando apenas a presença de fronteira. Assim como o modelo proposto por Ni et al. (2012), a ferramenta utiliza variáveis mistas de previsão para criação de modelos de detecção de fronteiras prosódicas. São consideradas variáveis de natureza acústica, como medidas de intensidade normalizada, *pitch* e duração normalizada, mas também são consideradas variáveis de natureza sintática, como etiquetas que indicam classes de palavras. Os resultados encontrados por meio de modelos *Hidden Markov* (HMM) mostram que a ferramenta detecta 86% das fronteiras prosódicas marcadas no conjunto de dados analisados, utilizando exclusivamente variáveis acústicas. Utilizando variáveis acústicas em conjunto com variáveis sintáticas, a ferramenta identifica 88% das fronteiras prosódicas anotadas previamente.

Posteriormente, Ananthakrishnan e Narayanan (2008) criaram mais uma ferramenta destinada à segmentação da fala. De modo geral, a ferramenta possui pressupostos teóricos e operacionais semelhantes aos pressupostos descritos em Ananthakrishnan e Narayanan (2005), isto é, não há distinção entre índices de fronteiras, são utilizados dados de fala retirados do *corpus* BURNC, são utilizadas variáveis preditoras acústicas e sintáticas. A principal diferença é dada pela inclusão de variáveis lexicais e pelo uso de análises estatísticas muito mais complexas. Os resultados apontam que o modelo baseado em variáveis acústicas prediz 89,99% das fronteiras prosódicas, o modelo que tem como base variáveis lexicais prediz 85,73% das fronteiras, o modelo que utiliza somente variáveis sintáticas prevê 87,99%. A combinação das variáveis gera modelos com altos índices de acerto: i) variáveis acústicas em conjunto com variáveis lexicais identificam 90,41% das fronteiras; ii) variáveis acústicas em conjunto com variáveis sintáticas predizem 91,61% das fronteiras; iii) variáveis acústicas em conjunto com variáveis lexicais e sintáticas preveem 91,38% das fronteiras. A figura 1 abaixo mostra um exemplo de segmentação obtida por meio do modelo (ii).

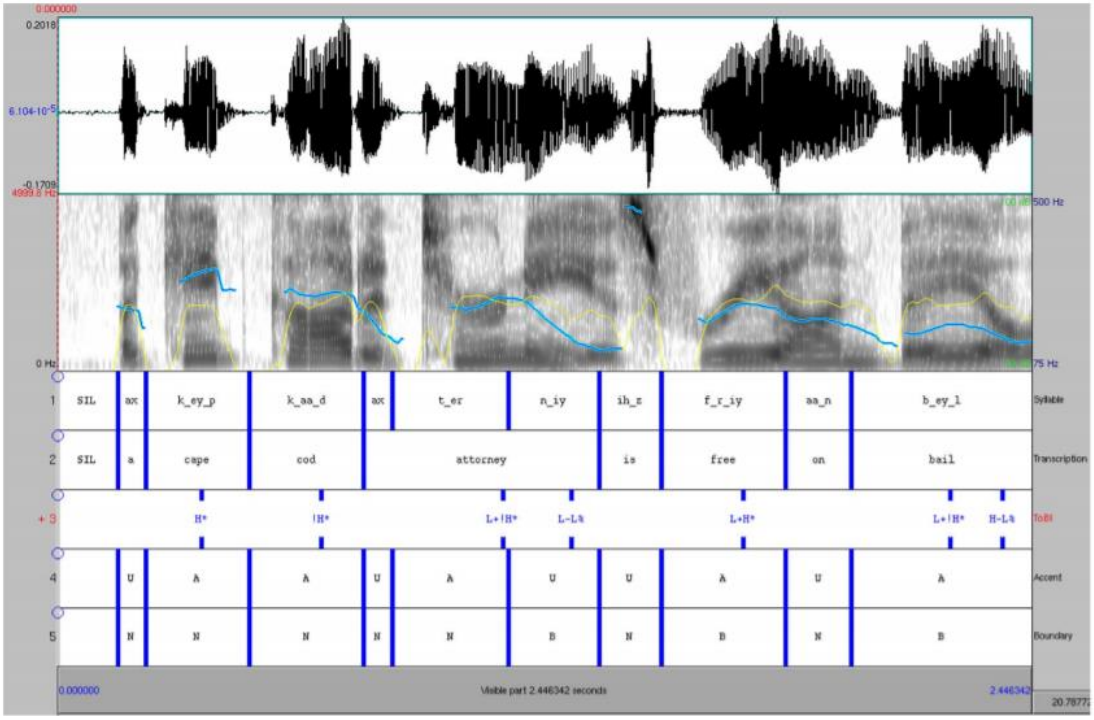


Figura 1: Output da ferramenta desenvolvida por Ananthakrishnan e Narayanan (2008)

A figura 1 mostra o output gerado pelo modelo ii ao segmentar um trecho de fala. De cima para baixo são visualizados a forma de onda, o espectrograma de banda larga e as camadas de anotação no Praat na ordem a seguir: 1) Segmentação e transcrição em nível de sílaba; 2) segmentação e transcrição em nível de palavra; 3) Marcação de tons de acordo com a anotação ToBI; 4) Eventos de acentuação atribuídos às sílabas segmentadas na camada 1 automaticamente através da ferramenta (U indica não acentuado; A indica acentuado); 5) Marcação automática das fronteiras atribuídas de acordo com a segmentação silábica realizada na camada 1 (N indica nenhuma fronteira; B indica fronteira).

O sistema desenvolvido por Ludusan e Dupoux (2014) detecta 98,3% das fronteiras de constituintes prosódicos anotadas previamente com os índices 3 e 4 no *corpus* Boston University Radio Speech Corpus (BURNC - Ostendorf et al., 1995), porém, a ferramenta não informa o índice da fronteira. Em outras palavras, a ferramenta detecta apenas a presença da fronteira e não distingue o índice 3 do índice 4. O sistema proposto por Mertens e Simons (2013) foi originalmente construído para detectar três níveis de força de fronteira prosódica. Neste estudo, os diferentes níveis de força foram estabelecidos perceptualmente, os dados foram previamente anotados por um segmentador *expert*. Os primeiros resultados encontrados em Mertens e Simons (2013) indicam que a detecção de fronteiras anotadas com o nível mais baixo é difícil de ser efetivamente viabilizada, porque o desempenho do sistema para detectar estas

fronteiras foi muito baixo, comprometendo os resultados da detecção das demais fronteiras. Por isso, para dar continuidade ao estudo, os autores optaram por considerar as fronteiras de nível mais baixo como não fronteiras (posições sem nenhuma fronteira). Após essa medida, a *performance* geral do sistema foi de 41,5% e 81,6% para a detecção de fronteiras anotadas respectivamente com os níveis de força 2 e 3.

O algoritmo ANALOR (A Tool for Semi-Automatic Annotation of French Prosodic Structure; Avanzi, Lacheret-Dujour e Victorri, 2008) detecta fronteiras prosódicas em francês por meio de parâmetros prosódicos como pausas físicas de pelo menos 300 ms, alongamentos silábicos, variações de f0, *resets* de f0 e proeminências. O grande atrativo do algoritmo em questão é a possibilidade de utilizá-lo em dados de fala lida e fala espontânea. Os resultados apontam uma convergência de 83,8% entre a detecção manual de segmentadores e a detecção automática realizada pela ferramenta. O algoritmo requer, no entanto, uma segmentação silábica prévia e detecta apenas a presença de fronteira prosódica, não distinguindo diferentes tipos de fronteira prosódica.

A ferramenta Prosody Tagger (Domínguez et al., 2016) identifica fronteiras e proeminências prosódicas. São apresentadas apenas as informações e considerações relativas à identificação automática das fronteiras prosódicas devido ao tema de interesse deste trabalho. A ferramenta em discussão funciona a partir da extração automática de parâmetros fonético-acústicos dos áudios. Os parâmetros extraídos são relativos à frequência fundamental, duração e intensidade. Em Domínguez et al. (2016), dois segmentadores *experts* anotaram as posições das fronteiras prosódicas. A anotação das fronteiras prosódicas não foi feita com base na percepção, mas sim com base em critérios estritamente formais. As fronteiras marcadas pelos segmentadores correspondiam às posições nas quais pelo menos um dos seguintes parâmetros ocorriam: pausa física; aumento de f0; diminuição brusca de intensidade. As fronteiras anotadas não foram diferenciadas entre si. Embora o trabalho tenha apresentado claramente quais fronteiras foram anotadas pelos segmentadores, a duração mínima de pausa considerada pelos segmentadores para anotar uma dada fronteira prosódica não foi mencionada.

Os elementos avaliados pelo Prosody Tagger para reconhecer as fronteiras não compreendem os mesmos elementos considerados pelos segmentadores para marcar as fronteiras. A pausa física foi um dos elementos formais que os segmentadores observaram ao anotarem manualmente as fronteiras, porém, ela não foi diretamente considerada pela ferramenta de detecção automática, pois a identificação automática das fronteiras foi feita a partir de apenas parâmetros de f0, duração e intensidade. A acurácia final da ferramenta foi medida comparando as fronteiras marcadas pela ferramenta com as fronteiras marcadas por

pelo menos um dos dois segmentadores. A acurácia final da ferramenta é de 0,84 com relação às fronteiras marcadas por pelo menos um segmentador.

Outras ferramentas seguem a proposta do Prosody Tagger no sentido de que identificam automaticamente fronteiras prosódicas atribuídas por segmentadores a partir de critérios formais. Estas ferramentas foram discutidas inicialmente na seção 2.5 (vide quadros 1 e 2) e são reapresentadas no quadro 3. No caso do quadro 3, a única diferença é o acréscimo da acurácia das ferramentas.

Quadro 3: Desempenho de algumas ferramentas de identificação automática de fronteiras prosódicas

Autor / Ano	**Amostra / Tamanho da amostra**	**Variáveis analisadas**	**Língua**	**Combinação de parâmetros associados às fronteiras**	**Acurácia**
Dominguéz, Farrús e Wanner / 2016	Fala espontânea / 5 minutos	F-AC	Espanhol Inglês	F0 Intensidade Duração	0,84
Apopei e Paduraru / 2015	Fala lida e fala eliciada / 10 minutos	F-AC	Romeno	F0 Intensidade Duração Pausa física	0,90
Beke, Szaszák e Váradi / 2013	Fala espontânea / 35 minutos	F-AC	Húngaro	F0 Intensidade	0,78
Rosenberg / 2009	Fala espontânea / 67 minutos	F-AC	Inglês	F0 Intensidade Duração	0,93
Hirschberg e Nakatani / 1998	Fala lida e fala eliciada	F-AC	Inglês	F0 Intensidade	0,83
Klejch, Bell e Renals / 2016	Mídia televisa / 15 horas	SIN LEX F-AC	Inglês	F0 Intensidade Duração Sintaxe Pausa física	0,87
Margolis, Ostendorf e Livescu / 2010	Fala radiofônica e fala telefônica / 2 horas	SIN LEX F-AC	Inglês	F0 Intensidade Duração Sintaxe	0,88
Heldner e Megyesi / 2003	Entrevista (Fala radiofônica) / 25 minutos	SIN LEX F-AC	Sueco	Duração Pausa física Sintaxe	0,85

Até o momento, todas as ferramentas apresentadas na seção 2.7 se baseiam em métodos de aprendizagem supervisionada. Os métodos de aprendizagem supervisionada analisam dados previamente anotados por segmentadores e identificam as características que melhor distinguem as categorias anotadas previamente. O objetivo, neste caso, é obter uma segmentação automática o mais compatível possível com a segmentação realizada pelos segmentadores. Com métodos não supervisionados ou semissupervisionados, os algoritmos apresentam as características que melhor distinguem os dados, porém, as categorias são

atribuídas pelo próprio algoritmo. Desta maneira, o algoritmo identifica sozinho um determinado número de categorias e divide os dados da forma que ele considera mais apropriada. No caso das fronteiras, a segmentação obtida pode ou não coincidir com a segmentação feita pelos segmentadores.

Em Jeon e Liu (2009), a identificação de fronteira é realizada através de um método de aprendizagem semissupervisionada. São utilizados dados provenientes do *corpus* BURNC. No trabalho em discussão, as posições são avaliadas por meio de algoritmos diferentes com visualizações semelhantes. Foram usados os algoritmos *Neural Network* (NN) e *Support Vector Machine* (SVM). O algoritmo NN foi utilizado para analisar apenas informações fonético-acústicas. Já o algoritmo SVM teve como foco a análise de informações sintáticas. No trabalho em discussão, cada algoritmo apresentou a sua versão de segmentação. Em seguida, as segmentações obtidas por meio dos algoritmos são comparadas. Caso os algoritmos concordem quanto ao estabelecimento de fronteira prosódica na mesma posição, a posição é, então, considerada uma boa candidata à anotação humana. Em Jeon e Liu (2009), a etiqueta final fruto do acordo entre os dois algoritmos foi comparada com a anotação manual do *corpus*. Para a identificação das fronteiras de índice 3 e 4 do sistema notacional ToBI, a acurácia da ferramenta com relação à anotação manual é de 86,7%.

Ferramentas de identificação automática de fronteiras prosódicas desenvolvidas a partir de métodos semissupervisionados ou não supervisionados não são muito comuns na literatura da área. Embora a segmentação perceptual não seja considerada durante o processo de desenvolvimento da ferramenta, geralmente, a segmentação atribuída autonomamente pelo algoritmo é comparada com a segmentação perceptual atribuída por segmentadores treinados ou não. Exemplos de ferramentas desenvolvidas a partir de métodos semissupervisionados que comparam a segmentação automática com a segmentação perceptual realizada por segmentadores são Chen et al. (2004) e Sridhar et al. (2008). Chen et al. (2004) usaram um modelo gaussiano para análise de informações fonético-acústicas e um modelo em rede neural para análise de informações sintáticas. A acurácia final da ferramenta de Chen et al. (2004) é de 90%, se comparada à segmentação perceptual dos dados. O trabalho de Sridhar et al. (2008) utiliza um modelo de entropia máxima para identificar as fronteiras. A precisão da ferramenta de Sridhar et al. (2008) é de 93%.

Atualmente, na literatura da área, estão disponíveis uma série de ferramentas de detecção automática de fronteira prosódica. De modo geral, as ferramentas oferecem uma boa contribuição a respeito da tarefa de identificação automática de fronteiras prosódicas. Ainda assim, é evidente que algumas questões devem observadas nas ferramentas.

A grande maioria das ferramentas citadas tem como foco a identificação de fronteiras prosódicas classificadas de acordo com os índices adotados no sistema notacional ToBI. No caso das ferramentas destinadas à identificação das fronteiras classificadas de acordo com o ToBI, é importante destacar que as ferramentas não costumam distinguir entre tipos de fronteiras prosódicas, isto é, elas indicam apenas a presença de fronteira prosódica, embora os dados utilizados no desenvolvimento das ferramentas sejam provenientes de um sistema de anotação com diferentes níveis de fronteiras prosódicas. Quando há alguma distinção, o desempenho das ferramentas não é alto.

Os métodos usados para o desenvolvimento das ferramentas de identificação automática de fronteira se dividem entre métodos supervisionados, semissupervisionados ou não supervisionados. Ferramentas desenvolvidas a partir de métodos supervisionados têm como grande vantagem refletir de alguma maneira a percepção de fronteira prosódica, caso as categorias observadas no processo de desenvolvimento da ferramenta tenham sido atribuídas perceptualmente. A desvantagem é a necessidade de ser realizado um trabalho extenso de preparação dos dados e desenvolvimento (treinamento) da ferramenta.

As ferramentas desenvolvidas a partir de métodos semissupervisionados ou não supervisionados são elaboradas mais rapidamente, pois não é necessária a segmentação dos segmentadores durante o processo de desenvolvimento da ferramenta. Geralmente, nestes casos, a segmentação automática é totalmente guiada pelas decisões tomadas exclusivamente pelo(s) algoritmo(s). A segmentação feita pelos segmentadores é importante apenas para avaliar em uma pequena quantidade de dados o quanto a segmentação automática se aproxima da segmentação dos segmentadores. Evidentemente, embora a segmentação automática obtida possa coincidir com a segmentação perceptual dos segmentadores, a percepção humana não é diretamente levada em consideração no processo de desenvolvimento da ferramenta.

O tamanho e o conjunto de dados usados no desenvolvimento das ferramentas também devem ser analisados. Em alguns trabalhos, a quantidade de dados é pequena. Frequentemente, os dados são repetitivos ou focados em uma tarefa específica. Por exemplo, o BURNC contém um número limitado de histórias repetidas por vários falantes. O Boston Directions Corpus (BDC) e Columbia Games Corpus (CGC) contêm respectivamente tarefas breves de orientação e comunicações provenientes de jogos específicos. A fala radiofônica é bastante utilizada. Porém, muitas vezes, as convenções sintáticas da escrita prevalecem e a prosódia se difere muito da prosódia da fala espontânea. Em alguns trabalhos, são utilizados dados utilizados dados de fala lida. Tanto para a fala radiofônica, quanto para a fala lida, as ferramentas podem

aprender características exclusivas daqueles dados e o poder de generalização das ferramentas pode ser muito comprometido.

Algumas ferramentas utilizam variáveis lexicais e sintáticas em conjunto com variáveis fonético-acústicas para identificar as fronteiras. Outros trabalhos, no entanto, realizam a identificação das fronteiras prosódicas com base apenas nos parâmetros fonético-acústicos. Os trabalhos de Ananthakrishnan e Narayanan (2005 e 2008) utilizaram variáveis de natureza sintática e/ou lexical para localizar as fronteiras prosódicas. Apesar dos modelos propostos em Ananthakrishnan e Narayanan (2005 e 2008) serem direcionados à língua inglesa e identificarem fronteiras tais como elas são concebidas no modelo de anotação ToBI, os resultados dos trabalhos em questão sugerem que ferramentas de identificação automática de fronteira prosódica podem ter um desempenho alto, usando apenas parâmetros fonético-acústicos extraídos do sinal sonoro, pois o uso de variáveis sintáticas e lexicais não representou um ganho muito expressivo no desempenho das ferramentas desenvolvidas pelos autores supracitados.

Atualmente, não está disponível uma ferramenta capaz de identificar automaticamente fronteiras prosódicas que veiculam a percepção de terminalidade ou continuidade em dados de fala espontânea em PB. Neste sentido, o trabalho aqui apresentado contribui na literatura sobre a segmentação automática da fala, pois auxilia no desenvolvimento de uma ferramenta com essa funcionalidade.

2.8. Ponto de partida

O fenômeno das fronteiras prosódicas é em si objeto de discussão. Nos parece que não podemos nos limitar a distinguir somente entre presença e ausência de fronteira ou distinguir diferentes níveis de saliência perceptual de fronteira, mas sim diferenciar as fronteiras qualitativamente (Moneglia e Cresti, 1997; Cresti, 2000; Pike, 1945; Schegloff, 1998; Szczepek Reed, 2004; Swerts, 1994; Campbell, 1993; Swerts et al., 1994; Degand e Simon, 2009). Como já foi dito, neste trabalho, as fronteiras prosódicas são abordadas do ponto de vista de uma perspectiva perceptual de descrição linguística. São avaliadas as fronteiras prosódicas que veiculam a percepção de conclusão ou continuidade. Portanto, adota-se a hipótese de que haja uma clara diferença perceptual entre fronteiras terminais e não-terminais. Todo esse conjunto de revisão de literatura embasa a premissa deste trabalho de que é possível segmentar o fluxo da fala em unidades entonacionais marcadas por fronteiras prosódicas terminais e não-terminais

codificadas ao longo do sinal sonoro por meio de configurações de parâmetros fonético-acústicos perceptivelmente relevantes para os ouvintes.

O presente trabalho é uma continuidade de um trabalho inicialmente desenvolvido em Teixeira (2018). O objetivo principal de Teixeira (2018) consistia em investigar os parâmetros fonético-acústicos associados à percepção de fronteiras prosódicas terminais e não-terminais. O trabalho em questão tinha como objetivo desenvolver modelos baseados em múltiplos parâmetros fonético-acústicos para detectar automaticamente fronteiras prosódicas terminais e não-terminais em dados de fala espontânea monológica masculina. O trabalho parte da hipótese da existência de dois grandes tipos de fronteiras: as terminais (TB), que delimitam unidades pragmática e prosodicamente autônomas, e as não terminais (NTB), que delimitam fronteiras internas à unidade terminada. A hipótese inicial do estudo consistia em desenvolver dois modelos distintos para detectar as fronteiras prosódicas, um dedicado à detecção fronteiras terminais, o outro dedicado às fronteiras não-terminais.

Em Teixeira (2018), foram utilizados sete trechos de fala monológica masculina compostos por aproximadamente 200 palavras. Os dados foram extraídos dos *corpora* de fala espontânea em português do Brasil (PB) C-ORAL-BRASIL I e II (Mello e Raso, 2012; Mello, Raso e Ferrari, em preparação). Quatorze segmentadores treinados, membros do Laboratório de Estudos Empíricos e Experimentais da Linguagem (LEEL) da Universidade Federal de Minas Gerais (UFMG), identificaram perceptualmente as fronteiras terminais e não-terminais nos trechos analisados. Os segmentadores foram treinados anteriormente pela equipe do laboratório e já apresentavam, mesmo que em diferentes graus, experiência em segmentação prosódica da fala. O acordo entre os segmentadores quanto à percepção das fronteiras foi avaliado por meio do coeficiente Kappa na versão Fleiss. Para os trechos de fala analisados no trabalho, o acordo geral obtido foi de 0,80 na anotação de fronteiras terminais percebidas e de 0,75 para fronteiras não-terminais, resultados considerados de grande acordo.

Os trechos foram anotados em cinco camadas de TextGrid do Praat (Boersma e Weenink, 2015) da seguinte forma: i) segmentação em unidades Vogal-Vogal (V-Vs); ii) Anotação das fronteiras não-terminais marcadas pelo grupo de segmentadores, informando o número de segmentadores que marcaram a fronteira; iii) Anotação das fronteiras terminais marcadas pelo grupo de segmentadores, informando o número de segmentadores que marcaram a fronteira; iv) Anotação de intervalos referentes às pausas físicas; v) Transcrição ortográfica. Para mais detalhes sobre o tratamento dos dados empregado, veja-se o capítulo 3.

O *script* para Praat *BreakDescriptor* (Barbosa, 2016) foi utilizado para extrair automaticamente 111 parâmetros fonético-acústicos em todas as unidades V-Vs em uma janela

centrada em toda fronteira de palavra fonológica. A janela inclui 10 unidades V-Vs à esquerda e 10 unidades V-Vs à direita de cada fronteira de palavra fonológica. Os parâmetros fonético-acústicos extraídos compreendem medidas de: i) taxa de elocução e articulação de fala (totalizando 6 medidas); ii) Duração normalizada e suavizada das unidades V-Vs (totalizando 34 medidas); iii) Frequência fundamental (totalizando 65 medidas); iv) Intensidade relativa (totalizando 4 medidas); v) Pausa física verificada no sinal sonoro (totalizando duas medidas). Estabeleceu-se nas definições de operação do *script* que as fronteiras prosódicas correspondem às posições nas quais pelo menos 50% dos anotadores indicaram uma fronteira do mesmo tipo. Isto é, o *script* considerou como fronteira as posições em que pelo menos 50% dos segmentadores perceberam uma fronteira da mesma natureza, isto é, fronteiras marcadas por pelo menos sete segmentadores. As demais posições são consideradas como posições sem nenhuma fronteira. Para mais detalhes sobre o BreakDescriptor, veja-se o capítulo 3.

Para gerar modelos compostos por múltiplos parâmetros fonético-acústicos que poderiam funcionar como preditores da realização das fronteiras, as fronteiras prosódicas percebidas por pelo menos 50% dos segmentadores e os parâmetros extraídos dos dados por meio do *script* foram submetidos aos algoritmos de classificação estatística *Random Forest* (RF) e *Linear Discriminant Analysis* (LDA) com o intuito de avaliar qual dos algoritmos apresentaria um melhor desempenho na tarefa de identificação automáticas das fronteiras em discussão. Os resultados mostram que ambos apresentam sucesso relativo na predição da realização das fronteiras prosódicas terminais e não-terminais. No entanto, o algoritmo LDA apresentou um melhor desempenho geral. Por isso, somente os modelos obtidos por meio do algoritmo LDA foram refinados e treinados a partir de diferentes estratégias.

Os modelos foram desenvolvidos através de um longo processo de treinamento heurístico composto por várias fases, avaliando o efeito da inclusão ou exclusão das diversas medidas extraídas pelo BreakDescriptor nas posições de fronteira e nas posições próximas à fronteira, mas também avaliando medidas extraídas nas posições sem nenhuma fronteira e nas posições próximas às posições sem nenhuma fronteira, porque ambas são essenciais para a classificação estatística. Foram desenvolvidos diferentes modelos designados para identificar automaticamente as fronteiras prosódicas. É importante esclarecer que os resultados expostos a seguir não foram obtidos por meio de uma ferramenta efetivamente implementada, que já realiza a detecção automática das fronteiras. Tais resultados expõem, de forma sucinta, modelos que podem ser usados como base para construir uma ferramenta com essa finalidade.

Para o caso das fronteiras terminais, foi analisada a oposição presença de fronteira terminal *versus* ausência de fronteira terminal com a categoria relativa à ausência englobando

posições de fronteira não-terminal e posições sem nenhuma fronteira. Para o caso das fronteiras não-terminais, inicialmente, foi analisada a oposição presença de fronteira não-terminal *versus* ausência de fronteira não-terminal com a categoria para tratar da ausência englobando posições de fronteira terminal e também posições sem nenhuma fronteira. Posteriormente, categoria ausência de fronteira não-terminal foi abandonada e passou-se a analisar a oposição fronteira não-terminal *versus* posição sem nenhuma fronteira. As justificativas para essa mudança na metodologia de desenvolvimento são expostas ao longo dos próximos parágrafos.

Foram selecionadas aleatoriamente 70% das posições para o desenvolvimento dos modelos. As demais posições (30% dos dados) foram usadas em uma avaliação sobre o poder de generalização dos modelos desenvolvidos com 70% das posições. A avaliação dos modelos trata-se de uma fase de validação e é muito importante para analisar o quanto o modelo aprendeu sobre o fenômeno de interesse. O modelo TB-nb foi desenvolvido a partir de 51 fronteiras terminais e 861 posições de ausência de fronteira terminal (70% dos dados). Na validação, foram usadas 22 fronteiras terminais e 369 posições de ausência de fronteira terminal que correspondem a 30% dos dados.

O modelo destinado à identificação automática das fronteiras terminais apresentou uma convergência de 80% em relação às fronteiras terminais identificadas pelos segmentadores durante a fase de desenvolvimento do modelo a partir de 70% dos dados e também na validação do modelo com 30% dos dados. Ao analisar NTB, TB e posições sem nenhuma fronteira, o modelo designado para identificar fronteiras não-terminais, chamado de NTB-1, apresentou uma convergência de 37,2% em relação às fronteiras não-terminais identificadas pelos segmentadores na fase de desenvolvimento e 47,7% na validação do modelo com 30% dos dados.

De fato, o total de fronteiras terminais devidamente reconhecidas pelo modelo destinado às fronteiras terminais é bem mais alto do que a quantidade total de fronteiras não-terminais identificadas pelo modelo destinado à identificação das fronteiras não-terminais. O modelo de identificação de fronteira não-terminal reconheceu em média 42,5% das fronteiras nas fases de desenvolvimento e validação. A composição detalhada dos modelos será abordada posteriormente ao término desta seção. A primeira observação que precisa ser feita sobre os modelos diz respeito aos parâmetros de pausa física. Nos dois modelos desenvolvidos, verificou-se que o papel relativo dos parâmetros de pausa física é altíssimo em ambos os casos. Também foi verificado que os parâmetros de pausa física parecem ser o principal motivo de obtenção de falsos alarmes. O modelo de identificação de fronteira terminal tende a marcar todas as posições de pausa física como fronteira terminal, mesmo quando a fronteira associada

à pausa física é não-terminal. O mesmo problema ocorre no modelo destinado às fronteiras não-terminais. Neste caso, o modelo tende a marcar as posições de pausa física como fronteira não-terminal, mesmo quando a fronteira associada à pausa física veicula a percepção de terminalidade para os segmentadores.

A pausa física é, obviamente, um parâmetro importante para a identificação perceptual e automática das fronteiras prosódicas, mas também parece ser responsável pela confusão entre os dois tipos fronteiras em discussão, explicando especialmente o baixo desempenho do modelo de identificação de fronteiras não-terminais. Por isso, a hipótese de partida foi parcialmente reformulada. A hipótese reformulada definiu que uma maior quantidade de modelos designados para identificar as fronteiras não-terminais seria desenvolvida, a hipótese inicial de que as fronteiras terminais poderiam ser detectadas através de um único modelo foi mantida. Cada modelo voltado para as fronteiras não-terminais teria como função detectar um subtipo de fronteira não-terminal, marcado por configurações de parâmetros diferentes entre si. Antes de efetivamente desenvolver uma maior quantidade de modelos NTB, o modelo NTB-1, desenvolvido a partir de posições NTB, TB e posições sem nenhuma fronteira, foi aplicado apenas em posições de NTB em conjunto com posições sem nenhuma fronteira. Os resultados mostram que exatamente o mesmo modelo NTB-1 identifica 68% das fronteiras não-terminais. Isso pode ser visto como evidências de que o modelo tem dificuldade para distinguir fronteiras não-terminais de terminais, pois a porcentagem de reconhecimento de NTB aumenta de uma média de 42,5% para 68% ao excluir as fronteiras terminais da análise.

O procedimento realizado para desenvolver novos modelos de identificação de NTB consistiu em eliminar as fronteiras reconhecidas pelo modelo NTB-1, desenvolver um segundo modelo (NTB-2) para identificar as fronteiras não-terminais não identificadas pelo modelo NTB-1, eliminar as fronteiras não-terminais devidamente identificadas pelo modelo NTB-2 e assim sucessivamente, quantas vezes fosse necessário realizar esse processo. Outra mudança diz respeito à base de dados usada no desenvolvimento do modelo. Para buscar minimizar a dificuldade de distinção entre fronteiras terminais e não-terminais com pausa física, foi estabelecido que as fronteiras terminais seriam completamente eliminadas do processo de desenvolvimento dos modelos destinados à identificação de NTB. Nesta fase de desenvolvimento de modelos voltados para as fronteiras não-terminais, 160 posições sem fronteira foram selecionadas aleatoriamente. Desta forma, os modelos passaram a ser treinados a partir da oposição fronteira não-terminal *versus* posição sem nenhuma fronteira prosódica, isto é, eliminando a categoria ausência de fronteira não-terminal. Foi estabelecido que todas as fronteiras não-terminais seriam utilizadas nesta fase de desenvolvimento dos modelos, o que

impossibilita a realização de validações em dados diferentes daqueles usados na fase de desenvolvimento.

As alterações na hipótese inicial e na base de dados usada no desenvolvimento dos modelos destinados ao reconhecimento de fronteiras não-terminais foram feitas com o intuito de aplicar os modelos de identificação de fronteiras prosódicas de forma hierárquica. Basicamente, a ideia de aplicação hierárquica consistia em aplicar o modelo de fronteira terminal, reconhecer as fronteiras terminais, excluir as fronteiras terminais devidamente identificadas, e, em seguida, identificar as fronteiras não-terminais analisando apenas as posições restantes, a começar pelo modelo NTB-1. Porém, ainda assim, eventualmente, fronteiras não-terminais acompanhadas por pausa física seriam indevidamente etiquetadas como terminal, porque, mesmo em menor proporção, o modelo de identificação de fronteira terminal também tende a supervalorizar os parâmetros relacionados à pausa física.

Ao término do processo, foram obtidos três modelos de identificação de fronteira não-terminal. Juntos, os três modelos apresentam uma convergência de 98% em relação às fronteiras não-terminais identificadas perceptualmente pelos segmentadores. Primeiramente, é apresentado o modelo de identificação de fronteira terminal. Em seguida, são apresentados os três modelos que visam reconhecer as fronteiras não-terminais. Convencionalmente, o modelo de identificação de fronteira terminal desenvolvido em Teixeira (2018) é chamado de TB-nb. Os três modelos de identificação de fronteiras não-terminais são chamados de NTB-1, NTB-2 e NTB-3.

Abaixo, nas tabelas 2 e 3, são apresentadas a composição e a *performance* detalhada do modelo TB-nb durante o desenvolvimento do modelo com 70% dos dados e a validação com 30% dos dados remanescentes.

Tabela 2: Desempenho do modelo TB-nb

Fase	**TB ident.**	**TB c/ pausa fís. ident.**	**TB s/ pausa fís. ident.**	**FA**	**FA c/ pausa fís.**	**FA s/ pausa fís.**
Des	80% (40)	100% (40)	NA	5% (43)	100% (43)	NA
Val	80% (17)	100% (17)	NA	8% (29)	100% (29)	NA

Tabela 3: Composição do modelo TB-nb

Rank	**Sigla**	**Parâmetro fonético-acústico**	**Peso**
1º	psdur	Duração da pausa física após a fronteira	2,641
2º	psp	Presença de pausa física após a fronteira	1,948
3º	f0meddloc	Reset de f0	0,329
4º	df0medr1	Inclinação média da f0 na 1ª unidade da janela à direita da fronteira	0,264
5º	df0medl	Inclinação média de f0 dentro da janela à esquerda da fronteira	0,257

6º	sddf0d	Diferença de variação das inclinações de f0 nas janelas dos dois lados da fronteira	0,157
7º	prd	Diferença de taxa de saliência duracional dos dois lados da fronteira	0,101
8º	sdf0l	Variação nas medianas da f0 na janela à esquerda da fronteira	0,091
9º	df0medl10	Inclinação média da F0 na 10ª unidade da janela à esquerda da fronteira	0,066
10º	f0rl	Taxa de produção de proeminência melódicas, associadas ao acento de *pitch* na janela à esquerda da fronteira	0,033
11º	df0meddloc	Diferença de inclinação média da f0 entre as unidades pós-fronteiriça e fronteiriça	0,032
12º	f0medd	Diferença entre medianas da F0 nas janelas dos dois lados da fronteira	0,029
13º	zl10	Duração normalizada média na unidade pré-fronteiriça	0,028
14º	skf0d	Diferença das caudas da distribuição de f0 dos dois lados da fronteira	0,025
15º	mzd	Diferença de duração normalizada média dos dois lados da fronteira	0,015
16º	skdf0d	Diferença de variação de cauda da distribuição de f0 nas janelas dos dois lados da fronteira	0,011
17º	SDzl	Variação da duração normalizada média na janela à esquerda da fronteira	0,010
18º	ard	Diferença de taxa de articulação dos dois lados da fronteira	0,003
19º	zdloc	Diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça	0,001
20º	emphl	Intensidade relativa (e volume, *loudness*) na janela à esquerda da fronteira	0,001

Na tabela 4, é apresentado o desempenho do modelo NTB-1, quando aplicado em todas as posições de fronteiras não-terminais e nas 160 posições sem nenhuma fronteira. Dentre as 244 fronteiras não-terminais, o modelo NTB-1 identificou 165 fronteiras não-terminais (68% de reconhecimento sob o total de fronteiras não-terminais). Com as 79 fronteiras não-terminais não reconhecidas pelo modelo NTB-1, desenvolveu-se o modelo NTB-2. Dentre as 79 fronteiras, o modelo NTB-2 identificou corretamente 61 fronteiras prosódicas, o que equivale a 25% de reconhecimento sob o total de fronteiras não-terminais. Com as fronteiras remanescentes não reconhecidas pelos modelos NTB-1 e NTB-2, desenvolveu-se o modelo NTB-3 que identificou 12 fronteiras não-terminais, o que equivale a 5% de reconhecimento sob o total de fronteiras não-terminais. Na tabela 4, também são apresentadas as informações relativas aos falsos alarmes gerados pelos modelos NTB-1, NTB-2 e NTB-3.

Tabela 4: Desempenho dos modelos NTB-1, NTB-2 e NTB-3

Modelo	**NTB ident.**	**NTB c/ pausa fís. ident.**	**NTB s/ pausa fís. ident.**	**FA**	**FA c/ pausa fís.**	**FA s/ pausa fís.**
NTB-1	68% (165)	62% (103)	38% (62)	22% (35)	NA	100% (35)
NTB-2	25% (61)	NA	100% (61)	20% (32)	NA	100% (32)
NTB-3	5% (12)	NA	100% (12)	12% (19)	NA	100% (19)

Na próxima página, a tabela 5 apresenta os parâmetros fonético-acústicos e pesos dos parâmetros que compõem os três modelos de identificação de fronteira não-terminal.

Tabela 5: Composição dos modelos NTB-1, NTB-2 e NTB-3

	Modelo NTB-1		**Modelo NTB-2**		**Modelo NTB-3**	
Rank	***Sigla*** ***Peso***	***Parâmetro fonético-acústico***	***Sigla*** ***Peso***	***Parâmetro fonético-acústico***	***Sigla*** ***Peso***	***Parâmetro fonético-acústico***
1º	zl0 4.5	Duração normalizada média na unidade fronteiriça	srl 0.72	Taxa de elocução dentro da janela à esquerda da fronteira	prl 151.6	Taxa de saliência duracional dentro da janela à esquerda da fronteira
2º	zr1 4.4	Duração normalizada média na unidade pós-fronteiriça	sddf0l 0.63	Variação das inclinações de f0 dentro da janela à esquerda da fronteira	prd 150.6	Diferença de taxa de saliência duracional dos dois lados da fronteira
3º	zdloc 4.2	Diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça	sdf0l 0.47	Variação nas medianas da f0 dentro da janela à esquerda da fronteira	prr 149.5	Taxa de saliência duracional dentro da janela à direita da fronteira
4º	psp 2.6	Presença de pausa física após a fronteira	ard 0.45	Diferença de taxa de articulação dos dois lados da fronteira	sdf0r 0.5	Variação nas medianas da f0 dentro da janela à direita da fronteira
5º	psdur 2.3	Duração da pausa física após a fronteira	f0medl 0.37	Mediana de f0 dentro da janela à esquerda da fronteira	SDzl 0.3	Variação na duração normalizada média dentro da janela à esquerda da fronteira
6º	ard 0.3	Diferença de taxa de articulação dos dois lados da fronteira	f0rd 0.21	Diferença de taxa de produção de proeminência melódicas dos dois lados da fronteira	df0medr1 0.3	Inclinação média de f0 na unidade pós-fronteiriça
7º	srd 0.3	Diferença de taxa de elocução dos dois lados da fronteira	f0meddloc 0.10	*Reset* de f0	df0medl10 0.2	Inclinação média de f0 na unidade pré-fronteiriça
8º	sdf0d 0.2	Diferença de variação das medianas de f0 dos dois lados da fronteira	f0med0 0.09	Mediana de f0 na unidade fronteiriça	df0meddloc 0.1	Diferença de inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça
9º	zl10 0.2	Duração normalizada média na unidade pré-fronteiriça	f0medr1 0.05	Mediana de f0 na unidade pós-fronteiriça	-	
10º	-		emphl 0.01	Intensidade relativa dentro da janela à esquerda da fronteira		

Em geral, os resultados sugerem que o uso de apenas dois modelos (um para TB; outro para NTB) é insuficiente para identificar automaticamente as fronteiras terminais e não-terminais. Desta forma, a detecção automática de fronteiras terminais parece ser mais simples. Os resultados mostram uma diferença clara entre os dois tipos de fronteiras quanto à configuração dos parâmetros fonético-acústicos. No modelo destinado às fronteiras terminais, os parâmetros principais do modelo, isto é, os parâmetros mais relevantes hierarquicamente compreendem parâmetros relacionados à pausa física e à mudança no contorno entonacional de f0. O modelo NTB-1, é o mais explicativo para a identificação das fronteiras não-terminais e observa principalmente parâmetros relacionados à duração normalizada e à pausa física. No modelo NTB-2, são utilizadas medidas de taxa de elocução, f0 e taxa de articulação. No modelo NTB-3, os principais parâmetros observados pelo modelo realizar suas predições são parâmetros de taxa de picos de saliência duracional das unidades V-Vs localizadas em torno das fronteiras.

Os quatro modelos desenvolvidos também mostram que os parâmetros mais relevantes hierarquicamente para identificar tanto as fronteiras terminais, quanto as fronteiras não-terminais, ocorrem principalmente na unidade imediatamente anterior à fronteira, na unidade fronteiriça e na unidade imediatamente posterior à fronteira. Assim, de modo geral, os fenômenos relevantes para a identificação das fronteiras prosódicas terminais e não-terminais tendem a ser localizados bem próximos da posição em que os segmentadores atribuem a percepção de fronteira. Este é um ponto importante, pois pode sugerir uma forte redução da extensão da janela de trabalho do *script* BreakDescriptor.

Durante o processo de desenvolvimento do modelo TB-nb, foi observado que parâmetros relacionados à duração não representam um ganho muito expressivo na *performance* do modelo de identificação de fronteiras terminais. Neste sentido, a adição de parâmetros duracionais no modelo de identificação de fronteira terminal não aumenta, mas também não diminui a porcentagem de fronteiras terminais devidamente identificadas. O cenário muda quando é analisada a importância dos parâmetros duracionais nos modelos de identificação de fronteira não-terminal. De fato, parâmetros relacionados à duração parecem ser fundamentais para a identificação de todas as fronteiras não-terminais, pois os modelos NTB-1, NTB-2 e NTB-3 incluem em sua composição parâmetros que representam fenômenos duracionais. A diferença basicamente reside no tipo de parâmetro duracional e no peso relativo dos parâmetros. Com as fronteiras não-terminais, geralmente, a eliminação de um ou alguns parâmetros duracionais afeta os modelos negativamente e a porcentagem de fronteiras identificadas reduz imediatamente.

Os parâmetros que tratam da frequência fundamental parecem ser mais importantes para identificar as fronteiras terminais, devido ao peso destes parâmetros. No modelo NTB-1, apenas um parâmetro relacionado à f0 é incluído com um peso bem baixo (8º parâmetro em um rank de 9 parâmetros). O quadro muda parcialmente se observamos o modelo NTB-2. Dentre os 10 parâmetros observados pelo modelo NTB-2, sete representam fenômenos relacionados à f0, porém, é importante ressaltar que este modelo identificou 25% das fronteiras percebidas pelos segmentadores. Dentre os oito parâmetros que compõem o modelo NTB-3, quatro são relacionados à frequência. Porém, o modelo identifica apenas 5% das fronteiras e a relevância hierárquica destes quatros parâmetros é baixíssima. Os parâmetros que representam fenômenos de intensidade são incluídos apenas nos modelos TB-nb e NTB-2 com um peso baixíssimo em ambos os modelos.

A segunda fase do estudo apresentado nesta seção foi realizada ao longo dos últimos quatro anos. Na segunda fase, os modelos de detecção automática de fronteiras (Teixeira, 2018) foram revalidados em novos dados. Em seguida, novos modelos para identificar automaticamente as fronteiras foram desenvolvidos a partir de estratégias e dados diferentes. Os novos modelos foram validados posteriormente em dados não utilizados originalmente no processo de desenvolvimento dos modelos. O desempenho de todos os modelos foi analisado minuciosamente e, em razão dos resultados obtidos, as fronteiras terminais e não-terminais foram refinadas internamente a partir de um teste de percepção auditiva de pausa.

O teste de percepção de pausa atendeu a dois propósitos. Por um lado, ele foi útil para aumentar o conhecimento acerca do fenômeno de percepção das pausas. Por outro, o teste foi importante para buscar reduzir os problemas apresentados pelos modelos de identificação das fronteiras terminais e não-terminais. A ferramenta de detecção automática das fronteiras não foi de fato implementada, mas o trabalho apresenta um direcionamento sobre o que precisa ser feito para implementar uma ferramenta computacional capaz de reproduzir, com a menor margem de erro possível, as decisões dos segmentadores quanto à percepção das fronteiras.

2.9. Resumo do capítulo

Neste capítulo, foi observado que os trabalhos da área podem ser divididos quanto aos diferentes objetos de análise. Entre os trabalhos com foco na análise das fronteiras, o pressuposto básico é identificá-las a partir da percepção auditiva, embora em alguns casos critérios formais sejam considerados no processo de identificação das fronteiras. Na literatura,

são propostas diferentes fronteiras prosódicas: i) fronteiras de diferentes níveis de força perceptiva; ii) fronteiras prosódicas dispostas em um *continuum* que varia gradualmente de percepção de ausência de fronteira à percepção de presença de fronteira; iii) fronteiras de diferentes valores comunicativos.

Há um bom conhecimento sobre os parâmetros fonético-acústicos envolvidos na percepção das fronteiras, mas a combinação e o peso relativo destes parâmetros não são bem compreendidos. A pausa física é uma condição suficiente para a percepção de uma fronteira, mas não é capaz de indicar sozinha o valor comunicativo da fronteira e algumas fronteiras não são acompanhadas pela pausa física. Atualmente, grande parte das ferramentas de detecção automática das fronteiras identifica fronteiras de diferentes níveis de força perceptiva e modelos treinados para identificar TB e NTB superestimam a pausa física como recurso de identificação das fronteiras, embora tais modelos não tenham sido implementados.

3. Panorama geral da metodologia empregada

As análises prosódicas das línguas podem ser feitas a partir de procedimentos metodológicos diferentes. Há análises baseadas em introspecção, experimentação, percepção ou abordagens construídas através de modelagens. O método por introspecção baseia-se na avaliação do pesquisador a respeito do objeto do estudo. Este tipo de método tem como grande problema ser intuitivo, não empírico e não testável por outros pesquisadores da área. O método por experimentação tem como característica analisar a forma (parâmetros fonético-acústicos) que se relaciona com as funções linguísticas estudadas. No entanto, muitas vezes os dados utilizados nesta abordagem são restritos a dados de fala que seguem o procedimento normalmente conhecido como *lab speech* que, embora possa ter um entendimento mais amplo do que simplesmente a leitura em laboratório (vide Xu, 2011), ainda assim têm algum grau de prejuízo quanto à validade ecológica. O método por percepção tem como característica analisar a sensação auditiva que é provocada pelo sinal sonoro nos ouvintes. O método por modelagem tem como característica a busca por modelos matemático-computacionais, capazes de descrever e/ou prever os fenômenos prosódicos. Contudo, podemos observar que os parâmetros utilizados nos trabalhos que utilizam métodos modelados matematicamente são muito variados entre si e correm o risco de não revelar como realmente a percepção humana de fronteira prosódica funciona, já que ela não costuma ser levada em consideração no momento da modelagem dos modelos.

Neste trabalho, a metodologia empregada pode ser considerada mista com o uso de análises perceptuais, experimentais e modelagens matemáticas mais transparentes, capazes de representar parâmetros fonético-acústicos relevantes para a produção e percepção de fronteiras terminais e não-terminais. As análises perceptuais foram feitas através da segmentação de vários segmentadores treinados para identificar perceptualmente as fronteiras terminais e não-terminais. Em um segundo momento, foram realizados o teste e a análise da percepção auditiva de pausa. Os dados utilizados são compostos por dados de fala espontânea decorrentes de situações comunicativas reais entre falantes. A análise fonético-acústica experimental avaliou os parâmetros fonético-acústicos sugeridos pela literatura tais como a pausa física, o *reset* de f0, a mudança na taxa de elocução e o alongamento pré-fronteiriço, mas também foram acrescentados outros parâmetros na análise com o objetivo de explorar aqueles que poderiam ser relevantes para a produção e percepção das fronteiras prosódicas. A modelagem matemático-computacional foi feita a partir de um Algoritmo de Aprendizagem de Máquina supervisionado. O algoritmo escolhido foi usado para desenvolver e validar modelos de identificação automática de fronteiras prosódicas percebidas pelos segmentadores.

A seguir, são apresentadas as informações mais gerais sobre a metodologia empregada. Tais informações compreendem os dados utilizados, o tratamento recebido pelos dados, o *script* usado para extrair dos áudios os parâmetros fonético-acústicos e a lista de quais parâmetros foram extraídos automaticamente pelo *script*. Todo o processo de desenvolvimento e validação dos modelos de identificação automática das fronteiras TB e NTB é apresentado à parte no capítulo 4.

3.1. Dados

Os dados estão disponíveis no seguinte link: https://drive.google.com/drive/folders/1DyBrgqM5ji-nZMkksrIFyyZvb5jjHHZc?usp=sharing e também podem ser acessados através do *QR Code* apresentado na figura 2.

Figura 2: *QR code* para acessar os áudios

Os dados compreendem trechos de fala monológica masculina com alta qualidade acústica. A fala monológica foi escolhida com o objetivo de evitar que as sobreposições de fala possam interferir nos parâmetros fonético-acústicos extraídos dos áudios. Em razão das diferenças de f0 conforme o sexo dos falantes, a opção de restringir os dados à fala feminina ou masculina é importante para garantir homogeneidade nos parâmetros de f0 extraídos. No caso deste trabalho, a fala masculina foi escolhida aleatoriamente. Os critérios de avaliação da qualidade acústica dos áudios foram: (i) resposta em frequência dos microfones; (ii) inspeção dos formantes; (iii) sobreposições de fala; (iv) ruído de fundo; (v) possibilidade de computação da curva de f0; (vi) para as gravações com menor qualidade acústica, nível de clareza da escuta. Estes critérios foram avaliados por um time de pesquisadores familiarizados com pesquisas prosódicas e fonéticas que classificaram os áudios em cinco faixas: A, AB, B, BC e C, sendo A classificada como qualidade ótima e C como a menor qualidade aceitável (pelo menos 60% da curva de f0 legível; Raso, 2012). A qualidade acústica de todos os áudios usados neste trabalho é A ou AB.

Os trechos são compostos por em média 192 palavras e são distribuídos em quatro amostras. Os dados foram extraídos tanto da parte informal, quanto da parte formal e de mídia do *corpus* C-ORAL-BRASIL I e II (Mello e Raso, 2012; Mello, Raso e Ferrari, em preparação). Os textos são, portanto, relativos a três *corpora* de fala espontânea: fala informal em contexto natural, fala formal em contexto natural e fala em mídia em contexto televisivo ou radiofônico. As quatro amostras são denominadas amostra I, amostra II, amostra total e amostra extra. Como o trabalho aqui desenvolvido é diretamente ligado ao trabalho desenvolvido em Teixeira (2018), definiu-se que a amostra coletada em Teixeira (2018) é chamada de amostra I. A amostra II é composta por uma segunda base de dados comparável com a amostra I. As amostras I e II

possuem sete trechos de fala cada. A amostra total compreende apenas os textos das amostras I e II juntos, totalizando, então, 14 trechos.

Portanto, a amostra total não engloba os trechos provenientes da amostra extra. A amostra extra é composta por 5 trechos. A amostra extra foi coletada no final do doutorado com o intuito de aumentar os dados para realizar o desenvolvimento e a validação de submodelos de identificação de fronteiras com e sem pausa percebida. Abaixo, a tabela 6 mostra a duração, a quantidade de palavras e uma breve descrição de todos os áudios.

Tabela 6: Descrição dos dados

Contexto	Amostra	Texto	Duração	Palavras	Descrição
Natural informal	I	bfammn11	01’11’’	189	Apresentação de memórias de infância
		bfammn24	00’58’’	151	Apresentação de aspirações profissionais de um jovem
	II	bpubmn12	01’26’’	198	Apresentação de atividades frequentemente realizadas em um setor de um tribunal de justiça
		bpubmn13	01’00’’	180	Apresentação da trajetória profissional de um artista plástico
	Extra	bfammn06	01’19”	207	Apresentação da trajetória profissional de um engenheiro elétrico
Mídia formal	I	bmidmasc01	01’23’’	212	Explicação sobre a capacidade de memória do ser humano (TV / Globo Ecologia)
		bmidmasc02	01’21’’	238	Explicação sobre o sistema nervoso do ser humano (TV / Globo Ecologia)
		bmidmasc03	01’07’’	183	Análise sobre a economia brasileira em anos de Copa do Mundo (TV / Jornal da Globo)
	II	bmedsp03_1a	01’02’’	206	Análise sobre a escalação de um time de futebol brasileiro (Rádio / Jogo Aberto)
		bmedsp03_1b	01’07’’	200	Explicação sobre o treinamento físico realizado por jogadores de um time de futebol brasileiro (Rádio / Jogo Aberto)
		bmedts10_1	01’11’’	180	Explicação sobre algumas características da música erudita dos séculos XVIII e XIX (TV / Globo Horizonte)
	Extra	bmedsc06	01’05”	172	Apresentação sobre o nível de proficiência em língua inglesa alcançado por brasileiros (TV / Conexão futura)
Natural formal	I	bnatmasc01	01’30’’	205	Explicação sobre a gestão científica, a produção acadêmica e o lugar da mimese nas universidades

	bnatmasc02	01’09’’	161	Reunião geral para apresentação de novas regras e processos de um hospital
II	bnatco03	01’00’’	202	Conferência sobre metodologia de pesquisa em Química
	bnatpr05	01’43’’	181	Discurso religioso sobre atitudes que devem ser evitadas
Extra	bnatco07	01’28”	212	Conferência sobre a governança de água no mundo
	bnatco08	01’23”	213	Conferência sobre a qualidade da ciência produzida no Brasil
	bnatte02	01’55”	207	Aula de Sintaxe no curso de graduação em Letras

De modo geral, os textos são muito diversificados e abrangem diferentes situações de comunicação entre falantes de PB. Os textos de mídia formal se dividem entre textos de mídia televisiva e radiofônica. Aproximadamente 28,5% da amostra total é composta por textos em contexto natural informal, 43% em contexto de mídia e 28,5% em contexto natural formal. Cerca de 20% da amostra extra é composta por textos em contexto natural informal, 20% em contexto de mídia e 60% em contexto natural formal. A origem dos falantes é apresentada abaixo na tabela 7.

Tabela 7: Origem dos falantes

Amostra	**Áudio**	**Origem do falante**
I	bfammn11	Rio de Janeiro / Rio de Janeiro
	bfammn24	Sete Lagoas / Minas Gerais
	bmidmasc01	Desconhecida
	bmidmasc02	São Paulo / São Paulo
	bmidmasc03	Desconhecida
	bnatmasc01	Rio Espera / Minas Gerais
	bnatmasc02	Rio Pomba / Minas Gerais
II	bmedsp03_1a	Diadema / São Paulo
	bmedsp03_1b	Desconhecida
	bmedts10_1	Desconhecida
	bnatco03	Rio de Janeiro / Rio de Janeiro
	bnatpr05	Belém / Pará
	bpubmn12	Belo Horizonte / Minas Gerais
	bpubmn13	Florianópolis / Santa Catarina
Extra	bfammn06	Rio de Janeiro / Rio de Janeiro
	bmedsc06	Rio de Janeiro / Rio de Janeiro
	bnatco07	Prudente de Moraes / Minas Gerais
	bnatco08	Belo Horizonte / Minas Gerais
	bnatte02	Salinas / Minas Gerais

A origem dos falantes é diversificada e abrange estados das regiões Sudeste, Sul e Norte do Brasil. A maioria dos falantes é de Minas Gerais, mas também há falantes provenientes do Rio de Janeiro, Pará, Santa Cantarina e São Paulo.

3.2. Tratamento dos dados

Os trechos das amostras I e II foram segmentados autonomamente por diferentes segmentadores, membros da equipe do Laboratório de Estudos Empíricos e Experimentais da Linguagem (LEEL) da Universidade Federal de Minas Gerais (UFMG). A amostra I foi segmentada por 14 segmentadores, a amostra II foi segmentada por 19 segmentadores. A amostra extra foi segmentada por apenas um segmentador devido à disponibilidade de tempo para realizar a percepção das fronteiras. Todos os segmentadores foram treinados anteriormente pela equipe do laboratório e já apresentavam, mesmo que em diferentes graus, experiência em segmentação prosódica da fala.

Entre os critérios de segmentação dos *corpora* C-ORAL-BRASIL, além das fronteiras terminais e não-terminais, estão previstas as marcações de duas outras fronteiras prosódicas denominadas retratações (*retractings*) e interrupções. Os *retractings* são unidades retratadas pelos falantes e são marcadas por fronteiras não-terminais. As interrupções compreendem unidades entonacionais em que o falante interrompe a unidade por problemas de planejamento, desempenho ou fatores externos, mas não retoma a unidade depois. Os exemplos 4 e 5 mostram casos de retratações e interrupções. Nos exemplos abaixo transcritos e segmentados de acordo com os critérios adotados nos *corpora* C-ORAL-BRASIL, a barra dupla é usada marcar a percepção de fronteira terminal (//), a barra simples marca a percepção de fronteira não-terminal (/), o símbolo ([/n]) marca a percepção de retratação (o número associado à barra entre colchetes indica o número de palavras retratadas pelo falante), o sinal de adição marca a percepção de interrupção (+), a sigla de três letras precedida por asterisco e seguida de dois pontos identifica o falante (*XXX:).

Exemplo 4 (bfammn01)
*MAI: aí quando ele [/2] quando ele lá vinha embora / ea começou a quebrar mato //

Exemplo 5 (bfamcv04)
*HEL: mas primeiro cê nũ fez assim / tipo + fez //

Na tarefa de percepção de fronteiras prosódicas, os segmentadores receberam os áudios e a sua correspondente transcrição (sem nenhuma anotação referente às fronteiras), os ouviram e marcaram na transcrição os símbolos designados para cada tipo de fronteira percebida. Os segmentadores ouviram os áudios quantas vezes julgassem necessário. A tarefa de percepção

de fronteiras nos áudios provenientes das amostras I, II e extra foi realizada no segundo semestre de 2016[9], 2018 e 2021 respectivamente.

Na amostra I, a tarefa dos segmentadores consistia em inserir na transcrição uma barra simples (/) para fronteira não-terminal ou uma barra dupla (//) para fronteira terminal nas posições em que perceberam as fronteiras. Abaixo, no quadro 4, são apresentadas as instruções fornecidas aos anotadores na tarefa de percepção de fronteiras da amostra I.

Quadro 4: Instruções fornecidas aos segmentadores na tarefa de percepção de fronteiras da amostra I

A tarefa consiste em segmentar (cada um independente do outro) trechos de fala espontânea da forma que fazemos habitualmente no laboratório. São trechos bem pequenos e a segmentação deve ser feita inserindo na transcrição:

- Uma barra simples (/) quando vocês percebem fronteira não-terminal;
- Uma barra dupla (//) quando vocês percebem fronteira terminal.

Naturalmente, vocês vão observar algum aspecto na transcrição que não concordam plenamente, mas peço para que não alterem nada na transcrição para que eu possa trabalhar com facilidade na base de dados depois.

Nas amostras II e extra, os segmentadores marcaram na transcrição as posições de percepção de fronteira terminal e não-terminal com os símbolos que representam as duas fronteiras em questão, mas também marcaram na transcrição as posições de percepção de retratações e interrupções com os símbolos correspondentes às disfluências. Para simplificar a tarefa realizada pelos segmentadores, solicitou-se que os segmentadores indicassem a percepção de retratações com uma barra simples entre colchetes, isto é, os segmentadores não precisavam indicar o número de palavras retratadas pelo falante ([/]). A anotação das retratações e interrupções percebidas foi solicitada na tarefa de percepção de fronteiras das amostras II e extra com o objetivo de assegurar que os parâmetros associados às fronteiras consideradas disfluências (fronteiras produzidas por razões não linguísticas) não fossem analisadas no processo de desenvolvimento dos modelos e submodelos de identificação de fronteiras.

No caso da amostra II, a razão principal pela qual a percepção de disfluências foi anotada pelos segmentadores é excluí-las da base de dados a partir de um critério que reflete a percepção da maioria dos segmentadores. No caso das amostras I e extra, as posições de disfluência também foram excluídas, porém, a exclusão foi baseada na percepção de apenas um segmentador. Em ambos os casos de exclusão, a ideia é impedir que as disfluências interfiram nos modelos e submodelos de identificação de fronteiras. Na tarefa de percepção de pausa, as

[9] Teixeira (2018).

disfluências foram reconsideradas (vide capítulo 5). Abaixo, no quadro 5, são apresentadas as instruções fornecidas aos anotadores na tarefa de percepção de fronteiras das amostra II e extra.

Quadro 5: Instruções fornecidas aos segmentadores na tarefa de percepção de fronteiras das amostras II e extra

A tarefa consiste em segmentar (cada um independente do outro) trechos de fala espontânea da forma que fazemos habitualmente no laboratório. São trechos bem pequenos e a segmentação deve ser feita inserindo na transcrição:

- Uma barra simples (/) quando vocês percebem fronteira não-terminal;
- Uma barra dupla (//) quando vocês percebem fronteira terminal;
- Uma barra simples entre colchetes ([/]) em casos de percepção de retratação;
- O sinal de adição (+) em casos de percepção de interrupção.

Naturalmente, vocês vão observar algum aspecto na transcrição que não concordam plenamente, mas peço para que não alterem nada na transcrição para que eu possa trabalhar com facilidade na base de dados depois.

A seguir, o exemplo 6 mostra a segmentação de um trecho da amostra II.

Exemplo 6 (bmedsp03_1b)

*WIL: ontem / foi folga / aqui no Corinthians / hoje dia de reapresentação / quer vim // vamo lá // com o Cabeça &zo [/] nosso cinegrafista / famoso Cabeça / aqui pa mostrar o Pato / &he / treinando / chutes a gol / nessa finalização do trabalho / nessa reapresentação do Corinthians / o Pato tá ali / finalizando o trabalho / quem veio pro campo hoje / foram aqueles jogadores / que não atuaram os noventa minutos / no jogo contra / o São Paulo / os que atuaram o jogo todo / participaram apenas / de um treinamento leve / de corridas em volta do campo / um trabalho lá dentro / aquele começo de semana / treino de / &he / treino regenerativo / né // então a gente tá vendo aí o Pato / porque é o centro das atenções / essa semana / aqui no Corinthians // na sexta feira / o Cristóvão Borges / na entrevista coletiva / já disse / que ele vai ser relacionado / pro jogo / contra o Figueirense // ele já poderia ser relacionado contra o São Paulo / não foi / pra ele ganhar mais uma semana / de treinamento / pra se recondicionar / melhor / e aí sim / vai pro jogo / contra o Figueirense // se vai ser titular / ou não / aí / essa semana / vai responder / essa questão / o Cristóvão Borges / portanto vai definir / como o Corinthians entra em campo / contra o Figueirense //

O acordo entre segmentadores foi avaliado por meio do coeficiente de Kappa (Fleiss, 1971). Com a amostra I, são consideradas apenas as categorias fronteira terminal, não-terminal e posição sem nenhuma fronteira, pois os segmentadores anotaram apenas a percepção de terminal e não-terminal. Com as amostras II e extra, os segmentadores anotaram a percepção de terminal, não-terminal, interrupção e *retracting*. Porém, optou-se por transformar as categorias retratação e interrupção em uma única categoria chamada "disfluência", porque as fronteiras de retratação e interrupção são consideradas fronteiras produzidas por razões não linguísticas. Naturalmente, no caso da amostra extra, não é possível avaliar nenhum tipo de acordo entre segmentadores, pois as fronteiras foram anotadas por um único segmentador.

Para a amostra I, o coeficiente Kappa geral relativo a todas as marcações é 0,80, já o coeficiente Kappa por tipo de marcação é exposto abaixo na tabela 8.

Tabela 8: Coeficiente Kappa das segmentações realizadas pelo grupo de 14 segmentadores

Categoria	Kappa
Fronteira terminal	0,80
Fronteira não-terminal	0,75
Posição sem nenhuma fronteira	0,84

Para amostra II, o coeficiente Kappa geral relativo a todas as marcações é 0,79, já o coeficiente Kappa por tipo de marcação é exposto abaixo na tabela 9.

Tabela 9: Coeficiente Kappa das segmentações realizadas pelo grupo de 19 segmentadores

Categoria	Kappa
Fronteira terminal	0,73
Fronteira não-terminal	0,72
Disfluência	0,78
Posição sem nenhuma fronteira	0,85

Os trechos da amostra I foram anotados em cinco camadas do objeto de anotação TextGrid do *software* para análises fonético-acústicas Praat (Boersma e Weenink, 2015). Foram adotadas as seguintes camadas:

1. Segmentação em unidades Vogal-Vogal (unidades V-Vs)[10] e etiquetagem[11] utilizando transcrição fonética larga em caracteres ASCII[12];
2. Anotação das fronteiras prosódicas não-terminais marcadas pelo grupo de segmentadores, informando o número de pessoas que marcaram a fronteira;
3. Anotação das fronteiras prosódicas terminais marcadas pelo grupo de segmentadores, informando o número de pessoas que marcaram a fronteira;
4. Anotação de intervalos referentes às pausas físicas;
5. Transcrição ortográfica do texto do enunciado.

As unidades V-Vs, anotadas na primeira camada de anotação do Praat, são delimitadas tomando o intervalo que vai do início de uma vogal ao início da vogal sucessiva, isto é, as unidades V-Vs compreendem o segmento acústico que vai de um *onset* vocálico até o *onset* seguinte, incluindo os segmentos consonantais e as pausas físicas presentes entre eles (Barbosa,

[10] A segmentação em unidades V-Vs foi feita automaticamente utilizando o *script* para praat BeatExtractor (Barbosa, 2004). Posteriormente, foi realizada uma revisão manual da segmentação gerada.
[11] A etiquetagem foi realizada manualmente por duas pessoas treinadas para realizar uma etiquetagem desse tipo.
[12] Para mais detalhes sobre a tabela de correspondência IPA versus ASCII, veja-se a documentação disponível em: https://github.com/pabarbosa/prosody-scripts/blob/master/ProsodyDescriptorExtractor/Documents/IPA-ASCIICorrespondanceTable_BP.pdf.

2006). A unidade V-V se diferencia da sílaba fonológica convencional, porque ela agrupa a rima de uma sílaba com o ataque da sílaba subsequente. Adota-se a segmentação em unidades V-Vs porque estas unidades permitem detectar os pontos de transição consoante-vogal (CV). A transição CV é muito importante para estudos sobre a estruturação rítmica de dados de fala, porque se observa uma maior precisão perceptiva do ouvido humano para perceber a transição entre consoante e vogal subsequente (CV) do que entre vogal e consoante seguinte (VC) (Marcus, 1981). A transição CV também é muito importante porque é o ponto de alinhamento entre o ritmo da fala e o ambiente externo, como mostram os experimentos com *perceptual-centers* (Marcus, 1981; Barbosa, 2006; Pompino-Marschall, 1989). A unidade V-V é útil, então, do ponto de vista da produção e percepção da fala, sendo especialmente utilizada para cálculos e avaliação da evolução da duração durante a fala.

As pausas físicas anotadas na quarta camada compreendem os intervalos de qualquer duração caracterizados pela interrupção da vocalização (fluxo de fala). Naturalmente, os períodos de silêncio referentes à oclusão das consoantes oclusivas não foram anotados na quarta camada do objeto de anotação TextGrid. Os textos das amostras II e extra foram anotados em seis camadas. Foram adotadas as cinco camadas de anotação utilizadas no tratamento de dados da amostra I e uma camada adicional para a anotação de disfluências marcadas pelo grupo de segmentadores da amostra II ou pelo segmentador da amostra extra. Para minimizar possíveis erros, as camadas de anotação referentes às amostras I, II e extra foram feitas por três diferentes segmentadores treinados, revisadas diversas vezes e, nos casos de maiores dúvidas, foi discutido o que seria mais adequado. Abaixo, a figura 3 apresenta as camadas de anotação relativas a algumas das unidades entonacionais marcadas pelos segmentadores no áudio bmedsp03_1a.

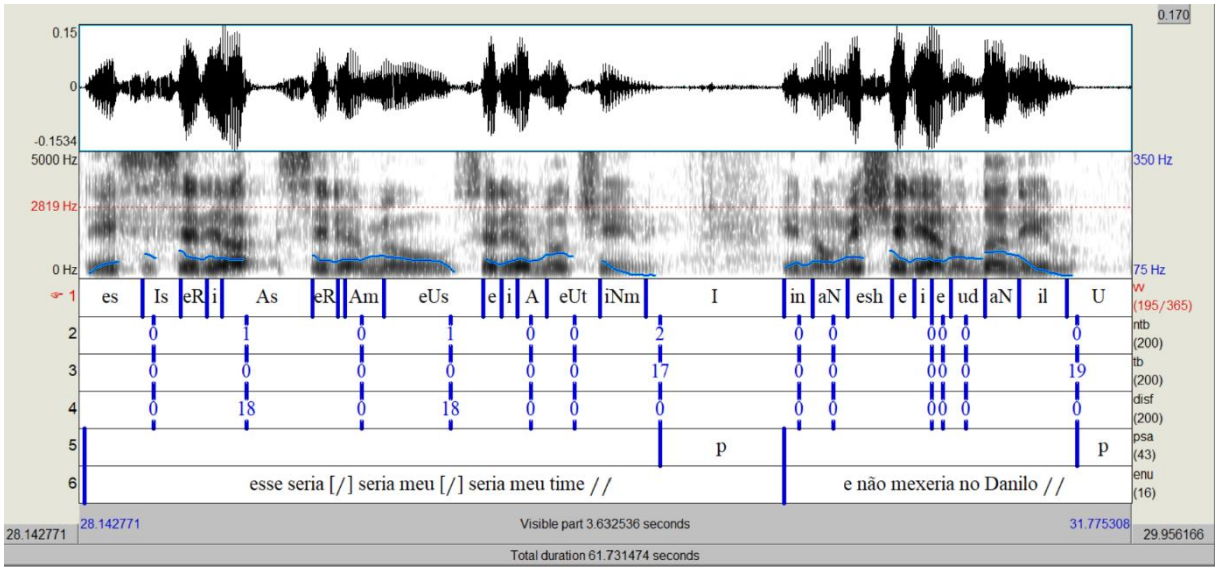


Figura 3: Camadas de anotação da tarefa de percepção de fronteira
Trecho: esse seria [/] seria meu [/] seria meu time // e não mexeria no Danilo //

Na figura 3, de cima para baixo, visualizamos a forma de onda, o espectrograma de banda larga e as camadas de anotação no Praat para um trecho de fala espontânea. Nas camadas 2, 3 e 4 são visualizadas as marcações atribuídas pelo grupo de segmentadores. Na camada 5, são visualizados os intervalos referentes às pausas físicas. Na última camada, o trecho é transcrito ortograficamente.

3.3. Princípios de funcionamento do *script* BreakDescriptor

A detecção automática de fronteiras prosódicas é basicamente uma tarefa de classificação, na qual cada fronteira de palavra fonológica é considerada uma potencial posição para se estabelecer uma fronteira prosódica, isto é, entre cada palavra fonológica pode ou não ocorrer uma fronteira prosódica de tipo terminal ou não-terminal. Com o intuito de analisar todas as potenciais posições de estabelecimento de fronteira prosódica, foi utilizado o *script* para Praat BreakDescriptor versão 2.0 (Barbosa, 2019) para extrair uma série de parâmetros fonético-acústicos em todas as unidades V-Vs em uma janela centrada em toda fronteira de palavra fonológica[13]. Isso inclui naturalmente as posições indicadas pelos segmentadores como posições de fronteiras prosódicas, mas inclui também as posições que os segmentadores

[13] Entende-se como palavra fonológica uma unidade prosódica identificada por recair um acento lexical primário.

consideraram posições sem nenhuma fronteira. O *script* extrai as medidas inclusive nas posições sem fronteira, porque o conjunto de parâmetros relativos às estas posições são fundamentais para a classificação estatística.

O funcionamento do BreakDescriptor é estabelecido a partir de dois critérios de referência. O primeiro critério consiste em definir a extensão da janela de análise do *script*; o segundo consiste em definir a porcentagem de acordo entre segmentadores a ser considerada para estabelecer quais são as posições de determinado tipo de fronteira, de acordo com a percepção dos segmentadores. O *script* permite que seja atribuída qualquer porcentagem de acordo e a janela de análise se move a cada fronteira de palavra fonológica, podendo ser composta por qualquer número natural entre 1 e 10.

Neste trabalho, no início, a janela de análise foi composta por 10 unidades anteriores e posteriores à posição analisada. Com esta opção de janela, a primeira posição para a qual o *script* extrai os parâmetros é a primeira posição em final de palavra fonológica anotada no objeto de anotação TextGrid a partir da 11º unidade V-V do áudio; a última posição em que os parâmetros são extraídos é a última posição em final de palavra fonológica na qual é possível contabilizar 10 unidades após ela. Neste sentido, não é feita a extração de parâmetros relativos a eventuais fronteiras localizadas nas 10 primeiras unidades V-V do áudio, mesmo que as fronteiras tenham sido percebidas pelos segmentadores e as informações referentes à percepção das fronteiras tenham sido anotadas no objeto de anotação TextGrid. Caso os segmentadores percebam fronteiras prosódicas em alguma das últimas 10 unidades V-V dos áudios, os parâmetros relativos às fronteiras em questão não são extraídos, porque não há 10 unidades após as fronteiras para o *script* analisar.

No total, com a janela de 10 unidades, para cada posição em final de palavra fonológica são extraídos os parâmetros em 21 unidades V-Vs (10 unidades anteriores, 10 posteriores e a própria unidade em análise). Caso o usuário do *script* opte pela redução das janelas para um valor menor do que 10, haverá uma maior quantidade de posições possíveis para se extrair os parâmetros, porque a redução da janela de análise permite a extração de parâmetros relativos a posições em final de palavra fonológica localizadas mais no início e fim do áudio.

As janelas de análise do BreakDescriptor foram reduzidas com o objetivo de extrair os parâmetros fonético-acústicos em posições mais próximas às posições em que os segmentadores marcaram a percepção de fronteira prosódica. Os resultados dos modelos desenvolvidos a partir de parâmetros extraídos da janela de 10 unidades indicaram que os parâmetros relevantes para a identificação automática das fronteiras se localizam nas próprias posições de atribuição de percepção de fronteira e em posições muito próximas às posições de

atribuição de percepção de fronteira, seja ela terminal ou não-terminal. As posições próximas compreendem as unidades V-V imediatamente anterior e posterior em relação à posição de atribuição da percepção de fronteira, isto é, as unidades pré e pós-fronteiriça. Apesar dos modelos desenvolvidos a partir de parâmetros extraídos das janelas de 10 unidades indicarem que não seria necessário analisar fenômenos que estão ocorrendo em posições muito distantes das posições de atribuição de percepção de fronteiras, para manter uma margem de segurança satisfatória, neste trabalho, as janelas de análise foram reduzidas para 7 e 3 unidades.

Este trabalho adota como referência o acordo de pelo menos 50% entre os segmentadores para estabelecer as posições de fronteira prosódica. Deste modo, para a amostra I que foi segmentada por 14 segmentadores, as posições de fronteira prosódica são aquelas em que pelo menos 7 segmentadores marcaram necessariamente uma fronteira do mesmo tipo. Na amostra I, as demais posições, também localizadas em limites de palavras fonológicas, são consideradas posições sem nenhuma fronteira prosódica. Um dos exemplos de empate é verificado quando 7 segmentadores marcam percepção de fronteira terminal e outros 7 marcam fronteira não-terminal em uma mesma posição. As situações de empate também podem ser verificadas quando 7 segmentadores marcaram a percepção de fronteira não-terminal ou terminal e outros 7 consideram que não há nenhuma fronteira em uma determinada posição.

Na amostra II, o acordo de pelo menos 50% entre os segmentadores para estabelecer as posições de fronteira prosódica é mantido. Como os áudios foram segmentados por 19 segmentadores, as posições de fronteira prosódica são aquelas em que pelo menos 10 segmentadores marcaram necessariamente uma fronteira do mesmo tipo. As demais posições são consideradas posições sem nenhuma fronteira prosódica. Na amostra extra, evidentemente, não é possível adotar o critério de acordo de pelo menos 50% para atribuir as posições de fronteira, pois somente um segmentador anotou as fronteiras percebidas. Desta forma, todas as fronteiras da amostra extra são aquelas percebidas pelo segmentador que realizou a tarefa de percepção.

Durante o processo de desenvolvimento e validação dos modelos de identificação automática de fronteira, as posições de empate no que diz respeito à natureza perceptual da fronteira foram consideradas ao mesmo tempo posições de fronteira terminal e não-terminal. Assim, as posições de empate quanto à natureza da fronteira foram usadas nas fases de desenvolvimento e validação dos modelos TB e NTB. Os empates obtidos quando 7 segmentadores marcaram percepção de fronteira não-terminal ou terminal e outros 7 segmentadores não perceberam nenhum tipo de fronteira prosódica são tratados apenas como terminal ou não-terminal. Na tarefa de percepção de pausa em posição de fronteira prosódica,

as posições de empate foram decididas por um segmentador *expert*. O segmentador *expert* foi escolhido com base no acordo entre a anotação de cada segmentador da amostra II e a anotação que é fruto do acordo de pelo menos 50% entre os segmentadores da amostra II. Mais detalhes sobre a escolha do segmentador *expert* são abordados posteriormente no capítulo 5.

Abaixo, a figura 4 ilustra o modo de operação do *script* a partir de uma porcentagem mínima de 50% de acordo entre segmentadores e janelas de análise compostas por 10 unidades.

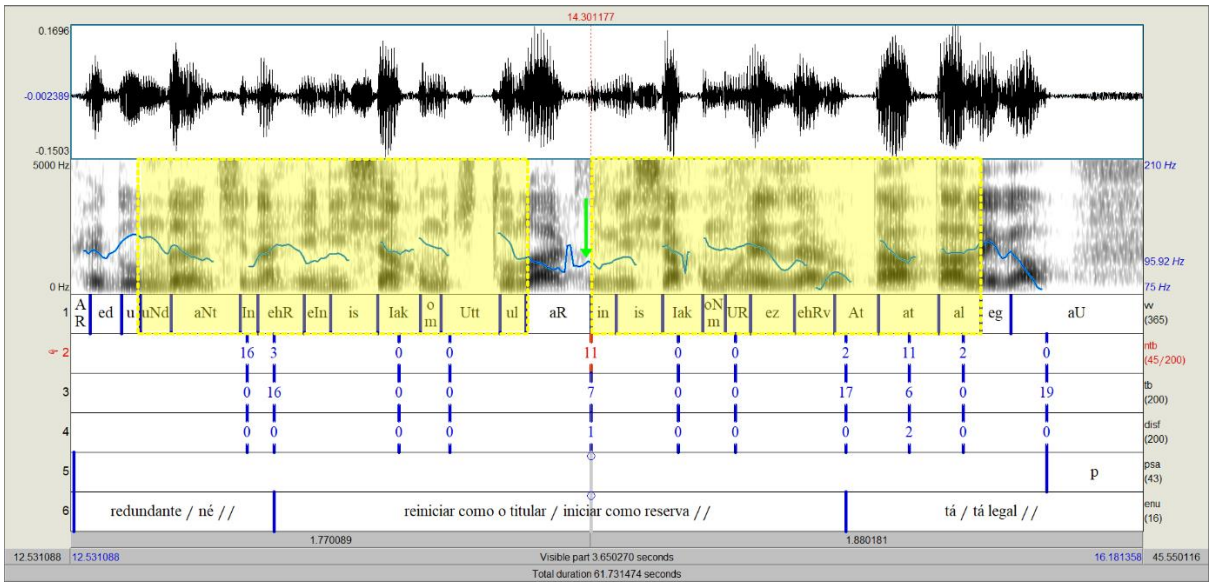


Figura 4: Funcionamento do BreakDescriptor 2.0
A fronteira destacada está inserida no enunciado: reiniciar como titular / iniciar como reserva //

Na figura 4, no centro, destaca-se uma unidade V-V em que 11 segmentadores marcaram como posição de NTB, 7 marcaram como TB e um marcou como disfluência. Devido ao critério de estabelecer as posições de fronteira a partir do acordo de pelo menos 50% entre segmentadores, a posição em questão é considerada pelo *script* uma posição de fronteira não-terminal, apesar de 7 segmentadores terem marcado fronteira terminal. Para a posição destacada com a seta verde, são calculados os parâmetros fonético-acústicos nas 10 unidades V-Vs anteriores (área sombreada do espectrograma), na unidade que marca a posição da fronteira e nas 10 unidades V-Vs posteriores à posição de análise (área sombreada do espectrograma).

Abaixo, a tabela 10 mostra as posições computadas a partir da janela de análise de 10 unidades e pelo menos 50% de acordo quanto à percepção das fronteiras. As disfluências definidas por um segmentador nas amostras I e extra e aquelas que foram percebidas por pelo menos 50% dos segmentadores na amostra II não são apresentadas, porque não foram

consideradas no processo de desenvolvimento e validação dos modelos de identificação automática de fronteira.

Tabela 10: Distribuição de posições contabilizadas a partir da janela de análise de 10 unidades

Posições	Total	Amostra	Frequência
Terminal	164	I	73
		II	46
		Extra	45
Não-terminal	726	I	244
		II	292
		Extra	190
Sem nenhuma fronteira	2172	I	986
		II	760
		Extra	426

Independentemente dos critérios de referência escolhidos (extensão da janela de análise e porcentagem de acordo entre segmentadores), as medidas extraídas automaticamente sempre compreendem medidas de natureza global e local. No que tange as medidas extraídas, a única diferença é a quantidade de medidas locais extraídas. Assim, sempre são extraídas 42 medidas globais. A quantidade de medidas locais extraídas é diversificada e proporcionalmente dependente da extensão de janela escolhida, de modo que quanto maior for a extensão da janela, maior é a quantidade de medidas locais extraídas.

Para a extensão de janela de 10 unidades V-Vs, que significa uma janela de 21 unidades (10 unidades V-Vs anteriores, 10 unidades V-Vs posteriores e a própria unidade V-V em análise), no total, são extraídas pelo *script* 111 medidas (42 globais e 69 locais) que permitem prever a possibilidade de fronteira prosódica para a unidade VV central. Para a menor possibilidade de extensão de janela, dada a forma como o *script* foi implementado (uma unidade V-V anterior, uma unidade V-V posterior e a própria unidade V-V em análise), no total, são extraídas 57 medidas (15 locais e 42 globais). As propriedades físicas relativas à produção dos sons extraídas automaticamente pelo *script* BreakDescriptor compreendem cinco grupos de parâmetros fonético-acústicos:

1. Medidas de taxa de elocução (*speech rate*) e ritmo (6 medidas);
2. Medidas de duração normalizada das unidades V-Vs[14] (no máximo, 34 medidas);

[14] As referidas medidas são obtidas automaticamente pelo *script* por meio de um procedimento de normalização básico em estatística, que consiste em obter a duração normalizada das unidades V-Vs (z−score estatístico que

3. Medidas de frequência fundamental (no máximo, 65 medidas);
4. Medidas de intensidade (4 medidas);
5. Medidas de pausa física (2 medidas).

A seguir, a tabela 11 mostra de forma sintética os parâmetros extraídos automaticamente pelo *script*.[15]

Tabela 11: Parâmetros fonético-acústicos extraídos automaticamente pelo *script* para Praat BreakDescriptor

Grupo	**Tipo**	**Medidas**
Taxa de elocução (*speech rate*) e ritmo	Global	Taxa de unidades V-V por segundo (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
		Taxa de unidades V-V não-salientes por segundo (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
Duração normalizada das unidades V-Vs	Local	Média de z-score suavizado de duração da unidade V-V (em cada unidade V-V que compõe as janelas à direita e à esquerda; diferença entre o z-score de duração da unidade V-V imediatamente após a posição de análise e o z-score de duração da unidade V-V analisada)
	Global	Média de z-score suavizado de duração da unidade V-V (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
		Desvio-padrão de z-score suavizado de duração das unidades V-Vs (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
		Assimetria de z-score suavizado de duração das unidades V-Vs (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
		Taxa de picos de z-score suavizado de duração das unidades V-Vs por segundo (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
Frequência fundamental	Local	Mediana de f0 em semitons re 1 Hz (em cada unidade V-V que compõe as janelas à direita e à esquerda; diferença entre a mediana de f0 da unidade V-V imediatamente após a posição de análise e mediana de f0 da unidade V-V analisada)
		Primeira derivada da mediana de f0 em semitons re 1 Hz (em cada unidade V-V que compõe as janelas à direita e à esquerda; diferença entre a primeira derivada da mediana de f0 da unidade V-V imediatamente após a posição de análise e a primeira derivada da mediana de f0 da unidade V-V analisada)
	Global	Média das medianas de f0 em semitons re 1 Hz (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
		Desvio-padrão das medianas de f0 em semitons re 1 Hz (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
		Assimetria das medianas de f0 em semitons re 1 Hz (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
		Média das primeiras derivadas das medianas de f0 em semitons re 1 Hz (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)

indica o afastamento do valor medido em relação a uma média em unidades de desvio-padrão). Em seguida, as curvas das durações normalizadas são suavizadas por meio de um procedimento padrão de média ponderada.

[15] Para obter maiores detalhes sobre cada uma das medidas extraídas automaticamente pelo *script*, veja-se Teixeira (2018).

		Desvio-padrão das primeiras derivadas das medianas de f0 em semitons re 1 Hz (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
		Taxa de picos de f0 suavizado por segundo (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
Intensidade	Local	Média de ênfase espectral em dB (diferença entre a média de ênfase espectral da unidade V-V imediatamente após a posição de análise e a média de ênfase espectral da unidade V-V analisada)
	Global	Média de ênfase espectral em dB (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
Pausa física	Local	Presença de pausa física após a fronteira (0 = ausência ou 1 = presença)
		Duração da pausa física após a fronteira

Uma última explicação é necessária para evitar dúvidas no que diz respeito ao parâmetro pausa física. Tendo em vista o modo de funcionamento do *script* para Praat BreakDescriptor, observa-se que algumas posições de fronteira localizadas no início ou fim do áudio são desconsideradas pelo *script* em virtude da extensão da janela de análise escolhida. Portanto, é natural que as pausas físicas localizadas bem no início ou fim do áudio não sejam computadas pelo *script.* Abaixo, a tabela 12 apresenta a distribuição de pausas físicas computados pelo *script* a partir da janela de análise composta por 10 unidades. A informação sobre a frequência de ocorrência das pausas físicas é apresentada, porque é fundamental para a compreensão dos resultados apresentados pelos modelos desenvolvidos a partir de medidas extraídas da janela análise composta por 10 unidades.

Tabela 12: Distribuição de pausas físicas contabilizadas a partir da janela de análise de 10 unidades

Posições	**Total**	**Amostra**	**Frequência**
Terminal com pausa física	137	I	58
		II	35
		Extra	44
Terminal sem pausa física	27	I	15
		II	11
		Extra	1
Não-terminal com pausa física	353	I	110
		II	129
		Extra	114
Não-terminal sem pausa física	373	I	134
		II	163
		Extra	76

Com a janela de análise de 10 unidades, dentre todas as fronteiras terminais contabilizadas nos dados, 83,5% são acompanhadas por pausa e 16,5% não são acompanhadas

por pausa. Dentre todas as fronteiras não-terminais, 48,6% são acompanhadas por pausa e 51,4% não são acompanhadas por pausa.

Com a janela de análise de 7 unidades, os parâmetros são extraídos em 15 unidades no total (7 unidades V-Vs anteriores, 7 unidades V-Vs posteriores e a própria unidade V-V em análise). Para o desenvolvimento de modelos a partir de medidas extraídas da janela de análise de 7 unidades, as posições de disfluências foram excluídas dos dados com base no mesmo critério de exclusão usado anteriormente. As posições contabilizadas a partir da janela de 7 unidades são expostas na tabela 13.

Tabela 13: Distribuição de posições contabilizadas a partir da janela de análise de 7 unidades

Posições	**Total**	**Amostra**	**Frequência**
Terminal	165	I	73
		II	47
		Extra	45
Não-terminal	746	I	250
		II	296
		Extra	200
Sem nenhuma fronteira	2198	I	1002
		II	769
		Extra	427

Com janelas compostas por 7 unidades, são extraídos 93 parâmetros fonético-acústicos no total. A seguir, a tabela 14 apresenta a distribuição de pausas físicas computadas pelo *script* a partir da janela de análise composta por 7 unidades.

Tabela 14: Distribuição de pausas físicas contabilizadas a partir da janela de análise de 7 unidades

Posições	**Total**	**Amostra**	**Frequência**
Terminal com pausa física	143	I	62
		II	37
		Extra	44
Terminal sem pausa física	22	I	11
		II	10
		Extra	1
Não-terminal com pausa física	367	I	116
		II	130
		Extra	121
Não-terminal sem pausa física	379	I	134
		II	166
		Extra	79

Com a janela de análise de 7 unidades, dentre todas as fronteiras terminais contabilizadas, 86,7% são acompanhadas por pausa e 13,3% não são acompanhadas por pausa. Dentre todas as fronteiras não-terminais, 49,2% são acompanhadas por pausa e 50,8% não são acompanhadas por pausa.

Com a janela de análise de 3 unidades, os parâmetros são extraídos em 7 unidades no total (3 unidades V-Vs anteriores, 3 unidades V-Vs posteriores e a própria unidade V-V em análise). Para o desenvolvimento de modelos a partir de medidas extraídas da janela de análise de 3 unidades, as posições de disfluências foram excluídas dos dados com base no mesmo critério de exclusão usado anteriormente (vide seção 3.2). As posições contabilizadas a partir da janela de 3 unidades são expostas na tabela 15.

Tabela 15: Distribuição de posições contabilizadas a partir da janela de análise de 3 unidades

Posições	**Total**	**Amostra**	**Frequência**
Terminal	168	I	76
		II	47
		Extra	45
Não-terminal	768	I	258
		II	302
		Extra	208
Sem nenhuma fronteira	2279	I	1030
		II	803
		Extra	446

No total, com janelas compostas por 3 unidades, são extraídos 69 parâmetros fonético-acústicos. A seguir, a tabela 16 apresenta a distribuição de pausas físicas computadas pelo *script* a partir da janela de análise composta por 3 unidades nas amostras.

Tabela 16: Distribuição de pausas físicas contabilizadas a partir da janela de análise de 3 unidades

Posições	**Total**	**Amostra**	**Frequência**
Terminal com pausa física	145	I	64
		II	37
		Extra	44
Terminal sem pausa física	23	I	12
		II	10
		Extra	1
Não-terminal com pausa física	374	I	116
		II	133
		Extra	125
Não-terminal sem pausa física	394	I	142

	II	169
	Extra	83

Com a janela de análise de 3 unidades, dentre as fronteiras terminais contabilizadas, 86,3% são acompanhadas por pausa física e 13,7% não são acompanhadas por este parâmetro. Dentre as fronteiras não-terminais, 48,6% são acompanhadas por pausa física e 51,5% não são acompanhadas por pausa física.

3.4. Resumo do capítulo

No capítulo 3, foram apresentados os dados utilizados neste trabalho, as tarefas de percepção de fronteiras realizadas pelos segmentadores, as camadas de TextGrid anotadas, os critérios de funcionamento do *script* usado para extrair uma série de parâmetros fonético-acústicos em uma janela à direita e à esquerda de cada fronteira de palavra fonológica e os parâmetros fonético-acústicos extraídos automaticamente pelo *script*.

4. Identificação automática das fronteiras prosódicas

Algoritmos de Aprendizagem de Máquina (em inglês, *Machine Learning*) são sistemas computacionais analíticos com alta capacidade de aprender, a partir da análise de dados, características relevantes sobre diversos fenômenos. Os Algoritmos de Aprendizagem de Máquina generalizam padrões sobre o fenômeno de interesse com o propósito de predizer a realização do fenômeno em outros dados. A principal vantagem do uso destes algoritmos é o desenvolvimento de modelos automáticos, que analisam uma grande quantidade de dados e executam tarefas de forma absolutamente consistente e muito mais rápida do que manualmente. Os algoritmos de *Machine Learning* se dividem em algoritmos: I) Supervisionados; II) Não supervisionados; III) Semissupervisionados.

Os algoritmos de aprendizagem supervisionada operam com um conjunto de dados previamente etiquetados, aprendendo formas de prever em novos dados a ocorrência de um certo fenômeno. Os algoritmos não supervisionados, também chamados de algoritmos de observação, são muito usados para observar similaridades entre objetos e incluí-los em grupos inferidos automaticamente pelo algoritmo. Os algoritmos não supervisionados não operam a partir de etiquetas já definidas previamente nos dados, eles analisam os dados, deduzem regras gerais (padrões) e encontram relações entre as regras. Posteriormente, este conhecimento é

usado para separar os dados em diferentes grupos que o próprio algoritmo encontrou. Algoritmos semissupervisionados representam uma junção dos algoritmos supervisionados com os não supervisionados. Estes algoritmos são ideais para o desenvolvimento de sistemas de previsão que necessitam analisar uma grande base de dados, mas, em contrapartida, dispõem de poucos dados devidamente etiquetados. A ideia geral dos algoritmos semissupervisionados é depreender as regras para a predição de dados já etiquetados, mas também depreender regras que explicam os dados não etiquetados, porque estes últimos também revelam informações úteis para o desenvolvimento de um sistema de previsão mais preciso.

Outro conceito importante na área de *Machine Learning* é a classificação estatística. A classificação estatística pode ser entendida como o processo em que etiquetas (categorias) são atribuídas automaticamente por meio de algoritmos de *Machine Learning* treinados para este fim. Tendo em vista os objetivos deste trabalho e o tratamento já recebido pelos dados, serão utilizados exclusivamente algoritmos supervisionados para realizar a classificação das fronteiras prosódicas. Deste modo, os algoritmos supervisionados atuarão no conjunto de dados aprendendo formas (equações) para predizer a realização de fronteira prosódica, de acordo com as categorias (etiquetas) de fronteira prosódica atribuídas perceptualmente pelos segmentadores.

Optou-se pelo uso de um algoritmo supervisionado que compara as diferenças (ou seja, observa a separação) entre grupos e depois classifica um objeto como pertencente a um grupo com perfil mais semelhante. Este algoritmo supervisionado é denominado *Linear Discriminant Analysis* (LDA) e, neste trabalho, ele foi usado para realizar a classificação de fronteiras prosódicas identificadas perceptualmente pelos segmentadores. Em linhas gerais, os objetivos do LDA são determinar se existem diferenças entre dois (ou mais) grupos já definidos, estimar a importância relativa de determinadas variáveis para diferenciar os grupos e predizer se os dados pertencem a um ou outro grupo. Deste modo, o LDA requer que os grupos (categorias) sejam conhecidos previamente com intuito de obter uma função descriminante a ser usada para alocar novos objetos.

O cálculo da função discriminante do LDA baseia-se em uma variável categórica dependente e em múltiplas variáveis independentes. No nosso caso, a variável categórica dependente engloba as fronteiras e posições sem fronteiras atribuídas pela maioria dos segmentadores; a variável independente engloba os parâmetros fonético-acústicos relativos às posições em final de palavra fonológica. O algoritmo LDA foi usado por meio do ambiente para computação estatística R (R Core Team, 2017) para atender a propósitos diferentes que se complementam entre si. O LDA foi usado para desenvolver os modelos de identificação

automática das fronteiras percebidas pela maior parte dos segmentadores, mas também foi usado para validar/ testar os modelos desenvolvidos em dados diferentes.

O trabalho de desenvolvimento e validação dos modelos de identificação automática das fronteiras foi dividido em três partes. Na primeira parte do capítulo (vide seção 4.1), são apresentados os procedimentos de desenvolvimento, validação e revalidação dos modelos desenvolvidos a partir de parâmetros extraídos da janela de análise de 10 unidades. Conforme é exposto a seguir na seção 4.1, os modelos desenvolvidos a partir dos parâmetros extraídos da janela composta por 10 unidades têm como ponto em comum incluir parâmetros relativos às unidades mais próximas ao final das palavras fonológicas. Tendo em vista este fato, as janelas de extração dos parâmetros fonético-acústicos através do BreakDescriptor foram reduzidas para 7 e 3 unidades. A vantagem dessa mudança é extrair com maior rigor os parâmetros relativos às fronteiras, pois eventuais configurações de parâmetros causadas por fatores não relacionados às fronteiras tendem a ser desconsideradas.

Na segunda parte do capítulo 4 (vide seção 4.2), são apresentados os procedimentos de desenvolvimento e validação dos modelos desenvolvidos a partir de parâmetros extraídos da janela de análise de 7 unidades. Na terceira parte do capítulo 4 (vide seção 4.3), são apresentados os procedimentos de desenvolvimento e validação dos modelos desenvolvidos a partir de parâmetros extraídos da janela de análise de 3 unidades.

Independentemente da extensão da janela de análise, algumas definições relativas às categorias são fixas. Em todos os modelos destinados à identificação das fronteiras terminais, foi analisada a oposição entre as categorias ausência de TB versus presença de TB, isto é, foram usadas as posições de TB, NTB e posições sem nenhuma fronteira (NB). Nos modelos destinados à identificação das fronteiras não-terminais, foi analisada a oposição NTB versus NB, devido a maior dificuldade do algoritmo para detectar as posições de NTB e diferenciá-las de TB (Teixeira, 2018). O processo de desenvolvimento dos modelos abordados nas seções 4.1 e 4.2 foi semelhante. Em ambos os casos, o desenvolvimento dos modelos foi feito heuristicamente, avaliando o efeito da inclusão e exclusão de cada medida no desempenho do modelo que estava sendo treinado naquela ocasião. Todas as medidas extraídas pelo BreakDescriptor foram avaliadas. Sob a condição das janelas de análise compostas por 10, 7 e 3 unidades, são analisadas 111, 93 e 69 medidas respectivamente.

4.1. Modelos de identificação de fronteiras desenvolvidos a partir de parâmetros extraídos de janelas de análise compostas por 10 unidades

O trabalho sobre os modelos de identificação automática de fronteiras desenvolvidos a partir de parâmetros extraídos das janelas de 10 unidades foi dividido em três partes principais. Na primeira parte (vide seção 4.1.1), os modelos desenvolvidos em Teixeira (2018) foram aplicados nas amostras II e total (amostra I acrescida da amostra II) para fins de revalidação em dados diferentes. Na segunda parte (vide seção 4.1.2), a partir de todas as posições em final de palavra fonológica disponíveis na amostra total, foram desenvolvidos dois modelos de identificação automática de fronteiras. Na terceira parte (vide seção 4.1.3), a partir de categorias balanceadas entre si, foram desenvolvidos novos modelos para identificar as fronteiras.

4.1.1. Revalidação de modelos anteriores de identificação de fronteira

Abaixo, as seções 4.1.1.1 e 4.1.1.2. apresentam respectivamente os resultados relativos à revalidação dos modelos anteriores de identificação de fronteiras terminais e não-terminais (Teixeira, 2018).

4.1.1.1. Revalidação do modelo anterior de identificação de fronteiras terminais

O procedimento adotado em Teixeira (2018) consistiu em desenvolver o modelo TB-nb destinado à identificação de fronteiras terminais a partir da escolha aleatória de 70% da amostra I e, depois, validar o modelo em 30% dos dados. Para fins de comparação, abaixo, na tabela 17, reportamos o desempenho alcançado pelo modelo TB-nb em Teixeira (2018) e o desempenho obtido na revalidação com as amostras II e total.

Tabela 17: Desempenho do modelo TB-nb nas fases de desenvolvimento, validação e revalidação

Fase	**Amostra**	**TB ident.**	**TB c/ pausa fís. ident.**	**TB s/ pausa fís. ident.**	**FA**	**FA c/ pausa fís.**	**FA s/ pausa fís.**
Des[16]	I	82% (36)	100% (36)	NA	5% (43)	100% (43)	NA
Val[17]	I	80% (20)	100% (20)	NA	8% (29)	100% (29)	NA

[16] Teixeira (2018).
[17] Teixeira (2018).

Rev-1[18]	II	70% (32)	100% (32)	NA	8,6% (90)	100% (90)	NA
Rev-2[19]	Total	74,8% (89)	100% (89)	NA	7,1% (162)	100% (162)	NA

O modelo TB-nb identifica em todas as amostras entre 70% e 82% das fronteiras terminais. Naturalmente, na amostra usada no desenvolvimento do modelo, o desempenho quanto à identificação das fronteiras terminais é melhor. Nas amostras II e Total, a porcentagem de fronteiras terminais reconhecidas é respectivamente 70% e 74,8%. No modelo TB-nb, os parâmetros fonético-acústicos que tratam da pausa física apresentam os maiores pesos relativos. Abaixo, a composição do modelo TB-nb é apresentada novamente na tabela 18.

Tabela 18: Composição do modelo TB-nb

Rank	**Sigla**	**Parâmetro fonético-acústico**	**Peso**
1º	psdur	Duração da pausa física após a fronteira	2,641
2º	psp	Presença de pausa física após a fronteira	1,948
3º	f0meddloc	Reset de f0	0,329
4º	df0medr1	Inclinação média da f0 na 1ª unidade da janela à direita da fronteira	0,264
5º	df0medl	Inclinação média de f0 dentro da janela à esquerda da fronteira	0,257
6º	sddf0d	Diferença de variação das inclinações de f0 nas janelas dos dois lados da fronteira	0,157
7º	prd	Diferença de taxa de saliência duracional dos dois lados da fronteira	0,101
8º	sdf0l	Variação nas medianas da f0 na janela à esquerda da fronteira	0,091
9º	df0medl10	Inclinação média da f0 na 10ª unidade da janela à esquerda da fronteira	0,066
10º	f0rl	Taxa de produção de proeminência melódicas, associadas ao acento de *pitch* na janela à esquerda da fronteira	0,033
11º	df0meddloc	Diferença de inclinação média da f0 entre as unidades pós-fronteiriça e fronteiriça	0,032
12º	f0medd	Diferença entre medianas da f0 nas janelas dos dois lados da fronteira	0,029
13º	zl10	Duração normalizada média na unidade pré-fronteiriça	0,028
14º	skf0d	Diferença das caudas da distribuição de f0 dos dois lados da fronteira	0,025
15º	mzd	Diferença de duração normalizada média dos dois lados da fronteira	0,015
16º	skdf0d	Diferença de variação de cauda da distribuição de F0 nas janelas dos dois lados da fronteira	0,011
17º	SDzl	Variação da duração normalizada média na janela à esquerda da fronteira	0,010
18º	ard	Diferença de taxa de articulação dos dois lados da fronteira	0,003
19º	zdloc	Diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça	0,001
20º	emphl	Intensidade relativa na janela à esquerda da fronteira	0,001

Nas amostras I, II e total, todas as fronteiras terminais identificadas pelo modelo foram imediatamente seguidas por pausa física. Logo, nenhuma fronteira terminal sem pausa física

[18] Revalidação.
[19] Revalidação.

foi identificada. O modelo foi desenvolvido para marcar todas as posições de NTB (com e sem pausa) como ausência de TB, já que a predição do modelo é feita com base em uma escolha binária e toda fronteira não-terminal, evidentemente, se trata de posição de ausência de fronteira terminal. Na revalidação do modelo TB-nb com as amostras II e total, todas as fronteiras não-terminais sem pausa foram corretamente marcadas como posições de ausência de fronteira terminal. Uma série de fronteiras não-terminais com pausa, no entanto, foi indevidamente marcada como fronteira terminal, gerando, assim, falsos alarmes. Nas revalidações realizadas neste trabalho, entre 26% e 68% das fronteiras não-terminais com pausa provenientes das amostras II e total foram marcadas indevidamente como fronteira terminal.

De fato, nas fases de desenvolvimento, validação e revalidação do modelo em discussão, todos os falsos alarmes obtidos foram devidos às fronteiras não-terminais com pausa. A tendência de marcar fronteira não-terminal com pausa como fronteira terminal se repete nos demais modelos de identificação de fronteira terminal e sugere que os parâmetros relativos à pausa física são superestimados; contudo, nos demais modelos, uma pequena quantidade de fronteiras terminais sem pausa física foi identificada a depender do modelo. Este aspecto será retomado posteriormente quando os modelos capazes de identificar algumas fronteiras terminais sem pausa forem apresentados.

O modelo TB-nb é composto por parâmetros de pausa física, f0, taxa de articulação e poucas medidas de duração normalizada. Dentre todos os parâmetros incluídos no modelo, o peso relativo dos parâmetros de “pausa” é claramente muito maior do que o peso relativo dos demais parâmetros. Isso significa que o modelo realiza as suas predições observando principalmente os parâmetros relacionados à pausa física. Assim, embora não seja desejável, é compreensível o porquê o modelo alcance os resultados acima.

Como os parâmetros de “pausa” tendem a ser superestimados no processo de atribuição/reconhecimento das fronteiras terminais, de fato, ainda não há conhecimento consistente sobre a contribuição efetiva dos demais parâmetros na percepção das fronteiras terminais, embora o modelo seja organizado com base nos pesos relativos de outros parâmetros fonético-acústicos. É possível que os outros parâmetros sejam sempre subordinados à presença/ausência da pausa física, e que não contribuam de maneira significativa ao modelo. Assim, na verdade, o modelo TB-nb não se trata de um modelo com excelente poder de predição, mas sim um modelo com foco na observação dos parâmetros de pausa física, que acaba, por sua vez, identificando muitas fronteiras terminais, porque a maioria delas é acompanhada pelas pausas físicas. Desta forma, é necessário desenvolver estratégias que

possam minimizar a superestimação da pausa física como recurso de identificação automática das fronteiras terminais.

Apesar do modelo ter como característica identificar as fronteiras terminais se baseando sobretudo nos parâmetros de pausa física, o modelo inclui em sua composição outros parâmetros. No caso de f0, são incluídas medidas locais e globais de f0, mas, naturalmente, o peso relativo de algumas delas é maior do que o peso de outras. As medidas de f0 com maior peso relativo são o reset de f0 entre as unidades pós-fronteiriça e fronteiriça, a inclinação média de f0 na unidade pós-fronteiriça e a inclinação média de f0 dentro da janela à esquerda da fronteira nesta ordem. O reset de f0 indica o reajuste da frequência fundamental da unidade pós-fronteiriça em relação à f0 da unidade fronteiriça (sabemos que isso é considerado um correlato importante da terminalidade), a inclinação média de f0 na unidade pós-fronteiriça é usada para observar a mudança no contorno entonacional de f0 na unidade pós-fronteiriça e a velocidade dessa mudança, a inclinação média de f0 dentro da janela à esquerda da fronteira é uma medida global que demonstra se a curva de f0 é preponderantemente ascendente ou descendente antes da fronteira. Tendo em vista que a declinação de f0 é definida como a tendência natural da curva de f0 descender dentro de um enunciado ou trecho de fala (t'Hart et al., 1968; Ladd, 1984), a inclinação média de f0 dentro da janela anterior à fronteira (à esquerda da fronteira) é uma medida que pode ser associada indiretamente ao fenômeno da declinação de f0, pois esta medida expressa de maneira global a evolução do contorno entonacional de f0 dentro da janela de 10 unidades. O peso relativo dos demais parâmetros do modelo TB-nb varia de 0,001 a 0,15 e compreende parâmetros de f0, duração normalizada e apenas um de intensidade relativa com o peso mínimo.

4.1.1.2. Revalidação dos modelos anteriores de identificação de fronteiras não-terminais

Em Teixeira (2018), o modelo NTB-1 refere-se ao modelo desenvolvido a partir das posições de fronteiras terminais, não-terminais e posições sem nenhuma fronteira. Porém, o modelo NTB-1 apresentou uma grande dificuldade para distinguir fronteiras terminais (acompanhadas por pausa física) das fronteiras não-terminais (também acompanhadas por pausa física), marcando-as indevidamente como fronteira não-terminal. Em razão dos resultados relatados, no trabalho em discussão, o modelo NTB-1 foi aplicado apenas nas posições de fronteira não-terminal e posições sem nenhuma fronteira para verificar como

ficariam as predições. Os resultados mostraram que o modelo identificou 68% das fronteiras não-terminais, logo após parar de analisar/ observar as fronteiras terminais.

Tendo em vista o último resultado relatado acima, Teixeira (2018) desenvolveu outros modelos NTB para identificar as fronteiras não-terminais remanescentes, isto é, aquelas que não foram identificadas pelo modelo NTB-1. O segundo modelo NTB foi desenvolvido a partir das fronteiras não-terminais não identificadas pelo modelo NTB-1 e é chamado de modelo NTB-2. O terceiro modelo, denominado NTB-3, foi desenvolvido com o restante das fronteiras NTB que não foi devidamente identificada pelos modelos NTB-1 e NTB-2. Como o modelo NTB-3, foi treinado a partir de uma quantidade de dados mais restrita, seus resultados devem ser considerados com ressalvas. No processo de desenvolvimento dos modelos NTB-2 e NTB-3, as fronteiras terminais não foram consideradas, pois foi observado que as fronteiras terminais dificultaram o processo de reconhecimento das fronteiras não-terminais através do modelo NTB-1.

No trabalho de 2018, não foi realizado nenhum tipo de validação dos modelos NTB-1, NTB-2 e NTB-3. Como já mencionado, foi feito o processo de aplicação do modelo NTB-1; eliminação das fronteiras reconhecidas pelo modelo NTB-1; aplicação do modelo NTB-2 nas fronteiras não identificadas pelo modelo NTB-1; eliminação das fronteiras identificadas pelo modelo NTB-2; e, por último, aplicação do modelo subsequente (NTB-3) nas fronteiras restantes. A rigor, o termo “revalidação” remete a algo que foi validado de alguma forma. Embora nenhuma validação dos modelos NTB-1, NTB-2 e NTB-3 tenha sido feita anteriormente, neste trabalho, é usado o termo “revalidação” para que as aplicações dos modelos NTB-1, NTB-2, NTB-3 e TB-nb (Teixeira, 2018) nas amostras II e total sejam chamadas da mesma forma.

A revalidação dos modelos NTB-1, NTB-2 e NTB-3 considerou todas as fronteiras não-terminais e posições sem nenhuma fronteira das amostras II e total. O procedimento de revalidação consistiu em aplicar o modelo NTB-1, eliminar as fronteiras reconhecidas pelo modelo NTB-1, aplicar o modelo NTB-2, eliminar as fronteiras identificadas pelo modelo NTB-2, e, por fim, aplicar o modelo NTB-3 nas fronteiras não-terminais não identificadas pelos modelos NTB-1 e NTB-2. Nas tabelas 19, 20 e 21, são apresentados os resultados da revalidação de cada modelo em discussão. Para os três modelos NTB, a porcentagem especificada na coluna “NTB ident.” foi calculada sobre o total de fronteiras não-terminais identificadas por cada modelo nas amostras II e total. Os falsos alarmes foram calculados sobre o total de posições sem nenhuma fronteira das amostras II e total.

Tabela 19: Desempenho do modelo NTB-1 nas fases de desenvolvimento e revalidação

Fase	Amostra	NTB ident.	NTB c/ pausa fís. ident.	NTB s/ pausa fís. ident.	FA	FA s/ pausa fís.
Des[20]	I	68% (165)	62% (103)	38% (62)	22% (35)	100% (35)
Rev-1	II	66% (193)	57% (110)	43% (83)	27,8% (211)	100% (236)
Rev-2	Total	67% (358)	67% (239)	33% (119)	25,6% (447)	100% (447)

Na tabela 20, são apresentados os resultados da revalidação do modelo NTB-2.

Tabela 20: Desempenho do modelo NTB-2 nas fases de desenvolvimento e revalidação

Fase	Amostra	NTB ident.	NTB s/ pausa fís. ident.	FA	FA s/ pausa fís.
Des[21]	I	25% (61)	100% (61)	20% (32)	100% (32)
Rev-1	II	19% (55)	100% (55)	47% (357)	100% (357)
Rev-2	Total	22% (116)	100% (116)	38,8% (678)	100% (678)

Na tabela 21, são apresentados os resultados da revalidação do modelo NTB-3.

Tabela 21: Desempenho do modelo NTB-3 nas fases de desenvolvimento e revalidação

Fase	Amostra	NTB ident.	NTB s/ pausa fís. ident.	FA	FA s/ pausa fís.
Des[22]	I	5% (12)	100% (12)	12% (19)	100% (19)
Rev-1	II	4,5% (13)	100% (13)	13,6% (103)	100% (103)
Rev-2	Total	4,7% (25)	100% (25)	16,4% (287)	100% (287)

Abaixo, a composição dos três modelos NTB é reapresentada novamente na tabela 22.

[20] Teixeira (2018).
[21] Teixeira (2018).
[22] Teixeira (2018).

Tabela 22: Modelos NTB-1, NTB-2 e NTB-3

	Modelo NTB-1		**Modelo NTB-2**		**Modelo NTB-3**	
Rank	***Sigla*** ***Peso***	***Parâmetro fonético-acústico***	***Sigla*** ***Peso***	***Parâmetro fonético-acústico***	***Sigla*** ***Peso***	***Parâmetro fonético-acústico***
1º	zl0 4.5	Duração normalizada média na unidade fronteiriça	srl 0.72	Taxa de elocução dentro da janela à esquerda da fronteira	prl 151.6	Taxa de saliência duracional dentro da janela à esquerda da fronteira
2º	zr1 4.4	Duração normalizada média na unidade pós-fronteiriça	sddf0l 0.63	Variação das inclinações de f0 dentro da janela à esquerda da fronteira	prd 150.6	Diferença de taxa de saliência duracional dos dois lados da fronteira
3º	zdloc 4.2	Diferença de duração normalizada entre as unidades pós-fronteiriça e fronteiriça	sdf0l 0.47	Variação nas medianas da f0 dentro da janela à esquerda da fronteira	prr 149.5	Taxa de saliência duracional dentro da janela à direita da fronteira
4º	psp 2.6	Presença de pausa física após a fronteira	ard 0.45	Diferença de taxa de articulação dos dois lados da fronteira	sdf0r 0.5	Variação nas medianas da f0 dentro da janela à direita da fronteira
5º	psdur 2.3	Duração da pausa física após a fronteira	f0medl 0.37	Mediana de f0 dentro da janela à esquerda da fronteira	SDzl 0.3	Variação na duração normalizada média dentro da janela à esquerda da fronteira
6º	ard 0.3	Diferença de taxa de articulação dos dois lados da fronteira	f0rd 0.21	Diferença de taxa de produção de proeminência melódicas dos dois lados da fronteira	df0medr1 0.3	Inclinação média de f0 na unidade pós-fronteiriça
7º	srd 0.3	Diferença de taxa de elocução dos dois lados da fronteira	f0meddloc 0.10	*Reset* de f0	df0medl10 0.2	Inclinação média de f0 na unidade pré-fronteiriça
8º	sdf0d 0.2	Diferença de variação das medianas de f0 dos dois lados da fronteira	f0med0 0.09	Mediana de f0 na unidade fronteiriça	df0meddloc 0.1	Diferença de inclinação média de f0 entre as unidades pós-fronteiriça (r1) e fronteiriça (l0)
9º	zl10 0.2	Duração normalizada média na unidade pré-fronteiriça	f0medr1 0.05	Mediana de f0 na unidade pós-fronteiriça	-	
10º	-		emphl 0.01	Intensidade relativa dentro da janela à esquerda da fronteira		

O modelo NTB-1 compreende nove parâmetros fonético-acústicos; o modelo NTB-2 compreende dez parâmetros e o modelo NTB-3 é composto por oito parâmetros. Nos modelos NTB-1 e NTB-2, não há um ou dois parâmetros que se destacam com relação aos outros parâmetros de maneira exagerada. No modelo NTB-3, três parâmetros se destacam demais possivelmente devido à limitada quantidade de dados utilizada no desenvolvimento do modelo.

O modelo NTB-1 apresenta seis parâmetros relacionados à duração normalizada, dois parâmetros de pausa física e apenas um parâmetro de f0 com o penúltimo menor peso relativo entre os nove parâmetros do modelo. Os parâmetros que compõem o modelo NTB-1 podem ser divididos em três grupos com base no peso relativo: um primeiro grupo englobando os três primeiros parâmetros de duração normalizada com peso variando de 4,2 a 4,5; um segundo grupo englobando os parâmetros de pausa física com os pesos de 2,6 e 2,3; um terceiro grupo incluindo os outros parâmetros com o peso entre 0,2 e 0,3. Os parâmetros de maior peso relativo incluído no primeiro grupo são: a duração normalizada média da unidade fronteiriça; a duração normalizada média da unidade pós-fronteiriça; diferença de duração normalizada entre as unidades pós-fronteiriça e fronteiriça. Os outros parâmetros do modelo NTB-1 são: a presença de pausa física, a duração da pausa física, a diferença de taxa de articulação dos dois lados da fronteira, a diferença de taxa de elocução dos dois lados da fronteira, a diferença de variação das medianas de f0 dos dois lados da fronteira e a duração normalizada média na unidade pré-fronteiriça.

Se compararmos a composição do modelo NTB-1 com a composição dos modelos TB, é possível verificar que os parâmetros de duração normalizada assumem papéis diferentes. No modelo NTB-1, os parâmetros de duração normalizada apresentam peso relativo alto. Nos modelos TB, às vezes, os parâmetros de duração normalizada sequer são incluídos. Quando são incluídos, apresentam peso relativo baixo. Os parâmetros de f0, por sua vez, possuem um peso relativo baixo no modelo NTB-1, mas apresentam pesos relativos maiores nos modelos TB. Em todos os modelos de identificação de TB, pelo menos um parâmetro de pausa física apresenta um peso relativo bem alto. No caso do modelo NTB-1, o peso dos parâmetros relativos à pausa física é mais baixo, se comparado ao peso destes parâmetros nos modelos TB.

Os três modelos NTB em discussão se diferenciam entre si. O modelo NTB-1 compreende basicamente parâmetros de duração normalizada e pausa física. O modelo NTB-2 compreende principalmente parâmetros de f0. O modelo NTB-3 compreende um mix de parâmetros de duração normalizada e f0, mas é um modelo que naturalmente possui um menor poder de generalização, porque foi desenvolvido a partir de poucas fronteiras não-terminais (fronteiras remanescentes não identificadas pelos modelos NTB-1 e NTB-2). Enquanto é

possível dividir os parâmetros do modelo NTB-1 em três grupos com base no peso relativo, como já foi dito, os parâmetros do modelo NTB-2 não apresentam uma diferença clara em termos de peso relativo. No modelo NTB-3, os parâmetros relacionados à duração normalizada apresentam um peso muito maior do que os parâmetros relacionados a f0 (que, por outro lado, estão em maior número).

Em Teixeira (2018), o modelo NTB-1 identificou 68% das fronteiras não-terminais. Na revalidação do modelo NTB-1, a porcentagem de fronteiras não-terminais identificadas corresponde a 66% e 67% nas amostras II e total respectivamente. A porcentagem de fronteiras não-terminais devidamente identificadas nas revalidações é inferior à porcentagem de fronteiras identificadas anteriormente, mas, ainda assim, muito próxima da porcentagem de reconhecimento alcançado durante o desenvolvimento do modelo, mostrando que a generalização do modelo se mantém em amostras diferentes daquelas em que o modelo foi originalmente treinado.

Nas fases de desenvolvimento e revalidação do modelo NTB-1, entre 57% e 67% das fronteiras não-terminais identificadas são acompanhadas por pausa física. Portanto, entre 33% e 43% das fronteiras identificadas não são acompanhadas por pausa física. Os falsos alarmes correspondem a 22% na fase de desenvolvimento do modelo e entre 25,6% e 27,8% nas revalidações. Tanto no desenvolvimento, quanto na revalidação do modelo NTB, os falsos alarmes obtidos ocorreram apenas em posição sem nenhuma fronteira, pois as fronteiras terminais não foram observadas. Independente disso, a quantidade exagerada de falsos alarmes obtidos pode ser considerada a limitação do modelo em discussão.

O modelo NTB-2 identificou 25% das fronteiras não-terminais em Teixeira (2018). A porcentagem de fronteiras não-terminais identificadas nas revalidações com as amostras II e total corresponde a 19% e 22% respectivamente. Em ambos os casos, a porcentagem de fronteiras não-terminais identificadas foi calculada com base no total de fronteiras não-terminais contabilizadas nas amostras. Mais uma vez, a porcentagem de fronteiras não-terminais identificadas nas revalidações é inferior à porcentagem de fronteiras identificadas quando o modelo foi desenvolvido, porém, essas porcentagens de reconhecimento são relativamente próximas. Neste caso, as fronteiras identificadas são todas sem pausa física, pois as fronteiras não-terminais acompanhadas pela pausa física foram todas identificadas anteriormente pelo modelo NTB-1. O principal problema do modelo NTB-2 é a quantidade exagerada de falsos alarmes em posição sem fronteira prosódica. Os falsos alarmes obtidos correspondem a 20% na fase de desenvolvimento do modelo e aumentam muito nas revalidações, totalizando 38,8% na revalidação com a amostra total e 47% na amostra II.

O modelo NTB-3 identificou apenas 5% das fronteiras não-terminais na fase de desenvolvimento em Teixeira (2018). Nas revalidações, a porcentagem de fronteiras não-terminais identificadas corresponde a 4,5% e 4,7% nas amostras II e total. Os falsos alarmes correspondem a 12% no desenvolvimento do modelo e entre 13,6% e 16,4% nas revalidações. Como já foi dito, o modelo NTB-3 possui um menor poder de generalização devido à limitada quantidade de dados que foi desenvolvido, mas a obtenção de falsos alarmes permanece.

Tanto na revalidação com a amostra II, quanto na revalidação com a amostra total, a porcentagem total de fronteiras identificadas pelos modelos NTB-1, NTB-2 e NTB-3 é alta. Na amostra II, os três modelos juntos identificaram 90% das fronteiras não-terminais. A porcentagem de fronteiras NTB identificadas pelos modelos NTB-1, NTB-2 e NTB-3 corresponde a 66%, 19,5% e 4,5% respectivamente. Na amostra total, os três modelos identificaram 93,7% das fronteiras NTB. Neste caso, a porcentagem de fronteiras identificadas pelos modelos NTB-1, NTB-2 e NTB-3 corresponde a 67%, 22% e 4,7% respectivamente. Em Teixeira (2018), a porcentagem de reconhecimento das fronteiras NTB corresponde a 98% (68% de reconhecimento com o modelo NTB-1; 25% com o modelo NTB-2 e 5% com o modelo NTB-3).

Nessas revalidações dos modelos NTB-1, NTB-2 e NTB-3, as posições analisadas compreendiam apenas posições sem nenhuma fronteira e NTB. Desta forma, os modelos não avaliaram as posições de TB e, evidentemente, não foram obtidos falsos alarmes em posição de TB com pausa física. Apesar da estratégia de restringir a aplicação dos modelos NTB às posições sem fronteira e NTB possivelmente facilitar o processo de reconhecimento de NTB, ainda assim, a quantidade de falsos alarmes obtidos é alta e, neste caso, os falsos alarmes ocorreram apenas em posição sem nenhuma fronteira prosódica.

Ao término da realização das revalidações apresentadas, os modelos NTB-1, NTB-2 e NTB-3 foram aplicados em todas as posições sem fronteira e NTB da amostra total, sem realizar eliminar progressivamente as fronteiras já reconhecidas. Em outras palavras, com exceção das posições de TB, todos os modelos NTB analisaram todas as posições. Os resultados mostraram que algumas fronteiras foram identificadas pelos modelos NTB-1, NTB-2 e NTB-3, outras foram identificadas por dois modelos, e apenas uma parte foi identificada apenas por um único modelo sozinho. No total, 191 fronteiras não-terminais foram identificadas por mais de um modelo. As fronteiras NTB identificadas por mais de um modelo correspondem a 36% das NTB contabilizadas na amostra total. Dentre elas, 123 são acompanhadas por pausa física. Desta forma, até mesmo os modelos que não incluem em sua composição parâmetros relativos à pausa física foram capazes de identificar algumas NTB com pausa física.

4.1.2. Modelos de identificação de fronteiras desenvolvidos a partir de todas as posições disponíveis

Foram desenvolvidos dois modelos a partir de todas as posições encontradas na amostra total. Um dos modelos é destinado à identificação automática das fronteiras terminais (vide seção 4.1.2.1). O outro é destinado às fronteiras não-terminais (vide seção 4.1.2.2). Em ambos os casos, a janela de análise escolhida para extrair os parâmetros fonético-acústicos é composta por 10 unidades. A estratégia de desenvolvimento de modelos a partir da base completa de dados é útil para testar o potencial máximo de acertos e erros dos modelos, embora não seja possível realizar a validação dos modelos em outros dados. Abaixo, as seções 4.1.2.1 e 4.1.2.2. apresentam respectivamente os modelos TB e NTB em questão bem como os seus resultados.

4.1.2.1. Modelo de identificação de fronteiras terminais

O modelo de identificação de fronteiras terminais desenvolvido a partir de todas as posições da amostra total é denominado TB-nbfull. Abaixo, na tabela 23, são apresentados os resultados do modelo TB-nbfull.

Tabela 23: Desempenho do modelo TB-nbfull na fase de desenvolvimento

Fase	Amostra	TB ident.	TB c/ pausa fís. ident.	TB s/ pausa fís. ident.	FA	FA c/ pausa fís.	FA s/ pausa fís.
Des	Total	65,5% (77)	100% (77)	NA	2,9% (66)	100% (66)	NA

A tabela 24 apresenta a composição do modelo TB-nbfull em termos de peso relativo de parâmetros fonético-acústicos observados para realizar as predições.

Tabela 24: Composição do modelo TB-nbfull

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psdur	Duração da pausa física após a fronteira	4,327162
2º	psp	Presença de pausa física após a fronteira	0,583102
3º	df0medr1	Inclinação média de f0 na unidade pré-fronteiriça	0,334025
4º	f0meddloc	Reset de f0	0,307601
5º	sddf0l	Variação das inclinações de f0 dentro da janela à esquerda da fronteira	0,274378
6º	skdf0l	Variação de cauda da distribuição de f0 dentro da janela à esquerda da fronteira	0,159998
7º	df0med0	Inclinação média de f0 na unidade fronteiriça	0,143298

8º	df0meddloc	Diferença de inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça	0,122687
9º	sdf0l	Variação nas medianas da f0 dentro da janela à esquerda da fronteira	0,117736
10º	f0medl	Mediana de f0 dentro da janela à esquerda da fronteira	0,111284
11º	f0rd	Diferença de taxa de produção de proeminências melódicas dos dois lados da fronteira	0,063397
12º	f0med0	Mediana de f0 na unidade fronteiriça	0,061654
13º	f0medd	Diferença de mediana de f0 dos dois lados da fronteira	0,056556
14º	df0medl10	Inclinação média de f0 na unidade pré-fronteiriça	0,046247
15º	df0medd	Diferença de inclinação média de f0 dentro das janelas dos dois lados da fronteira	0,039197
16º	df0medr2	Inclinação média de f0 na 2ª unidade após a fronteira	0,026527
17º	srd	Diferença de taxa de elocução dos dois lados da fronteira	0,010834

O modelo TB-nbfull é composto por 17 parâmetros fonético-acústicos. Dentre eles, a duração da pausa física se destaca claramente em razão do maior peso relativo. No que diz respeito aos parâmetros de pausa física, o modelo TB-nbfull se diferencia do TB-nb, comentado na seção 4.1.1.1. No modelo TB-nb, o peso relativo da duração da pausa física e da presença de pausa física são, em ambos os casos, muito altos (respectivamente 2,64 e 1,94). No modelo TB-nbfull, embora os dois parâmetros relacionados à pausa física estejam incluídos, apenas um deles (a duração da pausa física) se destaca com o peso relativo de 4,32 e este peso é inclusive maior do que o peso do mesmo parâmetro no modelo TB-nb.

Com exceção da diferença da taxa de elocução dos dois lados da fronteira e os parâmetros relacionados à pausa física, todos os parâmetros incluídos no modelo TB-nbfull compreendem exclusivamente parâmetros de f0. Após a duração da pausa física, o parâmetro presença de pausa física apresenta o segundo maior peso relativo de 0,58. O peso de 0,58 é inclusive próximo ao peso apresentado pelos demais parâmetro do modelo. O terceiro parâmetro relevante hierarquicamente é a inclinação média de f0 na unidade pós-fronteiriça, em seguida, o reset de f0 entre as unidades pós-fronteiriça e fronteiriça e a variação das inclinações de f0 dentro da janela à esquerda da fronteira nesta ordem. Se comparado ao peso relativo da duração da pausa física, o peso destes três parâmetros de f0 é bem menor e gira em torno de aproximadamente 0,3. Os outros 12 parâmetros do modelo TB-nbfull apresentam um peso relativo bem mais baixo, que varia de 0,01 a 0,15.

Apesar de diversos parâmetros fonético-acústicos serem incluídos no modelo em discussão, o parâmetro duração da pausa física se destaca muito claramente e pode ser o único parâmetro realmente considerado pelo modelo para reconhecer as fronteiras terminais, de acordo com os resultados apresentados. No total, o modelo identificou aproximadamente 66% das fronteiras terminais da amostra total. Todas as fronteiras terminais identificadas são

imediatamente seguidas por pausa física e a sua duração varia de 131 ms a 1520 ms com média de 632 ms e mediana de 563 ms. Isso significa que nenhuma fronteira terminal sem pausa física foi devidamente identificada. Em razão destes resultados relatados, é reforçada a hipótese de que o reconhecimento das fronteiras é baseado principalmente na pausa física.

Foram obtidos apenas 2,9% de falsos alarmes (66 casos de falsos alarmes que correspondem a 12% das fronteiras não-terminais contabilizadas na amostra total). Os falsos alarmes ocorreram apenas em posição de NTB com pausa física. A duração das pausas físicas que acompanham os falsos alarmes varia de 267 ms a 1475 ms com média de 654 ms e mediana de 621 ms. De fato, a duração máxima, média e mediana das pausas físicas que acompanham as fronteiras terminais reconhecidas (duração máxima de 1520 ms; média de 632 ms e mediana de 563 ms) é próxima à duração máxima, média e mediana das pausas físicas que acompanham os falsos alarmes (duração máxima de 1475 ms; média de 654 ms e mediana de 621 ms). As durações mínimas, no entanto, se diferem (131 ms no caso das fronteiras terminais identificadas e 267 ms no caso dos falsos alarmes). Tendo em vista as durações apresentadas, é justificável o porquê o modelo alcançou os resultados relatados, embora os falsos alarmes se configurem como predições erradas e exista esta pequena diferença quanto à duração mínima. Assim, mais uma vez, o conjunto de resultados apresentados pelo modelo pode ser visto como indício de que os parâmetros de pausa física são superestimados no processo de identificação automática de fronteira terminal.

Apesar do modelo em discussão parecer observar principalmente os parâmetros relativos às pausas físicas para realizar as suas predições, é importante destacar que determinadas fronteiras terminais acompanhadas por pausa física não foram devidamente identificadas, isto é, nem toda posição com pausa física é imediatamente marcada como fronteira terminal. As fronteiras terminais acompanhadas por pausa física, mas não identificadas pelo modelo, correspondem a 17,6% das fronteiras terminais contabilizadas na amostra total e a duração das pausas físicas varia de 212 ms a 566 ms com média de 358 ms e mediana de 319 ms. Em razão deste último resultado, é possível propor a hipótese de que o modelo observaria primeiramente a duração das pausas físicas.

Neste sentido, o modelo marcaria como TB as posições de pausa física com duração maior. O quanto maior, no entanto, não é claro e nem facilmente acessível. Em caso de duração menor, entrariam em jogo outros parâmetros para a tomada de decisão. Quanto aos demais parâmetros do modelo em discussão, duas hipóteses são possíveis. É possível que o modelo avalie/observe os demais parâmetros somente após analisar a duração das pausas físicas, mas

também é possível que os demais parâmetros não exerçam nenhum papel no reconhecimento automático das fronteiras e, se isso de fato ocorrer, trata-se de uma limitação do modelo.

Em geral, o modelo TB-nbfull se trata de um modelo com poder de predição intermediário, que realiza as suas predições observando principalmente a pausa física. O baixo peso relativo do parâmetro presença de pausa física junto com o elevado peso relativo da duração da pausa física reduzem um pouco a porcentagem de fronteiras terminais identificadas, mas, ao mesmo tempo, reduzem bastante os falsos alarmes devidos às fronteiras não-terminais acompanhadas por pausa física. Desta forma, o modelo, de fato, tende a marcar como TB as posições de pausa física, porém, o processo de identificação aparenta ser mais sofisticado do que marcar qualquer posição de pausa física como TB.

4.1.2.2. Modelo de identificação de fronteiras não-terminais

Foi desenvolvido um modelo de identificação de fronteiras não-terminais (denominado NTB-nbfull) a partir de todas as posições sem fronteira e todas as fronteiras não-terminais da amostra total, isto é, as posições de fronteira terminal não foram consideradas. A exclusão das posições de fronteira terminal é motivada pelos resultados dos modelos NTB-1, NTB-2 e NTB-3, que funcionaram melhor após deixarem de analisar as fronteiras terminais. Além disso, a exclusão das fronteiras terminais foi realizada com intuito de avaliar a possibilidade de reconhecer primeiramente as fronteiras terminais com um determinado modelo TB e, após o reconhecimento das fronteiras terminais, aplicar um dado modelo NTB nas posições restantes apenas para identificar as fronteiras não-terminais. Abaixo, a tabela 25 apresenta os resultados do modelo NTB-nbfull.

Tabela 25: Desempenho do modelo NTB-nbfull na fase de desenvolvimento

Fase	Amostra	NTB ident.	NTB c/ pausa fís. ident.	NTB s/ pausa fís. ident.	FA	FA s/ pausa fís.
Des	Total	40,5% (217)	99,5% (216)	0,5% (1)	0,3% (4)	100% (4)

Abaixo, a tabela 26 apresenta a composição do modelo NTB-nbfull em termos de peso relativo de parâmetros fonético-acústicos incluídos no modelo.

Tabela 26: Composição do modelo NTB-nbfull

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psp	Presença de pausa física após a fronteira	3,503017
2º	zdloc	Diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça	0,08795
3º	zl10	Duração normalizada média na unidade pré-fronteiriça	0,082874
4º	zl0	Duração normalizada média na unidade fronteiriça	0,078092
5º	ard	Diferença de taxa de articulação dos dois lados da fronteira	0,063516

O modelo NTB-nbfull é composto por parâmetros de pausa física, duração normalizada e taxa de articulação. No total, o modelo inclui apenas 5 parâmetros, se configurando como um modelo muito econômico no que diz respeito à quantidade de variáveis de predição. O parâmetro de maior peso relativo é a presença de pausa física com o peso de 3,5. O segundo parâmetro de maior peso relativo, mas com peso muito menor com relação à presença de pausa física, é a diferença de duração normalizada entre as unidades pós-fronteiriça e fronteiriça com o peso de 0,087. Os dois primeiros parâmetros seguidos pela duração normalizada média na unidade pré-fronteiriça, duração normalizada média na unidade fronteiriça e pela diferença de taxa de articulação dos dois lados da fronteira nesta ordem.

O modelo em discussão reconheceu um número reduzido de fronteiras não-terminais, se comparado ao modelo NTB-1. Ao analisar as posições sem fronteira e as NTB da amostra total, o modelo identificou aproximadamente 40,5% das fronteiras não-terminais das quais 99,5% são acompanhadas por pausa física. A capacidade de reconhecer principalmente NTB com pausa física junto com o não reconhecimento das NTB sem pausa física sugerem que a pausa física é superestimada no processo de reconhecimento das fronteiras. No total, o modelo obteve apenas 4 falsos alarmes, o que pode ser considerado um grande ponto positivo.

De modo geral, ainda que um parâmetro de duração normalizada seja o segundo de maior relevância com base no peso relativo, o modelo parece tomar as suas decisões pautando-se sobretudo na presença de pausa física, influenciando o funcionamento do modelo como um todo e, talvez, inibindo a contribuição dos demais parâmetros. O elevado peso da presença de pausa física é um ponto em comum entre o modelo NTB-nbfull e alguns modelos TB. Por isso, é natural que o modelo em análise marque TB com pausa física como NTB, caso ele seja aplicado em posições de TB também. O peso relativo dos demais parâmetros fonético-acústicos não se destaca e varia de 0,06 a 0,08.

4.1.3. Modelos de identificação de fronteiras desenvolvidos a partir de categorias balanceadas entre si

Foram desenvolvidos diversos modelos de identificação automática de fronteiras a partir de pequenas amostras compostas por categorias completamente balanceadas entre si. O balanceamento foi realizado de modo que uma série de amostras fossem geradas com um número equivalente de exemplares de cada categoria a ser treinada. Todo o processo de balanceamento das categorias foi feito aleatoriamente.

Para realizar o balanceamento das categorias, foi considerado o total de fronteiras terminais e não-terminais contabilizado a partir da janela de análise composta por 10 unidades. No caso das fronteiras terminais, o balanceamento dos dados foi feito observando o número total de fronteiras terminais encontradas em cada amostra e selecionando aleatoriamente uma quantidade equivalente de posições de ausência de TB. No caso de fronteiras não-terminais, o balanceamento foi feito observando o número total de fronteiras não-terminais encontradas em cada amostra e selecionando aleatoriamente a quantidade equivalente de posições sem nenhuma fronteira (NB). Abaixo, a tabela 27 apresenta uma visão geral das amostras balanceadas utilizadas para desenvolver os modelos.

Tabela 27: Visão geral das amostras balanceadas

Balanceamento em função do nº de fronteiras	Balanceamento em função da amostra	Amostra balanceada gerada	Etiqueta da amostra balanceada gerada
TB	I	Amostra balanceada em função do total de TB encontrado na amostra I	I-b (TB)
TB	II	Amostra balanceada em função do total de TB encontrado na amostra II	II-b (TB)
TB	Total	Amostra balanceada em função do total de TB encontrado na amostra total	FULL-b (TB)
NTB	I	Amostra balanceada em função do total de NTB encontrado na amostra I	I-b (NTB)
NTB	II	Amostra balanceada em função do total de NTB encontrado na amostra II	II-b (NTB)
NTB	Total	Amostra balanceada em função do total de NTB encontrado na amostra total	FULL-b (NTB)

Foram desenvolvidos seis modelos diferentes (um a partir de cada amostra apresentada na tabela 27). Posteriormente, os modelos desenvolvidos foram aplicados nas amostras I, II e

total não balanceadas para fins de validação. A validação dos modelos TB foi feita aplicando os modelos em discussão em todas as posições de TB, NTB e NB das amostras I, II e total. Os modelos NTB foram validados de duas formas diferentes. Primeiramente, a validação foi feita aplicando os modelos NTB em todas as posições de NTB e NB das amostras I, II e total. Depois, os modelos NTB foram aplicados em todas as posições de NTB, TB e NB das amostras I, II e total. Tanto na validação dos modelos TB, quanto nas validações dos modelos NTB, as amostras usadas são não balanceadas, visto que o número de exemplares de cada categoria é diferente. Os resultados relativos ao desenvolvimento de modelos a partir das amostras balanceadas bem como os resultados relativos às validações em dados não balanceados são expostos nas seções 4.1.3.1 e 4.1.3.2.

4.1.3.1. Modelos de identificação de fronteiras terminais

As próximas seções apresentam os modelos de identificação de fronteiras terminais desenvolvidos a partir das amostras balanceadas em função da quantidade de TB encontrada nas amostras I, II e total. Adicionalmente, são apresentados os resultados alcançados pelos modelos em discussão.

a. Modelo desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras terminais da amostra I

O modelo TB desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras terminais da amostra I é denominado TB-b1. A seguir, a tabela 28 apresenta o desempenho do modelo TB-b1 nas fases de desenvolvimento e aplicação nas amostras I, II e total não balanceadas.

Tabela 28: Desempenho do modelo TB-b1 nas fases de desenvolvimento e validações

Fase	Amostra	TB ident.	TB c/ pausa fís. ident.	TB s/ pausa fís. ident.	FA	FA c/ pausa fís.	FA s/ pausa fís.
Des	I-b (TB)	84,2% (61)	96,7% (59)	3,3% (2)	6,8% (5)	100% (5)	NA
Val-1	I	84,2% (61)	96,7% (59)	3,3% (2)	7,5% (92)	100% (92)	NA
Val-2	II	76,3% (35)	100% (35)	NA	12,4% (130)	97,7% (127)	2,3% (3)
Val-3	Total	80,7% (96)	97,9% (94)	2,1% (2)	9,7% (222)	98,6% (219)	1,4% (3)

Abaixo, a tabela 29 apresenta os parâmetros-fonético-acústicos e pesos dos parâmetros que compõem o modelo TB-b1.

Tabela 29: Composição do modelo TB-b1

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psp	Presença de pausa física após a fronteira	2,68983234
2º	df0medl	Inclinação média de f0 dentro da janela à esquerda da fronteira	0,59886888
3º	psdur	Duração da pausa física após a fronteira	0,1587707
4º	ard	Diferença de taxa de articulação dos dois lados da fronteira	0,13882117
5º	f0rd	Diferença de taxa de produção de proeminência melódicas dos dois lados da fronteira	0,10753909
6º	df0meddloc	Diferença de inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça	0,04081559
7º	f0meddloc	Reset de f0	0,02953472
8º	df0medl10	Inclinação média de f0 na unidade pré-fronteiriça	0,01667

O modelo TB-b1 pode ser descrito como um modelo composto por parâmetros de pausa física, f0 e taxa de articulação. O modelo inclui 8 parâmetros, se configurando como um modelo muito econômico no que diz respeito à quantidade de variáveis de predição. O parâmetro com peso relativo maior é a presença de pausa física. O segundo parâmetro com maior peso relativo, mas com peso muito menor com relação ao primeiro, é a inclinação média de f0 dentro da janela à esquerda da fronteira, mas não é claro o quanto o segundo parâmetro efetivamente contribui no reconhecimento das fronteiras. Assim, ainda que um parâmetro de f0 seja o segundo parâmetro mais relevante com base no peso relativo, corre-se o risco de o modelo tomar as suas decisões pautando-se sobretudo na pausa física, influenciando o funcionamento do modelo como um todo e, talvez, inibindo a contribuição dos demais parâmetros.

No modelo TB-b1, três parâmetros apresentam pesos relativos semelhantes. A duração da pausa física, a diferença da taxa de articulação dos dois lados da fronteira e a diferença da taxa de produção de proeminências melódicas se associam respectivamente aos pesos 0,15, 0,13 e 0,10. A diferença de taxa de articulação tem como correlato perceptivo básico a velocidade de fala mais lenta ou mais acelerada. A diferença de taxa de produção de proeminências melódicas nas janelas dos dois lados da fronteira se relaciona à produção de proeminências melódicas associadas ao acento de *pitch* (em inglês *pitch accent*). Geralmente, as proeminências são sinalizadas por um movimento de f0 (contorno de f0) e são consideradas um correlato da vivacidade da fala (em inglês, *liveliness*). Em razão dos pesos de 0,15, 0,13 e 0,10, a contribuição dos três parâmetros em discussão no funcionamento do modelo não parece ser expressiva.

Os três últimos parâmetros do modelo se relacionam à fenômenos locais de f0 e apresentam pesos relativos bem menores, se comparado ao peso dos cinco primeiros parâmetros do modelo. Estes parâmetros são a diferença de inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça, que expressa de forma aproximada a diferença de contorno entonacional de f0 da unidade pós-fronteiriça em relação ao contorno entonacional de f0 da unidade fronteiriça, o reset de f0 entre as unidades pós-fronteiriça e fronteiriça e a inclinação média de f0 na unidade pré-fronteiriça. Em todos os três parâmetros citados, os fenômenos de f0 ocorrem nas unidades imediatamente adjacentes às fronteiras, isto é, unidades mais próximas às posições em que os segmentadores atribuem a percepção das fronteiras.

O desempenho do modelo na tarefa de identificação das fronteiras terminais é quantitativamente bom, identificando em todas as amostras mais de 76% das fronteiras terminais percebidas pelos segmentadores. Na fase de desenvolvimento do modelo com a amostra balanceada em função da quantidade de fronteiras terminais da amostra I, 84,2% das fronteiras terminais foram identificadas. Nas aplicações do modelo nas amostras não balanceadas, o modelo identificou entre 76,3% e 84,2% das fronteiras terminais contabilizadas.

Na fase de desenvolvimento do modelo, foram obtidos apenas cinco falsos alarmes e eles ocorreram apenas em posição de NTB com pausa física. Nas validações, a porcentagem de falsos alarmes obtidos variou de 7,5% a 12,4% sob o total de posições de ausência de TB contabilizadas nas amostras I, II e total não balanceadas. Nas validações, os falsos alarmes ocorreram majoritariamente em posição de NTB acompanhada por pausa física.

Apesar de apresentar uma boa *performance* no que diz respeito ao número total e à porcentagem de TB identificada, o modelo parece identificar muitas fronteiras terminais porque realiza as suas predições observando principalmente a pausa física, e a maioria das fronteiras terminais, por sua vez, é acompanhada por este parâmetro. Outra questão que precisa ser pontuada no desempenho do modelo diz respeito às fronteiras terminais sem pausa física. Esta questão precisa ser observada com atenção, porque, possivelmente, compromete de forma negativa o desempenho geral do modelo. Assim como nos demais modelos TB já apresentados, o modelo não foi capaz de identificar a quase totalidade das fronteiras terminais sem pausa física. Nas fases de desenvolvimento e validação do modelo em discussão, foram identificadas apenas duas fronteiras terminais sem pausa física, o que pode limitar a *performance* do modelo ao aplicá-lo em dados de fala espontânea, pois aproximadamente 22% das fronteiras terminais não são acompanhadas pela pausa física.

A quantidade de falsos alarmes devidos às fronteiras não-terminais acompanhadas por pausa física, o baixo reconhecimento das fronteiras terminais sem pausa física bem como a

capacidade de identificar principalmente fronteiras com pausa física indicam que o modelo se baseou muito nas pausas físicas para realizar as suas predições.

b. Modelo desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras terminais da amostra II

O modelo TB desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras terminais da amostra II é denominado TB-b2. Abaixo, a tabela 30 apresenta o desempenho do modelo TB-b2 nas fases de desenvolvimento e validação nas amostras I, II e total não balanceadas.

Tabela 30: Desempenho do modelo TB-b2 nas fases de desenvolvimento e validações

Fase	Amostra	TB ident.	TB c/ pausa fís. ident.	TB s/ pausa fís. ident.	FA	FA c/ pausa fís.	FA s/ pausa fís.
Des	II-b (TB)	82,6% (38)	92,1% (35)	7,9% (3)	8,7% (4)	100% (4)	NA
Val-1	II	82,6% (38)	92,1% (35)	7,9% (3)	12,4% (130)	98,5% (128)	1,5% (2)
Val-2	I	80,8% (59)	98,3% (58)	1,7% (1)	7,9% (97)	94,9% (92)	5,1% (5)
Val-3	Total	81,5% (97)	95,9% (93)	4,1% (4)	9,9% (227)	96,9% (220)	3,1% (7)

Na tabela 31, são apresentados os parâmetros fonético-acústicos que compõem o modelo TB-b2.

Tabela 31: Composição do modelo TB-b2

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psp	Presença de pausa física após a fronteira	2,353419
2º	f0meddloc	Reset de f0	0,742071
3º	df0meddloc	Diferença de inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça	0,61342
4º	df0med0	Inclinação média de f0 na unidade fronteiriça	0,54787
5º	psdur	Duração da pausa física após a fronteira	0,487833

O modelo TB-b2 inclui em sua composição apenas cinco parâmetros fonético-acústicos, sendo um modelo ainda mais econômico do que o modelo TB-b1 em termos de quantidade de variáveis preditivas. Este modelo é composto por parâmetros relacionados à pausa física e à f0 das unidades imediatamente próximas à fronteira prosódica. Dentre os cinco parâmetros incluídos no modelo, a presença de pausa física é o parâmetro que se destaca claramente pelo

seu peso relativo. Desta forma, o elevado peso relativo da presença de pausa física é comum aos modelos TB-b2 e TB-b1 comentado na seção anterior.

O parâmetro presença de pausa física é seguido pelo reset de f0 entre as unidades pós-fronteiriça e fronteiriça, porém, o peso relativo do reset de f0 é muito menor que o peso do parâmetro de maior relevância hierárquica do modelo. Assim, apesar do modelo incluir parâmetros fonético-acústicos de f0 e um deles apresentar o segundo maior peso relativo, o parâmetro presença de pausa física se destaca muito claramente em virtude do seu peso relativo e pode, mais uma vez, ser o único parâmetro realmente considerado pelo modelo para reconhecer as fronteiras terminais.

O reset de f0 é seguido por mais dois parâmetros de f0. Os parâmetros de f0 são a diferença de inclinação média de f0 nas unidades pós-fronteiriça e fronteiriça e a inclinação média de f0 na unidade fronteiriça nesta ordem. Os parâmetros de f0 comuns aos modelos TB-b1 e TB-b2 compreendem o reset de f0 entre as unidades pós-fronteiriça e fronteiriça e a diferença de inclinação média da f0 nas unidades pós-fronteiriça e fronteiriça, além dos parâmetros de pausa física, como já foi dito. No caso da inclinação média de f0, o modelo TB-b1 inclui em sua composição a inclinação média de f0 na unidade pré-fronteiriça, já o modelo TB-b2 inclui em sua composição a inclinação média da f0 na unidade fronteiriça. O modelo TB-b1 inclui a diferença de taxa de produção de proeminência melódicas dos dois lados da fronteira. O modelo TB-b2, por sua vez, não possui nenhum parâmetro de duração normalizada.

O peso da duração da pausa física é o menor entre os cinco parâmetros do modelo TB-b2 e isso difere o modelo TB-b2 dos demais modelos destinados à identificação das fronteiras terminais. Enquanto a duração da pausa física apresenta o menor peso relativo do modelo TB-b2, nos outros modelos de identificação de fronteira terminal, este parâmetro é sempre o primeiro, segundo ou terceiro parâmetro mais relevante hierarquicamente e, na maioria das vezes, o peso relativo se destaca com relação ao peso dos outros parâmetros.

Em todas as amostras, o modelo TB-b2 identificou mais de 80% das fronteiras terminais. Na fase de desenvolvimento do modelo TB-b2, foram identificadas 82,6% das fronteiras terminais. Nas validações com as amostras não balanceadas, o modelo identificou entre 80,8% e 82,6% das fronteiras terminais. Tanto na fase de desenvolvimento, quanto na fase validação do modelo, pelo menos 90% das fronteiras devidamente identificadas eram acompanhadas por pausa física. No total, foram identificadas apenas 4 fronteiras terminais sem pausa física (três da amostra II e uma da amostra I), o que corresponde a 19% do número total de fronteiras terminais sem pausa física. As demais fronteiras terminais sem pausa física (17, no total) não foram identificadas pelo modelo TB-b2 em nenhum momento.

Outro elemento constante ao longo dos modelos de identificação de fronteira terminal é relativo à obtenção de falsos alarmes. Todos os outros modelos de identificação de fronteira terminal marcaram posições de NTB com pausa física como TB, gerando, assim, falsos alarmes em posição de NTB com pausa física. Alguns modelos TB realizaram a marcação indevida com maior frequência, mas isso ocorre em todos os modelos TB. No modelo TB-b2, foram obtidos entre 7,9% e 12,4% de falsos alarmes a depender da amostra, mas, em todas as amostras, pelo menos 94% dos falsos alarmes foram devidos às fronteiras não-terminais com pausa física.

No caso do modelo TB-b2, a quantidade de falsos alarmes devidos às fronteiras não-terminais acompanhadas por pausa física, o baixo reconhecimento das fronteiras terminais sem pausa física bem como a capacidade de identificar majoritariamente fronteiras com pausa física sugerem, aqui também, que a pausa física é superestimada no processo de reconhecimento automático das fronteiras terminais.

c. Modelo desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras terminais da amostra total

O modelo TB desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras terminais da amostra total é denominado TB-bfull. Abaixo, a tabela 32 apresenta o desempenho do modelo TB-bfull nas fases de desenvolvimento e validação nas amostras I, II e total não balanceadas.

Tabela 32: Desempenho do modelo TB-bfull nas fases de desenvolvimento e validações

Fase	Amostra	TB ident.	TB c/ pausa fís. ident.	TB s/ pausa fís. ident.	FA	FA c/ pausa fís.	FA s/ pausa fís.
Des	FULL-b (TB)	81,5% (97)	94,8% (92)	5,2% (5)	12,6% (15)	100% (15)	NA
Val-1	Total	81,5% (97)	94,8% (92)	5,2% (5)	9,2% (211)	94,8% (200)	5,2% (11)
Val-2	I	83,6% (61)	95% (58)	5% (3)	7,3% (91)	94,5% (86)	5,5% (5)
Val-3	II	78,3% (36)	94,4% (34)	5,6% (2)	11,4% (120)	95% (114)	5% (6)

A seguir, a tabela 33 apresenta a composição do modelo TB-bfull.

Tabela 33: Composição do modelo TB-bfull

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psp	Presença de pausa física após a fronteira	1,7343
2º	psdur	Duração da pausa física após a fronteira	1,2136

3º	f0meddloc	Reset de f0	0,1765
4º	df0medl10	Inclinação média de f0 na unidade pré-fronteiriça	0,1052
5º	df0medr1	Inclinação média de f0 na unidade pós-fronteiriça	0,0616
6º	zl10	Duração normalizada média na unidade pré-fronteiriça	0,0433

O modelo TB-bfull é composto por parâmetros de pausa física, f0 nas unidades mais próximas à fronteira e um parâmetro de duração normalizada em unidade pré-fronteiriça. No total, o modelo TB-bfull possui seis parâmetros. Neste sentido, o modelo é econômico em termos de quantidade de parâmetros fonético-acústicos analisados. Apesar da metade dos parâmetros do modelo TB-bfull serem parâmetros de f0, os parâmetros de pausa física se destacam com relação aos outros parâmetros do modelo em virtude do peso relativo deles, mas a presença de pausa física apresenta o peso relativo um pouco maior do que o peso relativo da duração da pausa física.

Após os dois parâmetros de pausa física, três parâmetros locais de f0 são os mais relevantes hierarquicamente com base no peso relativo. Em ordem de relevância hierárquica, os parâmetros são o reset de f0 entre as unidades pós-fronteiriça e fronteiriça, a inclinação média de f0 na unidade pré-fronteiriça e a inclinação média de f0 na unidade pós-fronteiriça. O peso relativo do reset de f0 entre as unidades pós-fronteiriça e fronteiriça, da inclinação média de f0 na unidade pré-fronteiriça e da inclinação média de f0 na unidade pós-fronteiriça é respectivamente 0,17, 0,10 e 0,06. O último parâmetro do modelo com o peso de 0,04 é a duração normalizada média da unidade pré-fronteiriça.

Tanto na fase de desenvolvimento, quanto nas validações, o modelo TB-bfull reconheceu pelo menos 78% das fronteiras terminais e mais de 90% das fronteiras devidamente identificadas eram acompanhadas por pausa física. O modelo identificou apenas 5 fronteiras terminais sem pausa física, o que equivale a 23% de reconhecimento sob o total de fronteiras terminais sem pausa física contabilizado na amostra total. Foram obtidos entre 7,3% e 12,6% de falsos alarmes a depender da amostra e pelo menos 94% dos falsos alarmes ocorreram em posição de NTB com pausa física.

No caso do modelo TB-bfull, a quantidade de falsos alarmes em posição de NTB com pausa física, o não reconhecimento das fronteiras terminais sem pausa física bem como a capacidade de identificar majoritariamente fronteiras com pausa física sugerem, mais uma vez, que os parâmetros relativos à pausa física são superestimados no processo de identificação automática de fronteira terminal.

4.1.3.2. Modelos de identificação de fronteiras não-terminais

Como já foi dito, o balanceamento dos dados em função da quantidade de NTB foi feito observando o número total de fronteiras não-terminais encontradas nas amostras I, II e total e selecionando aleatoriamente a quantidade equivalente de posições sem nenhuma fronteira (NB). Portanto, as fronteiras terminais não foram levadas em consideração na fase de desenvolvimento do modelo. Duas validações diferentes foram realizadas. Primeiramente, a validação foi feita aplicando os modelos desenvolvidos em todas as posições de NTB e NB das amostras I, II e total. Depois, os modelos foram aplicados em todas as posições de NTB, TB e NB das amostras I, II e total. Em ambos os casos de validação, as amostras usadas são amostras não balanceadas, visto que o número de exemplares de cada categoria era diferente.

As próximas seções apresentam os modelos de identificação de fronteiras não-terminais desenvolvidos a partir das amostras balanceadas em função da quantidade de NTB encontrada nas amostras I, II e total. Na seção 4.1.3.2, também são apresentados os resultados alcançados pelos modelos em discussão.

a. Modelo desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras não-terminais da amostra I

O modelo NTB desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras não-terminais da amostra I é denominado NTB-b1. Abaixo, a tabela 34 apresenta o desempenho do modelo NTB-b1 na fase de desenvolvimento.

Tabela 34: Desempenho do modelo NTB-b1 na fase de desenvolvimento

Fase	Amostra	NTB ident.	NTB c/ pausa fís. ident.	NTB s/ pausa fís. ident.	FA	FA s/ pausa fís.
Des	I-b (NTB)	70% (170)	54,2% (92)	45,8% (78)	23,4% (57)	100% (57)

Abaixo, a tabela 35 apresenta o desempenho do modelo NTB-b1 nas validações.

Tabela 35: Desempenho do modelo NTB-b1 nas fases de validação

Fase	Amostra	NTB ident.	NTB c/ pausa fís. ident.	NTB s/ pausa fís. ident.	FA	FA c/ pausa fís.	FA s/ pausa fís.
Val-1a	I	70% (170)	54,2% (92)	45,8% (78)	26,3% (259)	NA	100% (259)
Val-1b	II	71% (207)	58,9% (122)	41,1% (85)	35,5% (270)	NA	100% (270)
Val-1c	Total	70,3% (377)	56,8% (214)	43,2% (163)	30,3% (529)	NA	100% (529)
Val-2a	I	70% (170)	54,2% (92)	45,8% (78)	29,7% (314)	17,5% (55)	82,5% (259)
Val-2b	II	71% (207)	58,9% (122)	41,1% (85)	37,8% (305)	11,5% (35)	88,5% (270)
Val-2c	Total	70,3% (377)	56,8% (214)	43,2% (163)	33,2% (619)	14,5% (90)	85,5% (529)

A seguir, a tabela 36 apresenta a composição do modelo NTB-b1.

Tabela 36: Composição do modelo NTB-b1

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psp	Presença de pausa física após a fronteira	2,52829583
2º	psdur	Duração da pausa física após a fronteira	2,04480132
3º	ard	Diferença de taxa de articulação dos dois lados da fronteira	0,25644600
4º	zdloc	Diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça	0,22104138
5º	zl10	Duração normalizada média na unidade pré-fronteiriça	0,16382492
6º	zl0	Duração normalizada média na unidade fronteiriça	0,11697772
7º	df0med0	Inclinação média de f0 na unidade fronteiriça	0,10026581
8º	SDzd	Diferença de variação de duração normalizada média dos dois lados da fronteira	0,07818911
9º	srd	Diferença de taxa de elocução dos dois lados da fronteira	0,07686383
10º	df0meddloc	Diferença de inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça	0,05695446
11°	f0medl10	Mediana de f0 na unidade pré-fronteiriça	0,03560742

O modelo NTB-b1 é composto por onze parâmetros fonético-acústicos. Os parâmetros do modelo podem ser divididos em quatro grupos com base no peso relativo. Um primeiro grupo composto pelos parâmetros presença de pausa física e a duração da pausa física com os pesos relativo de 2,52 e 2,04 respectivamente. Um segundo grupo composto pelos parâmetros diferença de taxa de articulação dos dois lados da fronteira e diferença de duração normalizada entre as unidades pós-fronteiriça e fronteiriça com o peso relativo de 0,25 e 0,22 respectivamente. Um terceiro grupo incluindo os parâmetros duração normalizada média na unidade pré-fronteiriça, duração normalizada média na unidade fronteiriça e inclinação média de f0 na unidade fronteiriça com o peso relativo de 0,16, 0,11 e 0,10 respectivamente. Um último grupo composto pelos parâmetros diferença de variação normalizada média dos dois lados da fronteira, diferença de taxa de elocução dos dois lados da fronteira, diferença de

inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça e mediana de f0 na unidade pré-fronteiriça com o peso variando de 0,03 a 0,07.

Na fase de desenvolvimento, o modelo identificou 70% das fronteiras não-terminais das quais 54,2% são acompanhadas por pausa física e 45,8% não são acompanhadas por este parâmetro. Os falsos alarmes obtidos na fase de desenvolvimento do modelo correspondem a 23,4% da amostra balanceada utilizada e, evidentemente, eles ocorreram apenas em posição sem nenhuma fronteira, pois foram observadas apenas as posições de NTB e NB.

A porcentagem de fronteiras não-terminais reconhecidas se manteve estável nas aplicações do modelo nas diferentes amostras de dados não balanceados. Aproximadamente 70% das fronteiras não-terminais foram devidamente identificadas nas validações 1 e 2. Entre 54,2% e 58,9% das fronteiras identificadas são acompanhadas por pausa física. Portanto, entre 41,1% e 45,8% das fronteiras identificadas são fronteiras sem pausa física, o que pode ser considerado um ponto positivo do modelo em questão.

Os falsos alarmes obtidos nas validações 1 e 2 se diferem um pouco daqueles que foram obtidos na fase de desenvolvimento do modelo NTB-b1. Na validação 1, a porcentagem de falsos alarmes obtidos varia de 26,3% a 35,3%, a depender da amostra, e todos eles ocorreram em posição sem nenhuma fronteira, pois o modelo havia sido aplicado em todas as posições de NTB e NB das amostras I, II e total. Na validação 2, o cenário muda um pouco. A porcentagem de falsos alarmes obtidos varia de 29,7% a 38%, a depender da amostra, e os falsos alarmes ocorreram em posição sem nenhuma fronteira, mas também em posição de TB com pausa física, pois o modelo NTB-b1 foi aplicado em todas as posições de TB, NTB e NB das amostras I, II e total.

De modo geral, o modelo NTB-b1 apresentou uma boa capacidade para reconhecer NTB com e sem pausa física. Os resultados sugerem que o elevado peso relativo de um parâmetro de pausa física nem sempre impede a identificação de fronteiras sem pausa física. O problema do modelo NTB-b1 está relacionado aos falsos alarmes. Os falsos alarmes obtidos se dividem entre posições sem nenhuma fronteira e posições de fronteira terminal com pausa física. Os falsos alarmes em posição sem nenhuma fronteira ocorreram com maior frequência e isso pode ser considerado uma limitação do modelo. Em modelos destinados à identificação de NTB, é esperado obter uma maior porcentagem de falsos alarmes em posição de TB com pausa física do que em posição sem fronteira prosódica. Apesar da quantidade de falsos alarmes em posição sem nenhuma fronteira ser maior, mais de 90% das fronteiras terminais com pausa física das amostras I, II e total foram marcadas indevidamente como NTB. Desta forma, o modelo NTB-b1 também superestimou a importância da pausa física como recurso de identificação de NTB.

b. Modelo desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras não-terminais da amostra II

O modelo NTB desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras não-terminais da amostra II é denominado NTB-b2. Abaixo, a tabela 37 apresenta o desempenho do modelo NTB-b2 na fase de desenvolvimento.

Tabela 37: Desempenho do modelo NTB-b2 na fase de desenvolvimento

Fase	Amostra	NTB ident.	NTB c/ pausa fís. ident.	NTB s/ pausa fís. ident.	FA	FA s/ pausa fís.
Des	II-b (NTB)	68,5% (200)	57,5% (115)	42,5% (85)	17,5% (51)	100% (51)

A tabela 38 apresenta a seguir o desempenho do modelo NTB-b2 nas fases de validação.

Tabela 38: Desempenho do modelo NTB-b2 nas fases de validação

Fase	Amostra	NTB ident.	NTB c/ pausa fís. ident.	NTB s/ pausa fís. ident.	FA	FA c/ pausa fís.	FA s/ pausa fís.
Val-1a	I	75% (183)	49,2% (90)	50,8% (93)	38,2% (377)	NA	100% (377)
Val-1b	II	68,5% (200)	57,5% (115)	42,5% (85)	25,4% (193)	NA	100% (193)
Val-1c	Total	71,5% (383)	53,5% (205)	46,5% (178)	32,6% (570)	NA	100% (570)
Val-2a	I	75% (183)	49,2% (90)	50,8% (93)	40,8% (432)	12,7% (55)	87,3% (377)
Val-2b	II	68,5% (200)	57,5% (115)	42,5% (85)	28,3% (228)	15,4% (35)	84,6% (193)
Val-2c	Total	71,5% (383)	53,5% (205)	46,5% (178)	35,4% (660)	13,6% (90)	86,4% (570)

Abaixo, a tabela 39 apresenta os parâmetros fonético-acústicos que compõem o modelo NTB-b2.

Tabela 39: Composição do modelo NTB-b2

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	f0meddloc	Reset de f0	0,506241
2º	df0medr1	Inclinação média de f0 na unidade pré-fronteiriça	0,497678
3º	ard	Diferença de taxa de articulação dos dois lados da fronteira	0,243141
4º	zdloc	Diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça	0,189163
5º	zl0	Duração normalizada média na unidade fronteiriça	0,153507
6º	srd	Diferença de taxa de elocução dos dois lados da fronteira	0,143785
7º	zl9	Duração normalizada média na unidade anterior à unidade pré-fronteiriça (penúltima unidade antes da fronteira)	0,100575
8º	zl10	Duração normalizada média na unidade pré-fronteiriça	0,029064
9º	df0medl10	Inclinação média de f0 na unidade pré-fronteiriça	0,023524

10º	skzd	Diferenças das caudas de distribuição de duração normalizada média dos dois lados da fronteira	0,018309
11º	prd	Diferença de taxa de saliência duracional dos dois lados da fronteira	0,018035
12º	f0rd	Diferença de taxa de produção de proeminência melódicas dos dois lados da fronteira	0,001375
13º	df0meddloc	Diferença de inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça	0,00098

O modelo NTB-b2 é composto por 13 parâmetros fonético-acústicos, sendo que, desta vez, não há nenhum parâmetro relacionado à pausa física. Os parâmetros se organizam com base em uma escala mais gradual de peso relativo e podem ser divididos em três grupos, mas um dos parâmetros se afasta de todos os três grupos. O primeiro grupo se destaca com relação aos demais parâmetros e é composto pelos parâmetros reset de f0 e inclinação média de f0 na unidade pré-fronteiriça com o peso relativo de 0,50 e 0,49 respectivamente. O segundo grupo composto pelos parâmetros diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça, duração normalizada média na unidade fronteiriça e diferença de taxa de elocução dos dois lados da fronteira com o peso relativo variando de 0,14 a 0,18. O terceiro grupo composto pelos parâmetros duração normalizada média na unidade anterior à unidade pré-fronteiriça (penúltima unidade antes do final da palavra fonológica), duração normalizada média na unidade pré-fronteiriça, inclinação média de f0 na unidade pré-fronteiriça, diferenças das caudas de distribuição de duração normalizada média dos dois lados da fronteira, diferença de taxa de saliência duracional dos dois lados da fronteira, diferença de taxa de produção de proeminência melódicas dos dois lados da fronteira e diferença de inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça com o peso sempre abaixo de 0,1. A diferença de taxa de articulação dos dois lados da fronteira, por sua vez, tem o peso de 0,24 que não é alto o suficiente para ser incluído no primeiro grupo, mas também não é baixo o suficiente para ser incluído no segundo grupo.

Na fase de desenvolvimento, o modelo identificou 68,5% das fronteiras não-terminais. Aproximadamente 57,5% eram acompanhadas por pausa física e 42,5% não eram acompanhadas por pausa física. Os falsos alarmes obtidos correspondem 17,5% da amostra balanceada utilizada e, evidentemente, eles ocorreram apenas em posição sem nenhuma fronteira, pois, na fase de desenvolvimento, apenas as posições de NTB e NB foram observadas.

A porcentagem de fronteiras não-terminais reconhecidas se manteve estável nas aplicações do modelo nas diferentes amostras de dados não balanceados. Entre 68,5% e 75% das fronteiras não-terminais foram devidamente identificadas nas validações 1 e 2. Entre 49,2%

e 57,5% das fronteiras identificadas eram acompanhadas por pausa física. Portanto, entre 42,5% e 50,8% das fronteiras NTB identificadas não eram acompanhadas por pausa física.

Na validação 1, a depender da amostra, a porcentagem de falsos alarmes obtidos varia de 25,4% a 38,2%; todos os falsos alarmes ocorreram em posição sem nenhuma fronteira, pois o modelo havia sido aplicado apenas nas posições de NTB e NB das amostras I, II e total. Na validação 2, a porcentagem de falsos alarmes obtidos varia de 28,3% a 40,8%, a depender da amostra; os falsos alarmes ocorreram em posição sem nenhuma fronteira, mas também em posição de TB com pausa física, pois, nesta validação, o modelo foi aplicado em todas as posições de TB, NTB e NB das amostras I, II e total.

Em geral, o ponto positivo do modelo NTB-b2 está relacionado à capacidade de identificar uma quantidade expressiva de fronteiras não-terminais com e sem pausa física, nas fases de desenvolvimento e validações. O modelo teve dificuldade para distinguir as fronteiras não-terminais das fronteiras terminais com pausa física, frequentemente marcando TB com pausa física como NTB. O modelo também teve dificuldade para distinguir as fronteiras não-terminais das posições sem fronteira, marcando uma grande quantidade de NB como NTB. Os falsos alarmes obtidos ocorreram com maior frequência em posição sem nenhuma fronteira e essa marcação incorreta comprometeria como um todo uma ferramenta de identificação automática de fronteiras prosódicas. Quase todas as fronteiras terminais com pausa física geraram falsos alarmes e isso sugere que os parâmetros relativos à pausa física são superestimados até mesmo em um modelo que não inclui em sua composição parâmetros relativos à pausa física.

c. Modelo desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras não-terminais da amostra total

O modelo NTB desenvolvido a partir da amostra balanceada em função da quantidade de fronteiras não-terminais da amostra total é denominado NTB-bfull. Abaixo, a tabela 40 apresenta os resultados do modelo NTB-bfull na fase de desenvolvimento.

Tabela 40: Desempenho do modelo NTB-bfull na fase de desenvolvimento

Fase	Amostra	NTB ident.	NTB c/ pausa fís. ident.	NTB s/ pausa fís. ident.	FA	FA s/ pausa fís.
Des	FULL-b (NTB)	54,3% (291)	74,2% (216)	25,8% (75)	6,6% (33)	100% (33)

Abaixo, a tabela 41 apresenta o desempenho alcançado pelo modelo NTB-bfull nas fases de validação.

Tabela 41: Desempenho do modelo NTB-bfull nas fases de validação

Fase	Amostra	NTB ident.	NTB c/ pausa fís. ident.	NTB s/ pausa fís. ident.	FA	FA c/ pausa fís.	FA s/ pausa fís.
Val-1a	I	50,4% (123)	76,4% (94)	23,6% (29)	6,2% (61)	NA	100% (61)
Val-1b	II	57,5% (168)	72,6% (122)	27,4% (46)	7,8% (59)	NA	100% (59)
Val-1c	Total	54,3% (291)	74,2% (216)	25,8% (75)	6,9% (120)	NA	100% (120)
Val-2a	I	50,4% (123)	76,4% (94)	23,6% (29)	11% (116)	47,4% (55)	52,6% (61)
Val-2b	II	57,5% (168)	72,6% (122)	27,4% (46)	11,7% (94)	37,2% (35)	62,8% (59)
Val-2c	Total	54,3% (291)	74,2% (216)	25,8% (75)	11,3% (210)	42,9% (90)	57,1% (120)

A seguir, a tabela 42 apresenta a composição do modelo NTB-bfull.

Tabela 42: Composição do modelo NTB-bfull

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psp	Presença de pausa física após a fronteira	2,516793
2º	psdur	Duração da pausa física após a fronteira	1,816973
3º	df0medr1	Inclinação média de f0 na unidade pré-fronteiriça	0,514291
4º	f0medr1	Mediana de f0 na unidade pós-fronteiriça	0,41172
5º	f0med0	Mediana de f0 na unidade fronteiriça	0,377401
6º	ard	Diferença de taxa de articulação dos dois lados da fronteira	0,17864
7º	srd	Diferença de taxa de elocução dos dois lados da fronteira	0,151147
8º	zdloc	Diferença da duração normalizada nas unidades VV pós-fronteiriça e fronteiriça	0,139038
9º	f0meddloc	Reset de f0	0,120934
10º	zl10	Duração normalizada média na unidade pré-fronteiriça	0,11319
11º	zl0	Duração normalizada média na unidade fronteiriça	0,110301
12º	f0rd	Diferença de taxa de produção de proeminência melódicas dos dois lados da fronteira	0,067463
13º	df0medl10	Inclinação média de f0 na unidade pré-fronteiriça	0,048529
14º	sdf0d	Diferença de variação das medianas de f0 dos dois lados da fronteira	0,032558
15º	f0medd	Diferença de mediana de f0 dos dois lados da fronteira	0,02958
16º	f0medl10	Mediana de f0 na unidade pré-fronteiriça	0,025471

O modelo NTB-bfull é composto por 16 parâmetros fonético-acústicos. O modelo apresenta 9 parâmetros de f0, 3 de duração normalizada, 2 de pausa física, 1 de taxa de elocução e 1 de taxa de articulação. Os parâmetros podem ser divididos em três grupos com base no peso relativo. O primeiro grupo composto pelos parâmetros cujo peso relativo se destacam com relação a todos os outros. No primeiro grupo, são incluídos os parâmetros presença de pausa

física e duração da pausa física, mas há uma diferença de peso relativo entre eles (2,51 versus 1,81). O segundo grupo composto pelos parâmetros inclinação média de f0 na unidade pré-fronteiriça, mediana de f0 na unidade pós-fronteiriça e mediana de f0 na unidade fronteiriça com os pesos de 0,51, 0,41 e 0,37 respectivamente. O terceiro grupo composto pelos demais parâmetros do modelo com o peso relativo diminuindo lentamente e variando de 0,17 a 0,02 em ordem decrescente.

Na fase de desenvolvimento, o modelo identificou 54,3% das fronteiras não-terminais. Aproximadamente 74,2% eram acompanhadas por pausa física e 25,8% não eram acompanhadas por pausa física. Os falsos alarmes obtidos correspondem 6,6% da amostra balanceada destinada ao desenvolvimento do modelo e eles ocorreram apenas em posição sem nenhuma fronteira em razão das posições incluídas na amostra balanceada.

A porcentagem de fronteiras não-terminais reconhecidas se manteve estável nas validações 1 e 2. Entre 50,4% e 57,5% das fronteiras não-terminais foram devidamente identificadas nas validações 1 e 2. Entre 72,6% e 76,4% das fronteiras identificadas eram acompanhadas por pausa física. Portanto, entre 23,6% e 27,4% das fronteiras identificadas não eram acompanhadas por pausa física.

Na validação 1, a depender da amostra, a porcentagem de falsos alarmes obtidos varia de 6,2% a 7,8%; todos os falsos alarmes ocorreram em posição sem nenhuma fronteira, pois o modelo foi aplicado somente nas posições de NTB e NB das amostras I, II e total. Na validação 2, a porcentagem de falsos alarmes obtidos varia de 11% a 11,7%, a depender da amostra; os falsos alarmes ocorreram em posição sem nenhuma fronteira, mas também em posição de TB com pausa física, pois, nesta validação, o modelo foi aplicado em todas as posições de TB, NTB e NB das amostras I, II e total.

O modelo NTB-bfull pode ser considerado um modelo com desempenho intermediário, se considerarmos a porcentagem de fronteiras reconhecida nas fases de desenvolvimento e validações. Tanto na fase de desenvolvimento, quanto nas validações, a maioria das fronteiras reconhecidas é acompanhada por pausa física. O modelo teve dificuldade para distinguir as fronteiras não-terminais das posições sem fronteira, marcando NB como NTB, isto é, gerando falsos alarmes. O modelo também teve dificuldade para distinguir as fronteiras terminais com pausa física das fronteiras não-terminais, marcando TB com pausa física como NTB, isto é, gerando falsos alarmes.

Os falsos alarmes obtidos pelo modelo NTB-bfull ocorreram com maior frequência em posição sem nenhuma fronteira. Isso não significa que os falsos alarmes em posição de TB com pausa física são raros. Pelo contrário, praticamente todas as fronteiras terminais com pausa

física geraram falsos alarmes. Os falsos alarmes em posição sem fronteira são mais frequentes simplesmente porque os dados possuem uma maior quantidade de posição sem fronteira. Neste sentido, se 1000 fronteiras terminais com pausa física fossem analisadas pelo modelo, a quantidade de falsos alarmes seria altíssima. Quando comparado aos demais modelos NTB apresentados na seção 4.1.3.2, o modelo em discussão apresentou a menor porcentagem de falsos alarmes como um todo. De modo geral, a grande quantidade de falsos alarmes em posição de TB com pausa física junto com a capacidade de identificar majoritariamente fronteiras NTB com pausa física sugerem, mais uma vez, que os parâmetros relativos à pausa física são superestimados no processo de reconhecimento automático das fronteiras.

4.2. Modelos de identificação de fronteiras desenvolvidos a partir de parâmetros extraídos da janela de análise de 7 unidades

Foram desenvolvidos dois modelos diferentes para identificar automaticamente fronteiras terminais e não-terminais a partir de parâmetros fonético-acústicos extraídos da janela de análise composta por 7 unidades. Como já foi dito, a redução das janelas de análise influencia na quantidade de cálculo realizado pelo *script*, pois uma menor quantidade de medidas é extraída. O impacto de informações fonético-acústicas que podem causar ruído nas medidas globais extraídas também tende a ser reduzido, pois as medidas globais passam a se concentrar ainda mais nos fenômenos estabelecidos ao redor das posições em final de palavra fonológica.

Os modelos foram desenvolvidos a partir de duas amostras compostas por categorias completamente balanceadas entre si. O balanceamento foi realizado de modo que duas amostras, destinadas ao desenvolvimento dos modelos de identificação de TB e NTB, fossem geradas com um número equivalente de exemplares de cada categoria. Todo o processo de balanceamento das categorias foi feito aleatoriamente. A estratégia de balanceamento se difere um pouco da estratégia empregada anteriormente. A estratégia de balanceamento de categorias foi baseada em 70% das fronteiras terminais e não-terminais contabilizadas na amostra total a partir da janela de 7 unidades. No total, após a exclusão das disfluências, foram contabilizadas 120 fronteiras terminais e 546 fronteiras não-terminais na amostra total (vide tabela 14).

Para o desenvolvimento do modelo de identificação de fronteiras terminais a partir de parâmetros extraídos da janela de 7 unidades, 70% das fronteiras terminais contabilizadas na amostra total foram escolhidas aleatoriamente. A quantidade de fronteiras usadas no

desenvolvimento do modelo corresponde a 84. O balanceamento das categorias foi feito observando os 84 casos de fronteiras terminais escolhidos aleatoriamente e selecionando aleatoriamente uma quantidade equivalente de posições de ausência de fronteira terminal dividida proporcionalmente entre posições de fronteira não-terminal e posições sem nenhuma fronteira, isto é, 42 fronteiras não-terminais e 42 posições sem nenhuma fronteira. A amostra usada na fase de desenvolvimento do modelo de identificação de TB a partir de parâmetros extraídos da janela de 7 unidades é chamada de Full-7TB; o modelo desenvolvido, por sua vez, é chamado de TB-b7.

Para desenvolver o modelo de identificação de fronteiras não-terminais, foram escolhidas aleatoriamente 70% das fronteiras não-terminais contabilizadas na amostra total. A quantidade de fronteiras usadas na fase de desenvolvimento do modelo corresponde a 382. O balanceamento das categorias foi feito observando as 382 fronteiras não-terminais escolhidas aleatoriamente e selecionando aleatoriamente uma quantidade equivalente de posições sem fronteira prosódica. A amostra usada na fase de desenvolvimento do modelo de identificação de NTB a partir de parâmetros extraídos da janela de 7 unidades é chamada de Full-7NTB; o modelo desenvolvido, por sua vez, é chamado de NTB-b7.

Os modelos TB-b7 e NTB-b7 foram validados aplicando-os em amostras não balanceadas. A validação do modelo TB-b7 foi feita aplicando o modelo TB-b7 em todas as posições de TB, NTB e NB das amostras I, II e total. O modelo NTB-b7 foi validado de duas formas diferentes. Primeiramente, a validação foi feita aplicando o modelo NTB-b7 em todas as posições de NTB e NB das amostras I, II e total. Depois, o modelo NTB-b7 foi aplicado em todas as posições de NTB, TB e NB das amostras I, II e total. Tanto na validação do modelo TB-b7, quanto nas duas validações do modelo NTB-b7, as amostras usadas são não balanceadas, visto que o número de exemplares de cada categoria é diferente.

Os resultados relativos ao desenvolvimento de modelos a partir das amostras balanceadas bem como os resultados relativos à validação em dados não balanceados são expostos nas seções 4.2.1 e 4.2.2.

4.2.1. Modelo de identificação de fronteiras terminais

Abaixo, a tabela 43 apresenta o desempenho do modelo TB-b7 nas fases de desenvolvimento e validação.

Tabela 43: Desempenho do modelo TB-b7 nas fases de desenvolvimento e validação

Fase	Amostra	TB ident.	TB c/ pausa fís. ident.	TB s/ pausa fís. ident.	FA	FA c/ pausa fís.	FA s/ pausa fís.
Des	Full-7TB	85,7% (72)	88,9% (64)	11,1% (8)	6% (5)	100% (5)	NA
Val-1	I	82,2% (60)	91,7% (55)	8,3% (5)	7,9% (99)	53,5% (53)	46,5% (46)
Val-2	II	70,2% (33)	87,9% (29)	12,1% (4)	9,2% (98)	54% (53)	46% (45)
Val-3	Total	77,5% (93)	90,3% (84)	9,7% (9)	8,5% (197)	53,8% (106)	46,2% (91)

A seguir, a tabela 44 apresenta a composição do modelo TB-b7.

Tabela 44: Composição do modelo TB-b7

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psdur	Duração da pausa física após a fronteira	3.795666
2º	f0meddloc	Reset de f0	0.883585
3º	df0medr1	Inclinação média da f0 na 1ª unidade da janela à direita da fronteira	0.833643
4º	df0medl	Inclinação média de f0 dentro da janela à esquerda da fronteira	0.276789
5º	sdf0l	Variação nas medianas da f0 na janela à esquerda da fronteira	0.189375
6º	f0rl	Taxa de produção de proeminência melódicas, associadas ao acento de *pitch* na janela à esquerda da fronteira	0.151088
7º	df0medl7	Inclinação média de f0 na unidade pré-fronteiriça	0.114877
8º	df0meddloc	Diferença de inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça	0.101577
9º	zl7	Duração normalizada média na unidade pré-fronteiriça	0.079633
10º	f0medd	Diferença entre medianas da F0 nas janelas dos dois lados da fronteira	0.070854
11º	mzd	Diferença de duração normalizada média dos dois lados da fronteira	0.024620
12°	zdloc	Diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça	0.018060

O modelo TB-b7 é composto por 12 parâmetros fonético-acústicos. Os parâmetros do modelo em discussão podem ser divididos em quatro grupos com base no peso relativo. O primeiro grupo composto apenas pela duração da pausa física, que apresentou um peso relativo bem mais alto que o peso dos demais parâmetros. O segundo grupo composto pelo reset de f0 e inclinação média da F0 na 1ª unidade da janela à direita da fronteira com os pesos de 0,88 e 0,83 respectivamente. O terceiro composto pela inclinação média de f0 dentro da janela à esquerda da fronteira, variação nas medianas da F0 na janela à esquerda da fronteira, taxa de produção de proeminência melódicas na janela à esquerda da fronteira, inclinação média de f0 na unidade pré-fronteiriça e diferença de inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça com o peso variando de 0,27 a 0,1. O quarto grupo é composto pelos demais parâmetros do modelo e o peso varia de 0,07 a 0,01

Na fase de desenvolvimento do modelo, 85,7% das fronteiras terminais foram identificadas. Aproximadamente 89% das fronteiras devidamente reconhecidas eram acompanhadas por pausa física e 11% não eram acompanhadas por pausa física. Os falsos alarmes obtidos correspondem a 6% da amostra balanceada destinada ao desenvolvimento do modelo e todos eles ocorreram em posição de fronteira NTB com pausa física.

Na validação com as amostras não balanceadas, entre 70,2% e 82,2% das fronteiras terminais foram devidamente identificadas e mais de 87% das fronteiras identificadas eram acompanhadas por pausa física. A porcentagem de reconhecimento das fronteiras terminais sem pausa física foi sempre menor que 13%. Mesmo a porcentagem de reconhecimento de TB sem pausa física sendo relativamente baixa e indicar uma certa dificuldade para reconhecer estas fronteiras, é a maior porcentagem encontrada entre os modelos TB. Nas validações, a porcentagem de falsos alarmes obtidos variou de 8,3% a 12,1% a depender da amostra. Os falsos alarmes obtidos nas validações ocorreram de forma relativamente equilibrada em posição sem fronteira e em posição de NTB com pausa física.

A dificuldade de reconhecer TB sem pausa física bem como a capacidade de identificar majoritariamente fronteiras TB com pausa física são características comuns ao modelo TB-b7 e demais modelos TB já apresentados. A diferença do modelo TB-b7 com relação aos demais modelos TB está relacionada aos falsos alarmes obtidos nas validações. Entre todos os modelos TB desenvolvidos, o modelo TB-b7 apresentou a menor porcentagem de falsos alarmes em posição de NTB com pausa física. Neste sentido, o modelo em discussão apresentou a melhor capacidade para distinguir NTB e TB com pausa física, ainda que essa distinção não seja absoluta, pois a quantidade de falsos alarmes em posição de NTB com pausa física ainda pode ser considerada expressiva.

A dificuldade para reconhecer TB sem pausa física, a capacidade de identificar majoritariamente fronteiras TB com pausa física e os falsos alarmes em posição de NTB com pausa física continuam sendo indicativos da superestimação dos parâmetros de pausa física. A redução da extensão da janela de análise para sete unidades possibilitou desenvolver um modelo TB que contém apenas um parâmetro de pausa física. Em geral, o modelo TB-b7 possui uma melhor capacidade de identificar as fronteiras terminais sem pausa física, mas, em contrapartida, há um aumento de falsos alarmes em posição sem nenhuma fronteira. Assim, os falsos alarmes em posição sem nenhuma fronteira podem ser vistos como uma consequência do maior reconhecimento das fronteiras terminais sem pausa física.

4.2.2. Modelo de identificação de fronteiras não-terminais

A seguir, a tabela 45 apresenta o desempenho do modelo NTB-b7 na fase de desenvolvimento.

Tabela 45: Desempenho do modelo NTB-b7 na fase de desenvolvimento

Fase	Amostra	NTB ident.	NTB c/ pausa fís. ident.	NTB s/ pausa fís. ident.	FA	FA s/ pausa fís.
Des	Full-7NTB	51% (194)	93,3% (181)	6,7% (13)	4% (15)	100% (15)

Abaixo, a tabela 46 apresenta o desempenho alcançado pelo modelo NTB-b3 nas fases de validação.

Tabela 46: Desempenho do modelo NTB-b7 nas fases de validação

Fase	Amostra	NTB ident.	NTB c/ pausa fís. ident.	NTB s/ pausa fís. ident.	FA	FA c/ pausa fís.	FA s/ pausa fís.
Val-1a	I	49,6% (124)	92,7% (115)	7,3% (9)	5,7% (57)	NA	100% (57)
Val-1b	II	45,6% (135)	94,8% (128)	5,2% (7)	3,1% (24)	NA	100% (24)
Val-1c	Total	47,4% (259)	93,8% (243)	6,2% (16)	4,6% (81)	NA	100% (81)
Val-2a	I	49,6% (124)	92,7% (115)	7,3% (9)	10,9% (117)	51,3% (60)	48,7% (57)
Val-2b	II	45,6% (135)	94,8% (128)	5,2% (7)	7,5% (61)	60,7% (37)	39,3% (24)
Val-2c	Total	47,4% (259)	93,8% (243)	6,2% (16)	9,4% (178)	54,5% (97)	45,5% (81)

A tabela 47 a seguir apresenta a composição do modelo NTB-b7.

Tabela 47: Composição do modelo NTB-b7

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psdur	Duração da pausa física após a fronteira	4.059809
2º	zdloc	Diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça	0.141812
3º	zl7	Duração normalizada média na unidade pré-fronteiriça	0.115395
4º	zl0	Duração normalizada média na unidade fronteiriça	0.079704
5º	srd	Diferença de taxa de elocução dos dois lados da fronteira	0.071432
6º	skf0d	Diferença das caudas da distribuição de f0 dos dois lados da fronteira	0.043780
7º	f0medd	Diferença entre medianas da f0 nas janelas dos dois lados da fronteira	0.021719

O modelo NTB-b7 é composto por 7 parâmetros fonético-acústicos. Os parâmetros podem ser divididos em dois grupos com base no peso relativo. O primeiro grupo composto

apenas pela duração da pausa física com o peso de 4,05. Os parâmetros do segundo grupo compreendem parâmetros de duração normalizada nas unidades mais próximas ao final da palavra fonológica, a diferença de taxa de elocução e dois parâmetros globais de diferença de f0. O peso relativo dos parâmetros do segundo grupo diminui gradualmente e varia de 0,14 a 0,02.

Na fase de desenvolvimento, o modelo identificou 51% das fronteiras não-terminais. Aproximadamente 93,3% eram acompanhadas por pausa física e 6,7% não eram acompanhadas por pausa física. Os falsos alarmes obtidos correspondem a 4% das posições da amostra balanceada destinada ao desenvolvimento do modelo e todos ocorreram em posição sem fronteira prosódica em razão das posições incluídas na amostra de desenvolvimento.

Nas validações 1 e 2, entre 45,6% e 49,6% das fronteiras não-terminais foram devidamente identificadas e mais de 92% das fronteiras identificadas eram acompanhadas por pausa física. Portanto, menos de 8% das fronteiras NTB identificadas eram sem pausa física. Na validação 1, a depender da amostra, a porcentagem de falsos alarmes obtidos foi de 3,1% a 5,7%; todos os falsos alarmes ocorreram em posição sem nenhuma fronteira, pois a validação 1 avaliou apenas as posições de NTB e NB das amostras I, II e total. Na validação 2, a porcentagem de falsos alarmes obtidos varia de 7,5% a 10,9%, a depender da amostra; os falsos alarmes ocorreram principalmente em posição de TB com pausa física, mas também há falsos alarmes em posição sem nenhuma fronteira, pois, na validação 2, todas as posições de TB, NTB e NB das amostras I, II e total foram analisadas.

O modelo NTB-b7 pode ser considerado um modelo com desempenho intermediário no que diz respeito à porcentagem de fronteiras devidamente reconhecida nas fases de desenvolvimento e validações. Tanto na fase de desenvolvimento, quanto nas validações, a maioria das fronteiras reconhecidas é acompanhada por pausa física. O modelo teve dificuldade para distinguir as fronteiras não-terminais das posições sem fronteira, marcando NB como NTB nas fases de desenvolvimento e validações, isto é, gerando falsos alarmes. Na validação 2, o modelo também teve dificuldade para diferenciar as fronteiras terminais com pausa física das fronteiras não-terminais com pausa física, marcando TB com pausa física como NTB indevidamente. De modo geral, o conjunto de resultados do modelo NTB-b7 sugere, mais uma vez, que os parâmetros relativos à pausa física são superestimados no processo de reconhecimento automático das fronteiras, pois os falsos alarmes ocorreram principalmente em posição de TB com pausa física e a maioria das fronteiras NTB identificadas era acompanhada por pausa física.

4.3. Modelos de identificação de fronteiras desenvolvidos a partir de parâmetros extraídos da janela de análise de 3 unidades

Foram desenvolvidos dois modelos diferentes para identificar automaticamente fronteiras terminais e não-terminais a partir de parâmetros fonético-acústicos extraídos da janela de análise composta por 3 unidades. Os modelos foram desenvolvidos a partir de duas amostras compostas por categorias completamente balanceadas entre si. O balanceamento foi realizado de modo que duas amostras, destinadas ao desenvolvimento dos modelos de identificação de TB e NTB, fossem geradas com um número equivalente de exemplares de cada categoria. Todo o processo de balanceamento das categorias foi feito aleatoriamente. A estratégia de balanceamento de categorias foi baseada em 70% das fronteiras terminais e não-terminais contabilizadas na amostra total a partir da janela de 3 unidades. No total, após a exclusão das disfluências, foram contabilizadas 123 fronteiras terminais e 560 fronteiras não-terminais na amostra total (vide tabela 15).

Para o desenvolvimento do modelo de identificação de fronteiras terminais a partir de parâmetros extraídos da janela de 3 unidades, foram escolhidas aleatoriamente 70% das fronteiras terminais contabilizadas na amostra total. A quantidade de fronteiras usadas no desenvolvimento do modelo corresponde a 86. O balanceamento das categorias foi feito observando as 86 fronteiras terminais escolhidas aleatoriamente e selecionando aleatoriamente uma quantidade equivalente de posições de ausência de fronteira terminal dividida proporcionalmente entre posições de fronteira não-terminal e posições sem nenhuma fronteira, isto é, 43 fronteiras não-terminais e 43 posições sem nenhuma fronteira. A amostra usada na fase de desenvolvimento do modelo de identificação de TB a partir de parâmetros extraídos da janela de 3 unidades é chamada de Full-3TB; o modelo desenvolvido é chamado de TB-b3.

Para desenvolver o modelo de identificação de fronteiras não-terminais, foram escolhidas aleatoriamente 70% das fronteiras não-terminais contabilizadas na amostra total. A quantidade de fronteiras usadas na fase de desenvolvimento do modelo é 392. O balanceamento das categorias foi feito observando as 392 fronteiras não-terminais escolhidas aleatoriamente e selecionando aleatoriamente uma quantidade equivalente de posições sem fronteira prosódica. A amostra usada na fase de desenvolvimento do modelo de identificação de NTB a partir de parâmetros extraídos da janela de 3 unidades é chamada de Full-3NTB; o modelo desenvolvido, por sua vez, é chamado de NTB-b3.

Os modelos TB-b3 e NTB-b3 foram validados aplicando-os em amostras não balanceadas. A validação do modelo TB-b3 foi feita aplicando o modelo TB-b3 em todas as

posições de TB, NTB e NB das amostras I, II e total. O modelo NTB-b3 foi validado de duas formas diferentes. Primeiramente, a validação foi feita aplicando o modelo NTB-b3 em todas as posições de NTB e NB das amostras I, II e total. Depois, o modelo NTB-b3 foi aplicado em todas as posições de NTB, TB e NB das amostras I, II e total. Tanto na validação do modelo TB-b3, quanto nas diferentes validações do modelo NTB-b3, as amostras usadas são não balanceadas, visto que o número de exemplares de cada categoria é diferente.

Os resultados relativos ao desenvolvimento de modelos a partir das amostras balanceadas bem como os resultados relativos às aplicações em dados não balanceados são expostos nas seções 4.3.1 e 4.3.2.

4.3.1. Modelo de identificação de fronteiras terminais

Abaixo, a tabela 48 apresenta o desempenho do modelo TB-b3 nas fases de desenvolvimento e validação.

Tabela 48: Desempenho do modelo TB-b3 nas fases de desenvolvimento e validação

Fase	Amostra	TB ident.	TB c/ pausa fís. ident.	TB s/ pausa fís. ident.	FA	FA c/ pausa fís.	FA s/ pausa fís.
Des	Full-3TB	77,9% (67)	92,5% (62)	7,5% (5)	12,8% (11)	90,9% (10)	9,1% (1)
Val-1	I	81,6% (62)	96,8% (60)	3,2% (2)	7,2% (93)	72% (67)	28% (26)
Val-2	II	72,3% (34)	91,2% (31)	8,8% (3)	9% (99)	72,7% (72)	27,3% (27)
Val-3	Total	78% (96)	94,8% (91)	5,2% (5)	8% (192)	72,4% (139)	27,6% (53)

A seguir, a tabela 49 apresenta a composição do modelo TB-b3.

Tabela 49: Composição do modelo TB-b3

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psdur	Duração da pausa física após a fronteira	4,117467
2º	f0meddloc	Reset de f0	0,151306
3º	zl3	Duração normalizada média na unidade pré-fronteiriça	0,102239
4º	df0medl3	Inclinação média de f0 na unidade pré-fronteiriça	0,081472
5º	prd	Diferença de taxa de saliência duracional dos dois lados da fronteira	0,055019
6º	ard	Diferença de taxa de articulação dos dois lados da fronteira	0,016022

O modelo TB-b3 é econômico no que diz respeito à quantidade de parâmetros fonético-acústicos observados. O parâmetro com maior peso relativo é a duração da pausa física. A

duração da pausa física é seguida pelos parâmetros reset de f0 entre as unidades pós-fronteiriça e fronteiriça, duração normalizada média na unidade pré-fronteiriça, inclinação média de f0 na unidade pré-fronteiriça, diferença de taxa de saliência duracional dos dois lados da fronteira e diferença de taxa de articulação dos dois lados da fronteira nesta ordem. Com exceção da duração da pausa física, o peso do reset de f0 se destacou um pouco com relação aos demais parâmetros, mas não foi tanto quanto a duração da pausa física. O peso relativo dos quatro últimos parâmetros é bem menor e diminui gradualmente de 0,10 a 0,01.

Na fase de desenvolvimento do modelo, 77,9% das fronteiras terminais foram identificadas. Aproximadamente 93% das fronteiras devidamente reconhecidas eram acompanhadas por pausa física e 7% não eram acompanhadas por pausa física. Os falsos alarmes obtidos correspondem a 12,8% da amostra balanceada destinada ao desenvolvimento do modelo e a maioria deles ocorreram em posição de fronteira não-terminal com pausa física.

Na validação com as amostras não balanceadas, entre 72,3% e 81,6% das fronteiras terminais foram devidamente identificadas e mais de 90% das fronteiras identificadas eram acompanhadas por pausa física. Portanto, a porcentagem de reconhecimento de fronteiras terminais sem pausa física foi sempre menor que 10%. Nas validações, a porcentagem de falsos alarmes obtidos variou de 7,2% a 9% a depender da amostra. Nas validações, os falsos alarmes ocorreram principalmente em posição de NTB acompanhada por pausa física.

Em alguns aspectos, o modelo TB-b3 se parece muito com outros modelos de identificação de fronteira terminal já apresentados. A quantidade de falsos alarmes em posição de NTB com pausa física, o baixo reconhecimento das fronteiras terminais sem pausa física bem como a capacidade de identificar majoritariamente fronteiras com pausa física sugerem que, neste modelo também, os parâmetros relativos à pausa física são superestimados no processo de identificação automática de fronteira terminal. A diferença do modelo TB-b3 com relação à maioria dos demais modelos TB está relacionada aos falsos alarmes obtidos nas validações. Com exceção do modelo TB-b7, nas validações de todos os outros modelos TB já apresentados, os falsos alarmes em posição de NTB com pausa física correspondem a pelo menos 90% dos casos. Nas validações do modelo TB-b3, a porcentagem corresponde a aproximadamente 70% dos casos. Desta forma, o modelo em discussão apresentou uma melhor capacidade para distinguir NTB e TB com pausa física. Porém, ainda assim, a quantidade de marcações erradas em posição de NTB com pausa física é alta.

A redução da extensão da janela de análise para três unidades possibilitou desenvolver um segundo modelo TB que contém apenas um parâmetro de pausa física. Esta mudança na composição de parâmetros aumentou a porcentagem de fronteiras terminais sem pausa física

identificada, mas, em contrapartida, houve um aumento de falsos alarmes em posição sem nenhuma fronteira. Portanto, assim como no modelo TB-b7, os falsos alarmes em posição sem nenhuma fronteira podem ser vistos como uma consequência do maior reconhecimento das fronteiras terminais sem pausa física.

4.3.2. Modelo de identificação de fronteiras não-terminais

Abaixo, a tabela 50 apresenta o desempenho do modelo NTB-b3 na fase de desenvolvimento.

Tabela 50: Desempenho do modelo NTB-b3 na fase de desenvolvimento

Fase	Amostra	NTB ident.	NTB c/ pausa fís. ident.	NTB s/ pausa fís. ident.	FA	FA s/ pausa fís.
Des	Full-3NTB	54% (212)	83% (176)	17% (36)	7,1% (28)	100% (28)

Abaixo, a tabela 51 apresenta o desempenho alcançado pelo modelo NTB-b3 nas fases de validação.

Tabela 51: Desempenho do modelo NTB-b3 nas fases de validação

Fase	Amostra	NTB ident.	NTB c/ pausa fís. ident.	NTB s/ pausa fís. ident.	FA	FA c/ pausa fís.	FA s/ pausa fís.
Val-1a	I	57% (147)	77% (113)	23% (34)	11,5% (118)	NA	100% (118)
Val-1b	II	52% (157)	81,5% (128)	18,5% (29)	8,1% (65)	NA	100% (65)
Val-1c	Total	54,3% (304)	79,3% (241)	20,7% (63)	10% (183)	NA	100% (183)
Val-2a	I	57% (147)	77% (113)	23% (34)	16,3% (180)	34% (62)	66% (118)
Val-2b	II	52% (157)	81,5% (128)	18,5% (29)	12% (102)	36,3% (37)	63,7% (65)
Val-2c	Total	54,3% (304)	79,3% (241)	20,7% (63)	14,4% (282)	35,1% (99)	64,9% (183)

A seguir, a tabela 52 apresenta a composição do modelo NTB-b3.

Tabela 52: Composição do modelo NTB-b3

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psdur	Duração da pausa física após a fronteira	4,034147
2º	prd	Diferença de taxa de saliência duracional dos dois lados da fronteira	0,157067
3º	zdloc	Diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça	0,127506

4º	srd	Diferença de taxa de elocução dos dois lados da fronteira	0,072652
5º	SDzd	Diferença de variação de duração normalizada média dos dois lados da fronteira	0,049456
6º	zl3	Duração normalizada média na unidade pré-fronteiriça	0,041732
7º	mzd	Diferença de duração normalizada média dos dois lados da fronteira	0,039886
8º	ard	Diferença de taxa de articulação dos dois lados da fronteira	0,034892
9º	f0meddloc	Reset de f0	0,016706

O modelo NTB-b3 é composto por 9 parâmetros fonético-acústicos. Os parâmetros podem ser divididos em três grupos com base no peso relativo. O primeiro grupo composto pela duração da pausa física com o peso de 4,03. O segundo grupo composto pela diferença de taxa de saliência duracional dos dois lados da fronteira e diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça com os pesos de 0,15 e 0,12 respectivamente. O terceiro grupo composto pelos demais parâmetros do modelo com o peso relativo diminuindo lentamente e variando de 0,07 a 0,01 em ordem decrescente.

Na fase de desenvolvimento, o modelo identificou 54% das fronteiras não-terminais. Aproximadamente 83% eram acompanhadas por pausa física e 17% não eram acompanhadas por pausa física. Os falsos alarmes obtidos correspondem 7,1% da amostra balanceada destinada ao desenvolvimento do modelo e todos ocorreram em posição sem nenhuma fronteira em razão das posições incluídas na amostra de desenvolvimento.

Nas validações 1 e 2, entre 54,3% e 57% das fronteiras não-terminais foram devidamente identificadas. Dentre as fronteiras identificadas, entre 77% e 81,5% das fronteiras identificadas eram acompanhadas por pausa física. Portanto, entre 18,5% e 23% das fronteiras identificadas não eram acompanhadas por pausa física.

Na validação 1, a depender da amostra, a porcentagem de falsos alarmes obtidos varia de 8,1% a 11,5%; todos os falsos alarmes ocorreram em posição sem nenhuma fronteira, pois a validação 1 observou apenas as posições de NTB e NB das amostras I, II e total. Na validação 2, a porcentagem de falsos alarmes obtidos varia de 12% a 16,3%, a depender da amostra; os falsos alarmes ocorreram em posição sem nenhuma fronteira, mas também em posição de TB com pausa física, pois a validação 2 considerou todas as posições de TB, NTB e NB das amostras I, II e total.

O modelo NTB-b3 pode ser considerado um modelo com desempenho intermediário no que diz respeito à porcentagem de NTB reconhecida nas fases de desenvolvimento e validações. Tanto na fase de desenvolvimento, quanto nas validações, a maioria das fronteiras reconhecidas é acompanhada por pausa física. O modelo teve dificuldade para distinguir as fronteiras não-terminais das posições sem fronteira, marcando NB como NTB nas fases de desenvolvimento

e validações, isto é, gerando falsos alarmes. O modelo também teve dificuldade para distinguir as fronteiras terminais com pausa física das fronteiras não-terminais, marcando TB com pausa física como NTB na validação 2, isto é, gerando falsos alarmes.

De modo geral, tanto nas fases de desenvolvimento, quanto nas fases de validação, os falsos alarmes do modelo NTB-b3 se concentraram principalmente em posição sem nenhuma fronteira, o que pode ser visto como uma limitação. Porém, é importante destacar que praticamente todas as fronteiras terminais com pausa física geraram falsos alarmes na validação 2. Desta forma, mais uma vez, a quantidade de falsos alarmes em posição de TB com pausa física e a capacidade de identificar majoritariamente fronteiras NTB com pausa física indicam que os parâmetros relativos à pausa física são superestimados.

4.4. Resumo do capítulo

No capítulo 4, primeiramente, os modelos TB e NTB anteriores foram revalidados. Em seguida, outros modelos TB e NTB foram desenvolvidos e validados a partir de diferentes estratégias. Os modelos foram desenvolvidos a partir de parâmetros extraídos de janelas de análise de diferentes extensões e amostras compostas por categorias balanceadas entre si. Após o processo de treinamento, foi realizada a validação dos modelos em dados sem nenhum tipo de balanceamento de categorias.

Foram destacados os resultados alcançados pelos modelos, os parâmetros e as características dos modelos. A principal característica comum aos modelos TB e NTB é a importância do parâmetro pausa física no processo de identificação das fronteiras. Em razão desta característica, os modelos TB reconheceram sobretudo as fronteiras terminais com pausa física, deixaram de identificar aquelas que não apresentam pausa física e ainda obtiveram muitos falsos alarmes em posição de NTB com pausa física. Os modelos NTB apresentam um problema semelhante, isto é, identificaram sobretudo as fronteiras não-terminais com pausa física, deixaram de identificar as fronteiras sem pausa física e geraram diversos falsos alarmes em posição de TB com pausa física.

5. Procedimento de investigação da percepção de pausa

Com objetivo de compreender mais sobre a percepção de pausa e buscar reduzir a importância da pausa física nos modelos de identificação automática de fronteira, neste

trabalho, foi realizada uma investigação sobre a percepção de pausa em posição de fronteira por parte de um grupo de anotadores. A investigação sobre a pausa percebida parte da premissa de que ela deve ser definida do ponto de vista da percepção dos ouvintes, pois se trata de um fenômeno perceptual. Todos os detalhes do procedimento de investigação sobre a percepção de pausa são apresentados ao longo do capítulo 5.

Na seção 5.1, são apresentados os dados analisados na tarefa de percepção de pausa. Na seção 5.2, são apresentados os anotadores que realizaram a tarefa. Na seção 5.3, as instruções fornecidas aos anotadores são apresentadas. Na seção 5.4, são apresentados os acordos relativos à percepção de pausa em posição de fronteira prosódica. Na seção 5.5, é apresentada a análise da duração mínima de pausa física percebida como pausa pelos ouvintes. Na seção 5.6, é apresentada a análise dos fatores potencialmente responsáveis pela percepção de pausa.

5.1. Dados

Os dados analisados na tarefa de percepção de pausa compreendem todas as fronteiras terminais e não-terminais percebidas como fronteiras do mesmo tipo por pelo menos 50% dos segmentadores na tarefa de percepção de fronteiras das amostras I e II. Os dados também compreendem as fronteiras TB e NTB atribuídas por um segmentador *expert* nos casos de empate entre segmentadores. O segmentador *expert* foi escolhido com base no cálculo do coeficiente Kappa (Fleiss, 1971) que comparou a anotação individual de cada segmentador com a anotação que é fruto do acordo de pelo menos 50% entre os segmentadores. Por questões operacionais relacionadas à disponibilidade dos segmentadores, foram avaliados os segmentadores que realizaram a segmentação da amostra II. Portanto, os 19 segmentadores da amostra II eram potenciais candidatos à função de segmentador *expert*. Os segmentadores foram chamados com números de 1 a 19, de acordo com o coeficiente Kappa de cada um deles. Abaixo, a tabela 53 mostra o coeficiente Kappa relativo à concordância entre a anotação de cada segmentador e a anotação fruto do acordo de pelo menos 50% entre segmentadores.

Tabela 53: Resultados do Kappa de concordância entre a anotação de cada segmentador e a anotação fruto do acordo de pelo menos 50% entre segmentadores

Segmentador	Geral	TB	Disfluência	NTB	Posição sem nenhuma fronteira
Segmentador 1	0,92	0,86	0,87	0,92	0,94
Segmentador 2	0,91	0,80	0,94	0,89	0,94
Segmentador 3	0,91	0,87	0,96	0,88	0,93

Segmentador 4	0,91	0,83	0,95	0,88	0,93
Segmentador 5	0,89	0,82	0,91	0,88	0,92
Segmentador 6	0,89	0,88	0,91	0,87	0,92
Segmentador 7	0,89	0,88	0,92	0,86	0,91
Segmentador 8	0,88	0,87	0,91	0,86	0,89
Segmentador 9	0,88	0,83	0,96	0,83	0,90
Segmentador 10	0,87	0,77	0,96	0,83	0,91
Segmentador 11	0,87	0,68	0,99	0,83	0,90
Segmentador 12	0,87	0,68	0,99	0,83	0,90
Segmentador 13	0,87	0,82	0,92	0,83	0,89
Segmentador 14	0,85	0,80	0,81	0,82	0,89
Segmentador 15	0,84	0,75	0,95	0,79	0,88
Segmentador 16	0,80	0,80	0,84	0,76	0,81
Segmentador 17	0,78	0,87	0,34	0,73	0,88
Segmentador 18	0,76	0,80	0,30	0,70	0,89
Segmentador 19	0,75	0,71	0,87	0,67	0,79

Conforme mostra a tabela 53, o acordo entre a anotação do segmentador 1 e a anotação fruto do acordo de pelo menos 50% entre os 19 segmentadores é 0,92. Por se tratar do maior valor de acordo encontrado, foi estabelecido que o segmentador 1 seria o segmentador *expert* e decidiria a melhor categoria para as posições de empate. No total, considerando os textos das amostras I e II, o segmentador *expert* decidiu a categoria perceptual de 19 posições.

Na tarefa de percepção de pausa, além das fronteiras percebidas como fronteiras do mesmo tipo por pelo menos 50% dos segmentadores e fronteiras atribuídas pelo segmentador *expert* quando necessário, são analisadas também as disfluências definidas por um segmentador na amostra I e aquelas que foram percebidas por pelo menos 50% dos segmentadores na amostra II. Isso significa que a quantidade total de posições analisadas na tarefa de percepção de pausa difere-se da quantidade total de posições analisada no desenvolvimento dos modelos. O principal fator que explica esta diferença é a exclusão das disfluências no processo de desenvolvimento e validação dos modelos. Além disso, a diferença é devida à eventual eliminação de fronteiras a depender da extensão da janela de análise do *script* que extrai os parâmetros fonético-acústicos em posição de fronteira e em seu entorno (Barbosa, 2016). Apesar da diferença de fato existir, é importante ressaltar que o total de posições usadas no desenvolvimento dos modelos não se difere muito do total de posições avaliadas no teste de percepção de pausa.

Com todas as fronteiras retratadas ao longo desta subseção, obtém-se, então, uma segmentação “ideal” para os áudios das amostras I e II. A segmentação ideal compreende as posições de TB e NTB percebidas por pelo menos 50% dos segmentadores nas amostras I e II,

as disfluências atribuídas por um segmentador na amostra I, as disfluências percebidas por pelo menos 50% dos segmentadores na amostra II e as decisões tomadas pelo segmentador *expert*, quando necessário. Esta segmentação ideal é a segmentação utilizada na tarefa de percepção de pausa em posição de fronteira prosódica; mais detalhes sobre a tarefa de percepção de pausa são apresentados nas próximas subseções. Abaixo, a tabela 54 apresenta a distribuição de posições analisadas na tarefa de percepção de pausa.

Tabela 54: Distribuição de posições analisadas na tarefa de percepção de pausa

Posições	Total	Amostra	Frequência
Terminal	141	I	86
		II	55
Não-terminal	570	I	263
		II	307
Disfluência	126	I	53
		II	73
Sem nenhuma fronteira	1834	I	931
		II	903

As pausas físicas encontradas nos áudios foram anotadas em uma camada do objeto de anotação TextGrid do software para análises fonético-acústicas Praat (Boersma e Weenink, 2015. Foram anotadas as pausas físicas de qualquer duração. Naturalmente, os silêncios devidos às oclusões não foram anotados. A seguir, a tabela 55 apresenta a distribuição de posições analisadas na tarefa de percepção de pausa.

Tabela 55: Distribuição de posições analisadas na tarefa de percepção de pausa

Posições	Total	Amostra	Frequência
Terminal com pausa física	115	I	70
		II	45
Terminal sem pausa física	26	I	16
		II	10
Não-terminal com pausa física	259	I	119
		II	140
Não-terminal sem pausa física	311	I	144
		II	167
Disfluência com pausa física	40	I	15
		II	25
Disfluência sem pausa física	86	I	38
		II	48

Abaixo, na tabela 56, é apresentada a duração das pausas físicas.

Tabela 56: Duração das pausas físicas analisadas na tarefa de percepção de pausa

Duração da pausa física (ms)													
54	136	214	257	288	324	363	404	445	498	545	613	738	915
60	140	214	258	288	325	363	407	450	498	545	614	740	921
61	141	214	258	288	326	364	408	451	500	545	620	743	922
63	144	218	259	291	327	365	410	452	500	549	621	745	960
64.4	147	221	260	291	328	366	410	453	501	549	623	746	963
66	148	222	262	292	328	369	411	454	504	550	624	752	965
67	150	222	263	292	330	370	413	454	505	558	624	753	968
68	153	224	264	294	333	371	413	455	505	559	635	760	971
74	159	225	265	294	334	373	414	455	505	562	642	761	983
77	160	225	266	297	335	373	416	459	506	566	645	763	1016
82	160	227	267	298	336	373	417	461	509	567	651	768	1074
88	161	232	268	299	337	374	418	461	511	568	655	771	1091
89	161	232	269	299	338	376	419	461	513	570	657	793	1113
102	165	232	269	301	340	376	419	463	521	574	657	797	1151
102	178	232	269	304	340	381	421	464	525	575	660	798	1199
103	182	232	270	304	341	382	421	465	526	579	660	802	1322
103	183	237	271	306	343	384	421	467	527	580	661	823	1341
104	188	237	273	307	344	387	425	468	528	580	662	827	1342
106	188	238	273	309	344	388	425	473	529	584	663	828	1352
109	188	238	275	311	348	388	425	474	529	589	664	829	1363
115	190	241	277	311	350	389	426	475	532	590	668	837	1437
118	191	244	277	317	352	390	427	475	532	591	669	850	1475
121	193	245	277	317	353	392	430	479	532	592	677	855	1520
126	197	245	278	317	353	393	431	480	534	592	679	859	1849
127	202	248	278	318	354	397	431	481	534	593	684	861	
131	203	250	280	319	354	398	441	483	534	597	694	865	
131	204	251	281	319	359	400	442	486	536	599	695	869	
132	207	252	284	320	359	402	443	487	537	605	705	880	
133	209	254	284	321	360	404	443	491	539	607	707	891	
133	212	256	286	322	361	404	444	492	541	611	719	900	

Dentre as 414 pausas físicas analisadas, 401 possuem duração superior a 100 ms. A duração das demais pausas físicas varia de 54 ms a 89 ms.

5.2. Anotadores

Os anotadores da tarefa de percepção de pausa foram escolhidos com base no mesmo coeficiente Kappa usado para escolher o segmentador *expert* (vide seção 5.1). Portanto, os anotadores são chamados com números tal como na tabela 53. Dentre os 19 anotadores avaliados, os resultados obtidos mostram que o Kappa de 15 anotadores é acima de 0,8.

Inicialmente, a tarefa de percepção de pausas seria realizada pelos 15 anotadores cujo Kappa é maior que 0,8. No entanto, quando a tarefa de percepção de pausas foi realizada, alguns anotadores não participavam mais do laboratório em que a pesquisa estava sendo feita. Por esse motivo, foi necessário realizar alguns ajustes no grupo que seria submetido à tarefa de percepção de pausas.

Primeiramente, dois anotadores cujo Kappa é menor 0,8 foram incluídos no grupo que realizaria a tarefa de percepção de pausas. Neste caso, os anotadores incluídos são os anotadores 17 e 19. Além disso, três anotadores que originalmente não segmentaram os dados foram incluídos no grupo e são chamados de anotadores 20, 21 e 22. Estes três últimos anotadores são membros mais novos do laboratório e foram incluídos nessa fase do trabalho em virtude do alto coeficiente Kappa calculado entre eles após o período de formação de anotadores do *corpus* C-ORAL-BRASIL. No total, após os devidos ajustes, o grupo de anotadores que fez a tarefa de percepção de pausas é composto por 14 anotadores: anotadores 1, 2, 4, 5, 6, 7, 9, 10, 15, 17, 19, 20, 21 e 22.

5.3. Tarefa de percepção de pausas

Na tarefa de percepção de pausas, os anotadores receberam novamente os áudios e a transcrição dos áudios. Junto aos áudios e à transcrição, os anotadores receberam a segmentação ideal. Na transcrição, os símbolos designados para representar fronteira terminal, não-terminal e disfluência já estavam marcados, de acordo com a segmentação ideal considerada. Basicamente, a tarefa de percepção de pausas consistia em ouvir os áudios e anotar as posições de percepção de pausa em correspondência com as fronteiras, que já estavam anotadas previamente e eram iguais para todos. Os anotadores ouviram os áudios quantas vezes julgassem necessário e não foram treinados para anotar as pausas percebidas. Abaixo, as instruções fornecidas aos anotadores na tarefa de percepção de pausa são apresentadas no quadro 6.

Quadro 6: Instruções fornecidas aos segmentadores na tarefa de percepção de pausa

A tarefa consiste em ouvir os áudios (sem observar a forma de onda e o espectrograma), e acrescentar no arquivo txt onde você efetivamente percebeu pausas. Estou enviando três exemplos. Com estes exemplos, vocês vão entender a importância de não observar a forma de onda e espectrograma durante essa tarefa de segmentação.
No exemplo 1, temos uma pausa que provavelmente vai ser percebida por todos em virtude da sua duração relativamente longa (quase 2s). No exemplo 2, temos uma pausa mais curta que possivelmente vai ser percebida por uma menor quantidade de anotadores em virtude da sua duração bem menor. No exemplo 3, não há pausas e os anotadores possivelmente devem perceber que a fronteira estabelecida no trecho não é seguida por pausa.
Tendo isso em mente, nesta tarefa, vocês devem apenas informar se percebem uma pausa após as fronteiras (terminais, não-terminais e disfluências) previamente marcadas nos arquivos txt. As pausas percebidas devem ser marcadas com a letra **P** logo após os símbolos de fronteira terminal, não-terminal e disfluência.

Por exemplo, no trecho:
*XXX: João / gosta de viajar //

Após ouvir áudio, você deve informar se:

1) percebeu uma pausa após as duas fronteiras em questão;
2) percebeu uma pausa após a fronteira terminal;
3) percebeu uma pausa após a fronteira não-terminal;
4) não percebeu nenhuma pausa.

Caso perceba uma pausa (independentemente de sua duração) após uma fronteira previamente anotada, o arquivo deve ser alterado da seguinte forma:

*XXX: João **/P** gosta de viajar**//P** (Se percebeu pausa após ambas as fronteiras)
*XXX: João **/P** gosta de viajar// (Se percebeu uma pausa apenas após a fronteira não-terminal)
*XXX: João / gosta de viajar**//P** (Se percebeu uma pausa apenas após a fronteira terminal)
*XXX: João / gosta de viajar // (Se não percebeu pausa após nenhuma fronteira)

5.4. Acordos relativos à percepção de pausa em posição de fronteira prosódica

Após a finalização da tarefa de percepção de pausas, calculou-se o coeficiente Kappa (Fleiss, 1971) referente às marcações de pausas percebidas pelo grupo de anotadores. O coeficiente Kappa foi calculado de diferentes formas. Primeiramente, o coeficiente Kappa foi calculado de forma geral considerando todos os 14 anotadores e todas as 837 posições de fronteira prosódica discriminadas na tabela 54 (141 fronteiras terminais, 570 não-terminais e 126 disfluências). Com o primeiro cálculo geral, o coeficiente Kappa geral é 0,65. O coeficiente Kappa para cada categoria é exposto abaixo na tabela 57.

Tabela 57: Coeficiente Kappa relativo à percepção de pausa em todas as posições de fronteira prosódica

Categoria	Kappa
Fronteira terminal com pausa percebida	0,79
Fronteira terminal sem pausa percebida	0,53
Fronteira não-terminal com pausa percebida	0,57
Fronteira não-terminal sem pausa percebida	0,62
Disfluência com pausa percebida	0,57
Disfluência sem pausa percebida	0,82

Um segundo coeficiente Kappa foi calculado de forma geral considerando todos os 14 anotadores, porém, excluindo as posições de fronteira prosódica nas quais todos os 14 anotadores concordam em perceber como posições de fronteira sem pausa. Essa exclusão fundamenta-se no princípio de que o acordo total no que diz respeito à percepção de ausência de pausa pode ser desconsiderado para que somente as posições de fronteira prosódica com discussão sobre a percepção de pausa sejam avaliadas no cálculo de coeficiente Kappa. Naturalmente, no segundo tipo de cálculo de coeficiente geral de Kappa, os valores do coeficiente são menores, mas mais rigorosos. No total, para o segundo tipo de cálculo geral de coeficiente Kappa, são excluídas 95 posições. Portanto, são consideradas 742 posições de fronteira prosódica. Com o segundo cálculo geral, o coeficiente Kappa geral é 0,60 o coeficiente Kappa para cada categoria é exposto abaixo na tabela 58.

Tabela 58: Coeficiente Kappa relativo à percepção de pausa - exclusão das fronteiras que 14 anotadores concordam quanto à ausência de pausa percebida

Categoria	Kappa
Fronteira terminal com pausa percebida	0,79
Fronteira terminal sem pausa percebida	0,53
Fronteira não-terminal com pausa percebida	0,55
Fronteira não-terminal sem pausa percebida	0,56
Disfluência com pausa percebida	0,57
Disfluência sem pausa percebida	0,71

Um terceiro coeficiente Kappa foi calculado de forma geral considerando todos os 14 anotadores, porém, excluindo as fronteiras que 11 ou mais anotadores concordam quanto à ausência de pausa percebida. No total, são excluídas 348 posições. Portanto, são consideradas 489 fronteiras. O coeficiente Kappa geral é 0,52 o coeficiente Kappa para cada categoria é exposto abaixo na tabela 59.

Tabela 59: Coeficiente Kappa relativo à percepção de pausa - exclusão das fronteiras que 11 ou mais anotadores concordam quanto à ausência de pausa percebida

Categoria	Kappa
Fronteira terminal com pausa percebida	0,79
Fronteira terminal sem pausa percebida	0,35
Fronteira não-terminal com pausa percebida	0,54
Fronteira não-terminal sem pausa percebida	0,24
Disfluência com pausa percebida	0,64
Disfluência sem pausa percebida	0,40

Para cada anotador, o coeficiente Kappa foi calculado individualmente. Neste caso, o Kappa mediu a concordância entre as pausas percebidas (por cada anotador) e as pausas físicas (observadas no sinal sonoro). Todas as fronteiras foram consideradas no cálculo do coeficiente Kappa individual. Abaixo, a tabela 60 apresenta os resultados obtidos.

Tabela 60: Coeficiente Kappa relativo à percepção de pausa de cada anotador

Anotador	Geral	Disfluência sem pausa percebida	Disfluência com pausa percebida	TB sem pausa percebida	TB com pausa percebida	NTB sem pausa percebida	NTB com pausa percebida
Anot. 1	0,58	0,86	0,56	0,48	0,69	0,59	0,38
Anot. 2	0,90	0,95	0,91	0,77	0,93	0,90	0,89
Anot. 4	0,89	0,93	0,88	0,72	0,91	0,90	0,90
Anot. 5	0,82	0,88	0,71	0,59	0,83	0,84	0,81
Anot. 6	0,86	0,85	0,76	0,79	0,96	0,84	0,84
Anot. 7	0,82	0,88	0,67	0,75	0,92	0,82	0,79
Anot. 9	0,70	0,85	0,60	0,47	0,69	0,72	0,67
Anot. 10	0,58	0,74	0,21	0,50	0,79	0,59	0,42
Anot. 15	0,61	0,85	0,49	0,45	0,65	0,64	0,49
Anot. 17	0,86	0,85	0,76	0,77	0,95	0,86	0,86
Anot. 19	0,60	0,80	0,50	0,43	0,85	0,51	0,50
Anot. 20	0,56	0,70	0,64	0,19	0,89	0,36	0,51
Anot. 21	0,88	0,92	0,84	0,79	0,95	0,86	0,86
Anot. 22	0,92	0,93	0,86	0,79	0,93	0,93	0,93

De acordo com os resultados expostos na tabela 60, os anotadores podem ser divididos em dois grupos claramente distintos. O grupo 1, composto por 6 anotadores, com Kappa mais baixo variando de 0,56 a 0,7 e o grupo 2, composto por 8 anotadores, com Kappa mais alto variando de 0,82 a 0,92. Tendo em vista a possível separação dos grupos com base na percepção de pausa, também foi calculado o Kappa geral de cada grupo. Para o grupo 1, no cálculo do Kappa, são observados os anotadores 1, 9, 10, 15, 19 e 20. Para o grupo 2, são observados os anotadores 2, 4, 5, 6, 7, 17, 21 e 22. Todas as fronteiras foram consideradas no cálculo do

coeficiente Kappa dos grupos. A tabela 61 apresenta a seguir o coeficiente Kappa relativo aos grupos 1 e 2.

Tabela 61: Coeficiente Kappa relativo à percepção de pausa nos grupos 1 e 2

Categoria	**Kappa / Grupo 1**	**Kappa / Grupo 2**
Fronteira terminal com pausa percebida	0,69	0,89
Fronteira terminal sem pausa percebida	0,50	0,67
Fronteira não-terminal com pausa percebida	0,34	0,81
Fronteira não-terminal sem pausa percebida	0,48	0,82
Disfluência com pausa percebida	0,36	0,74
Disfluência sem pausa percebida	0,79	0,87
Geral	0,51	0,82

O Kappa geral do grupo 2 é muito superior ao Kappa geral do grupo 1. O acordo quanto à percepção de todas as categorias avaliadas na tarefa de percepção de pausa é maior entre os anotadores do grupo 2. Os resultados comprovam a tese de que os anotadores podem ser divididos em dois grupos com base na percepção auditiva de pausa. Evidentemente, uma análise sobre a percepção de pausa em dois grupos completamente distintos possui um potencial muito grande de contribuir nas discussões da literatura da área, mas, por se tratar da primeira investigação sobre a percepção de pausa em posição de fronteiras terminais e não-terminais, optou-se por analisar todos os anotadores e respectivas anotações neste trabalho.

5.5. Análise da duração mínima de pausa percebida

A percepção de pausa dos ouvintes pode ser estimada através da quantidade de marcações de percepção de pausa em uma dada posição. A duração das pausas físicas, por sua vez, é medida diretamente no sinal sonoro. Na literatura, ainda não há um amplo consenso sobre o que constitui uma pausa percebida. Por isso, é fundamental investigar a duração mínima de pausa percebida que minimiza de um lado a perda de pausas percebidas e de outro a confusão com a fase oclusiva dos segmentos. Naturalmente, para aprofundar o estudo e minimizar erros, devem ser analisados em conjunto a duração das pausas físicas verificadas no sinal sonoro e o nível de saliência perceptual das pausas percebidas, calculado com base no número de anotadores que marcaram a percepção de pausa em uma mesma posição. Quanto maior o número de anotadores que anotaram a percepção de pausa em uma posição, maior pode ser considerada a saliência perceptual daquela pausa percebida.

Resolvemos considerar de alta saliência as pausas percebidas por pelo menos 8 anotadores e de baixa saliência aquelas percebidas por 7 ou menos anotadores. As pausas percebidas foram divididas com base na duração das pausas físicas. Neste momento, optamos por avaliar apenas as pausas percebidas que estão associadas às pausas físicas no sinal sonoro, embora a percepção de pausa também tenha ocorrido em posição de fronteira sem pausa física. Abaixo, na tabela 62, são apresentadas as pausas percebidas pelos anotadores.

Tabela 62: Pausas percebidas pelos anotadores

Posições	**Frequência**	**Alta saliência perceptual**	**Baixa saliência perceptual**
Pausas percebidas com duração acima de 100 ms	401	373 (93%)	28 (7%)
Pausas percebidas com duração menor ou igual a 100 ms	13	10 (77%)	3 (23%)
Total de pausas percebidas	**414**	**383 (93%)**	**31 (7%)**

No total, os anotadores analisaram 837 fronteiras prosódicas. Dentre elas, 414 eram acompanhadas por pausa física no sinal sonoro. Os resultados mostram que 93% das pausas físicas, independentemente da sua duração, foram percebidas como pausa por pelo menos 8 anotadores. A quase totalidade das pausas físicas com duração acima de 100 ms (93%) foi percebida como pausa por pelo menos 8 anotadores. Além disso, os resultados evidenciam que 77% das pausas físicas com duração menor ou igual a 100 ms foram percebidas como pausa por pelo menos 8 anotadores.

Embora o conjunto de resultados relatados mostre claramente que as pausas físicas com duração acima de 100 ms têm um alto nível de saliência perceptual, não é possível afirmar com segurança que a duração mínima de pausa física percebida auditivamente como pausa é 100 ms. Em razão disso, foram analisadas as pausas físicas de menor duração encontradas nos dados. A duração das menores pausas físicas varia de 54 ms a 89 ms. No total, são 13 pausas físicas com duração entre 54 ms e 89 ms. Abaixo, na tabela 63, são apresentadas as pausas físicas de menor duração, as fronteiras associadas às menores pausas físicas e o número de marcações de pausa percebida em cada posição de pausa física com duração entre 54 ms e 89 ms.

Tabela 63: Pausas físicas com duração entre 54 ms e 89 ms

Duração da pausa física	Fronteira	Número de marcações de pausa percebida
64.4	TB	5
74	TB	11
54	NTB	11
60	NTB	8
61	NTB	8
63	NTB	8
66	NTB	8
67	NTB	6
77	NTB	9
82	NTB	8
88	NTB	9
89	NTB	8
68	DISF	2

As posições retratadas na tabela 63 foram divididas com base em diferentes intervalos de duração e níveis de saliência perceptual, conforme mostra a tabela 64. Os intervalos de duração adotados para a divisão das pausas percebidas aumentam de 10 em 10 ms a partir de 50 ms até 90 ms.

Tabela 64: Pausas percebidas com duração entre 50 ms e 90 ms

Duração da pausa percebida	Frequência	Alta saliência perceptual	Baixa saliência perceptual
Entre 50 ms e 60 ms	2	2 (100%)	0 (0%)
Entre 61 ms e 70 ms	6	3 (50%)	3 (50%)
Entre 71 ms e 80 ms	2	2 (100%)	0 (0%)
Entre 81 ms e 90 ms	3	3 (100%)	0 (0%)

Os resultados mostram que as duas pausas físicas com duração entre 50 ms e 60 ms foram percebidas como pausa por pelo menos 8 anotadores. Dentre as seis pausas físicas com duração entre 61 ms e 70 ms, três foram percebidas como pausa por pelo menos 8 anotadores. As duas pausas físicas com duração entre 71 e 80 ms foram percebidas como pausa por pelos menos 8 anotadores. As três pausas físicas com duração entre 81 ms e 90 ms foram percebidas como pausa por no mínimo 8 anotadores. Desta forma, com exceção das pausas físicas com duração entre 61 ms e 70 ms, todas as pausas físicas com duração entre 50 ms e 90 ms foram percebidas como pausa por pelo menos 8 anotadores.

Não é simples definir a duração mínima de pausa física, que é percebida como pausa pelos ouvintes e, ao mesmo tempo, evita confusão com as oclusivas. Dentre as 13 pausas físicas de menor duração avaliadas, três são imediatamente seguidas por oclusiva não vozeada. Se analisarmos as pausas físicas seguidas por oclusiva não vozeada, qualquer consideração sobre a duração mínima é arbitrária, pois é impossível mensurar o quanto da percepção de pausa é

atribuível à pausa física e o quanto é atribuível à oclusão do fone após a pausa física. A pausa física de 64,5 ms foi percebida como pausa por cinco anotadores. Além de a pausa física de 64,5 ms não possuir alta saliência perceptual, ela está localizada no final do áudio, o que impossibilita estabelecer a duração mínima baseando-se neste caso. Embora tenham sido percebidas pela maioria dos anotadores, as demais pausas físicas com duração entre 50 e 90 ms não podem tomadas como referência para escolher a duração mínima, porque muitas oclusivas poderiam ser consideradas pausas percebidas, gerando evidentemente muitos erros.

Como já foi dito, parte dos resultados deste trabalho indica que 93% das pausas físicas com duração acima de 100 ms foram percebidos pela maioria dos anotadores. Devido aos resultados e considerações apresentadas ao longo desta subseção, entendemos que é prudente propor a duração mínima de pausa de 100 ms com o intuito de evitar confusão com a fase oclusiva dos segmentos e, ao mesmo tempo, evitar que pausas relevantes perceptualmente deixem de ser identificadas. Com esta proposta de duração, se evitam que sejam consideradas pausas percebidas silêncios devidos à fase de oclusão das plosivas, cuja duração média se encontra na faixa de 100 ms (veja-se Barbosa, 2006). Isso, a nosso ver, gera uma perda mínima de pausas com duração menor, mas evita que muitas oclusões sejam consideradas pausas indevidamente.

5.6. Análise de fatores potencialmente responsáveis pela percepção de pausa

A análise de fatores potencialmente responsáveis pela percepção de pausa tem como objetivo investigar quais elementos, além da duração da pausa física, podem explicar e/ou condicionar a percepção auditiva de pausa em posição de fronteira prosódica. Os fatores diversos analisados são os seguintes:

1. Duração da pausa física;

   A duração de quase todas as pausas físicas foi extraída/medida automaticamente a partir das camadas de anotação TextGrid analisadas pelo BreakDescriptor;

   A duração das pausas físicas localizadas bem no início ou bem no final dos áudios foi medida manualmente, porque estas posições podem ser eventualmente desconsideradas pelo BreakDescriptor em razão da extensão da janela de análise do *script*.

2. Tipo de fronteira (se TB, NTB ou disfluência);

3. Taxa de articulação dos áudios;

A taxa de articulação foi calculada dividindo o número total de unidades V-Vs do áudio pela duração dos trechos de fala, isto é, desconsiderando a duração das pausas físicas anotadas em uma camada de anotação do Praat;

Abaixo, a tabela 65 apresenta a taxa de articulação dos textos.

Tabela 65: Taxa de articulação dos textos

Texto	Taxa de articulação (V-V/s)
bfammn11	4,73
bfammn24	4,54
bmidmasc01	4,97
bmidmasc02	5,17
bmidmasc03	5,03
bnatmasc01	4,65
bnatmasc02	5
bmedsp03_1a	6,79
bmedsp03_1b	6,53
bmedts_10_1	6,1
bnatco03	5,99
bnatpr05	4,04
bpubmn12	4,79
bpubmn13	5,37

4. Parâmetros segmentais em posições anteriores e posteriores às pausas físicas:
    a. Duração bruta da vogal antes da pausa física;
    b. Duração bruta da coda imediatamente anterior à pausa física (se houver);
    c. Duração bruta da rima imediatamente anterior à pausa física;
    d. Duração bruta da consoante imediatamente posterior à pausa física (se houver);
    e. Coda [w], /R/ ou /S/ antes da pausa física;
    f. Fone imediatamente posterior à pausa física.

As durações brutas dos itens a, b, c e d foram medidas manualmente no Praat sem nenhum auxílio de *scripts*. A opção de analisar as durações brutas foi escolhida para manter a unidade de medida usada para medir a duração das pausas físicas (milissegundos). As oclusivas desvozeadas após as pausas físicas foram desconsideradas na variável duração bruta da consoante imediatamente posterior à pausa física (item d), pois é impossível mensurar o quanto

da percepção de pausa é atribuível à pausa física e o quanto é atribuível à oclusão do fone após a pausa física.

Os itens e e f foram representados através de categorias para tratar a ausência ou a presença dos fenômenos em questão.

No caso da coda antes da pausa física (item e), as categorias adotadas são “R”, “S”, “W” para representar as codas possíveis em português do Brasil ou “nenhuma” para representar a ausência de coda antes da pausa física.

No caso do fone imediatamente posterior à pausa física (item f), as categorias adotadas são “vogal”, “consoante vozeada” ou “consoante desvozeada”. Portanto, os diferentes modos de articulação não são analisados no item f.

5. Parâmetros suprassegmentais em posições anteriores e posteriores às pausas físicas:
    a. Reset de f0 entre a fronteira analisada e o início da próxima unidade entonacional;
    b. Shift de f0 entre a fronteira analisada e o início da próxima unidade entonacional.

No caso do reset (item a) e shift de f0 (item b) entre a fronteira analisada e o início da próxima unidade entonacional, as categorias são “SIM” e “NÃO” para representar respectivamente a presença e a ausência do fenômeno em questão. A atribuição das categorias foi feita observando no Praat o contorno entonacional de f0.

A anotação dos fatores de 1 a 5 foi feita para todas as pausas físicas disponíveis nas amostras I e II, independentemente da respectiva duração e nível de saliência perceptual. Foi utilizada a regressão logística através de um Modelo Linear Generalizado (GLM) para compreender quais elementos têm impacto na percepção de pausa. A variável dependente investigada compreende a proporção de anotadores que perceberam pausa nas posições de fronteira com pausa física verificada no sinal sonoro. As variáveis independentes são aquelas apresentadas nos itens de 1 a 4. Foram desenvolvidos três modelos diferentes para investigar a percepção de pausa em posição de TB, NTB e disfluência com pausa física. Estes modelos visam identificar as melhores variáveis preditoras dos níveis de saliência perceptual das pausas. Desta forma, em termos gerais, os modelos permitem investigar o efeito das variáveis diversas no nível de saliência perceptual das pausas.

Primeiramente, são apresentados os fatores correlacionados com a percepção de pausa em posição de TB, NTB e disfluência com pausa física. Por último, é apresentada uma discussão mais ampla sobre os fatores correlacionados com a percepção de pausa em posição de fronteira com pausa física.

5.6.1. Fatores potencialmente responsáveis pela percepção de pausa em posição de fronteira terminal com pausa física

Os fatores relevantes para explicar a percepção de pausa em posição de TB com pausa física são apresentados na tabela 66.

Tabela 66: Fatores relevantes para a percepção de pausa em posição de TB com pausa física

Fenômeno	Coeficiente estimado	Valor de p (p-value)
Taxa de articulação	0,3230	0,005114
Duração da pausa física	0,0004	< 0,0000000000000002
Duração da coda antes da pausa física	0,0007	0,021658
Duração da consoante após a pausa física	0,0089	0,0000679
Presença de coda antes da pausa física - R	-0,8064	0,507475
Presença de coda antes da pausa física - S	-1,1213	0,030538
Presença de coda antes da pausa física – W	-0,7959	0,007443
Reset de f0 - Sim	0,5244	0,000374

A percepção de pausa em posição de TB com pausa física é explicada através de oito fatores. As variáveis numéricas significativas são a duração da pausa física, a taxa de articulação, a duração da coda antes da pausa física e a duração da consoante após a pausa física. Os resultados indicam que quanto maior a taxa de articulação, maior é o número de anotadores que marcam a percepção de pausa em posição de TB. Esta mesma relação ocorre com a duração da pausa física, com a duração da consoante após a pausa física e também com a duração de qualquer coda antes da pausa física.

Uma única variável relacionada ao reset de f0 é significativa. Quando comparada com a ausência de reset de f0, a presença de reset de f0 entre a fronteira com pausa física e o início da próxima unidade entonacional favorece a percepção de pausa, aumentando o número de anotadores que marcam a percepção de pausa em posição de TB com pausa física.

Os resultados mostraram que duas variáveis relacionadas às codas são significativas para explicar a percepção de pausa em posição de TB com pausa física. Com a variável duração da coda antes da pausa física de TB, é analisado o quanto a duração da coda, seja ela [w], /R/ ou /S/, contribui na percepção de pausa na posição em discussão. Em outras palavras, a contribuição individual da duração de cada coda possível em PB não é analisada nesta variável.

Com a variável presença de coda [w], /R/ ou /S/, é analisado o efeito individual da presença de cada coda, mas a duração dos sons não é observada diretamente.

No caso da variável que trata do efeito individual das codas antes de TB com pausa física, observa-se que a presença de [w], /R/ ou /S/ é significativa para explicar a percepção da pausa. A correlação é inversamente proporcional e o número de anotadores que indicam a percepção de pausa diminui em virtude da presença de [w], /R/ ou /S/ antes de TB com pausa física. As variáveis relativas às codas, quando analisadas em conjunto, indicam que o aumento na duração de qualquer coda antes da pausa física aumenta o número de anotadores que percebem a pausa, porém, a variável que codifica a presença de coda [w], /R/ ou /S/ antes da pausa física diminui este número em razão da correlação inversamente proporcional observada. Desta forma, os resultados sugerem que o aumento na duração de qualquer coda antes da pausa física aumenta o número de anotadores que indicam a percepção de pausa em posição de TB com pausa física, porém, não é possível afirmar qual coda (se [w]/, /R/ ou /S/) é mais relevante.

5.6.2. Fatores potencialmente responsáveis pela percepção de pausa em posição de fronteira não-terminal com pausa física

Os fatores relevantes que explicam a percepção de pausa em posição de NTB com pausa física são apresentados na tabela 67.

Tabela 67: Fatores relevantes para a percepção de pausa em posição de NTB com pausa física

Fenômeno	Coeficiente estimado	Valor de p (p-value)
Taxa de articulação	0,1224	0,02105
Duração da pausa física	0,0034	< 0,0000000000000002
Duração da vogal antes da pausa física	0,0012	0,00391
Fone imediatamente posterior à pausa física – Consoante vozeada	0,2186	0,05479

A percepção de pausa em posição de NTB é condicionada por quatro fenômenos. As variáveis numéricas significativas são a taxa de articulação, a duração da pausa física e a duração da vogal antes da pausa física. Assim como em TB, quanto maior a duração da pausa física, maior é o número de anotadores que marcam a percepção de pausa. A mesma relação de proporção é evidenciada com a taxa de articulação e a duração da vogal antes da pausa física.

O fone imediatamente posterior à pausa física também é importante para compreender a percepção auditiva de pausa em posição de NTB com pausa física. Isso significa que a presença de um determinado tipo de fone logo após a pausa física está correlacionada com a percepção de pausa na posição em discussão. Após a pausa física relativa a NTB, quando é analisado o contraste entre a presença de consoante não vozeada e a presença de consoante vozeada, a presença de consoante vozeada, independentemente do modo de articulação, é significativa. A presença de consoante vozeada após a pausa física contribui aumentando o número de anotadores que indicam a percepção de pausa (nível de saliência perceptual da pausa percebida).

5.6.3. Fatores potencialmente responsáveis pela percepção de pausa em posição de disfluência com pausa física

Os fatores relevantes para explicar a percepção de pausa em posição de disfluência com pausa física são apresentados na tabela 68.

Tabela 68: Fatores relevantes para a percepção de pausa em posição de disfluência com pausa física

Fenômeno	Coeficiente estimado	Valor de p (p-value)
Taxa de articulação	-0,2859	0,02350
Duração da pausa física	0,0016	0,0000402
Duração da vogal antes da pausa física	0,0022	0,02032
Duração da consoante após a pausa física	-0,0158	0,00219
Fone imediatamente posterior à pausa física – Consoante vozeada	1,4693	0,00087

A percepção de pausa em posição de disfluência com pausa física é explicada por cinco fenômenos. Naturalmente, a duração da pausa física é significativa e quanto maior a pausa física, maior o número de anotadores que percebem a pausa. A taxa de articulação também é significativa. Neste caso, quanto menor a taxa de articulação, maior é o número de anotadores que marcam a percepção de pausa. O fato de as pausas físicas diante das disfluências serem percebidas com maior facilidade devido à menor taxa de articulação é uma diferença entre o modelo em discussão e os demais modelos, pois, nos modelos para a percepção de pausa em posição de TB e NTB, ocorre o inverso. A duração da vogal antes da pausa física das

disfluências é outro elemento significativo e, quanto maior a duração da vogal, maior a saliência perceptual da pausa.

As pausas percebidas em posição de disfluência e TB com pausa física têm um ponto em comum. Para a percepção de pausa em posição TB, duas variáveis que tratam das codas são significativas: uma analisa a duração das codas de modo geral e outra analisa a contribuição individual de cada coda possível em PB. Para a percepção de pausa em posição de disfluência, ocorre algo semelhante, pois são significativas duas variáveis que tratam de um único elemento de ponto de vistas diferentes. Uma variável analisa de modo geral a duração das consoantes após a pausa física. A outra analisa a contribuição individual da presença dos fones diversos que podem estar localizados após as pausas físicas das disfluências.

Como já foi dito, os resultados indicam que a duração das consoantes após a pausa física das disfluências é significativa e, quanto menor a duração, maior é o número de anotadores que percebem a pausa. O tipo de fone após a pausa física das disfluências também é importante. Após a pausa física das disfluências, quando é analisado o contraste entre a presença de consoante não vozeada e a presença de consoante vozeada, a presença de consoante vozeada, independentemente do seu modo de articulação, é significativa.

Os resultados relativos às duas variáveis discutidas no parágrafo anterior, quando interpretados em conjunto, sugerem como ocorre a interação entre elas. Por um lado, os resultados mostram que quanto menor a duração de qualquer consoante após a pausa física, maior o número de anotadores que percebem a pausa. Por outro, os resultados indicam que a presença de uma consoante vozeada após a pausa física aumenta este número. Portanto, para a percepção de pausa em posição de disfluência com pausa física, é importante que a duração de qualquer consoante após a pausa física seja menor, mas é possível inferir que os fones após a pausa física que favorecem a percepção da pausa são consoantes vozeadas.

5.6.4. Parâmetros potencialmente responsáveis pela percepção de pausa em posição de fronteira com pausa física

Em tese, os parâmetros comuns aos três modelos apresentados estariam relacionados à percepção de pausa em posição de fronteira com pausa física de modo geral. Os parâmetros exclusivos de um dado modelo, por sua vez, estariam relacionados à percepção de pausa em posição de um tipo específico de fronteira com pausa física. Com a mesma correlação, o único parâmetro comum aos três modelos de percepção de pausa é a duração da pausa física. Neste

sentido, evidentemente, a duração da pausa física é considerada o único parâmetro relacionado à percepção de pausa em posição de qualquer fronteira prosódica acompanhada por pausa física.

Os parâmetros relevantes apenas para a percepção de pausa em posição de TB com pausa física são a duração da coda antes da pausa física, a presença de coda [w], /R/ ou /S/ antes da pausa física e a presença do reset de f0. O único parâmetro relevante apenas para a percepção de pausa em posição de disfluência com pausa física é a taxa de articulação, que apresenta um coeficiente negativo e indica uma correlação inversamente proporcional entre a taxa de articulação e a saliência da pausa percebida.

Não há nenhum parâmetro exclusivo do modelo de percepção de pausa em posição de NTB com pausa física. Isso significa que não é possível determinar quais fatores são relevantes apenas para a percepção de pausa em posição de NTB com pausa física. Neste caso, é possível estimar quais parâmetros são relevantes para a percepção da pausa em posição de NTB com pausa física, quando se analisa o contraste entre NTB com pausa física e outro tipo de fronteira com pausa física. Entre NTB e TB, os parâmetros distintivos são a duração da coda antes da pausa física, a duração da consoante após a pausa física, a presença de coda [w], /R/ ou /S/ antes da pausa física, o reset de f0, a duração da vogal antes da pausa física e a presença de consoante vozeada após a pausa física. Entre NTB e disfluência, os parâmetros distintivos são a taxa de articulação e a duração da consoante após a pausa física.

5.7. Resumo do capítulo

O capítulo 5 apresenta o procedimento de percepção auditiva de pausa realizado neste trabalho. O capítulo descreve em detalhes os dados utilizados, os anotadores que realizaram a tarefa de percepção e como esta tarefa foi feita. Foi calculado o acordo quanto à percepção de pausa em posição de fronteira prosódica percebida, foi analisada a duração mínima de pausa percebida pela maioria ouvintes e foi investigado quais elementos, além da pausa física, podem explicar a percepção auditiva de pausa.

Os resultados mostraram que os anotadores podem ser divididos em dois grupos com base no coeficiente Kappa. É proposta a duração mínima de pausa percebida de 100 ms com intuito de evitar confusão com a fase oclusiva dos segmentos e garantir que a escolha da duração mínima seja atribuída com base na percepção da maioria dos ouvintes. Os resultados mostram que a duração da pausa física não é o único elemento relevante para explicar a percepção de interrupção do fluxo da fala e esses fatores são diversificados a depender da natureza da

fronteira correlacionada à pausa física. Nos dados, uma pausa física de 54 ms foi percebida como pausa por 11 dos 14 anotadores. Este caso específico reforça a tese de que a duração da pausa física não é o único elemento relevante para a pausa ser percebida, pois se trata da menor pausa física encontrada nos dados e, mesmo assim, o nível de saliência perceptual é altíssimo.

6. Identificação automática de fronteiras com e sem pausa percebida

Evidentemente, para desenvolver submodelos para identificar fronteiras com e sem pausa percebida, é fundamental estabelecer o que se entende com o conceito de pausa percebida, isto é, qual seria a duração mínima de pausa física que pode ser percebida como pausa pelos ouvintes, minimizando os erros. Neste trabalho, como já foi dito, a duração proposta é 100 ms (vide capítulo 5). De acordo com esta proposta de duração, as fronteiras com pausa percebida compreendem, então, as fronteiras acompanhadas por pausa física de pelo menos 100 ms. As fronteiras sem pausa percebida, por sua vez, compreendem as fronteiras não acompanhadas por pausa física bem como as fronteiras acompanhadas por pausa física com duração inferior a 100 ms.

O desenvolvimento de submodelos para identificar fronteiras com e sem pausa percebida implica que cada submodelo reconhece um número menor de fronteiras, mas supostamente, com maior acurácia geral. Foram avaliadas duas propostas diferentes de ordenação hierárquica de submodelos de identificação de fronteiras com e sem pausa percebida. As propostas foram elaboradas com a finalidade de avaliar o que funcionaria melhor na ferramenta de identificação automática de fronteira. Sendo assim, é fundamental verificar se o desempenho final obtido por meio de um método de ordenação é melhor ou pior que o desempenho final obtido através do outro método de ordenação dos submodelos. Portanto, o processo de identificação das fronteiras com e sem pausa percebida depende da análise interna às propostas de ordenação hierárquica, mas também depende da análise entre elas.

Com a primeira proposta de ordenação hierárquica dos submodelos, o processo de identificação das fronteiras é dividido em quatro etapas sequencialmente ligadas entre si. Neste caso, a aplicação dos submodelos é organizada com base em uma ordem predeterminada. A primeira proposta de ordenação é denominada ordenação hierárquica com base em quatro etapas interligadas. Embora sejam necessárias quatro etapas, três submodelos são requeridos nesta ordem: o primeiro para identificar TB e NTB com pausa percebida; o segundo para identificar fronteira sem pausa percebida e posição sem fronteira; o terceiro para identificar se

as fronteiras sem pausa percebida identificadas pelo submodelo anterior são TB ou NTB sem pausa percebida.

Com a segunda proposta de ordenação hierárquica dos submodelos, o processo de identificação das fronteiras é dividido em três etapas sequencialmente ligadas entre si. A aplicação dos submodelos também é organizada com base em uma ordem predeterminada. A segunda proposta de ordenação é denominada ordenação hierárquica com base em três etapas interligadas. Embora sejam necessárias três etapas, apenas dois submodelos são requeridos nesta ordem: o primeiro para identificar TB e NTB com pausa percebida; o segundo para identificar ao mesmo tempo posição sem fronteira, TB e NTB sem pausa percebida.

Para fins de comparação entre as duas possibilidades de ordenação hierárquica de submodelos, foi observado/simulado o desempenho final obtido através das propostas de ordenação. O objetivo das simulações é avaliar qual proposta de ordenação hierárquica apresenta o melhor resultado de modo geral. Idealmente, a melhor proposta de ordenação deve apresentar a maior capacidade de reconhecer as fronteiras e, ao mesmo tempo, apresentar a menor quantidade de erros. Se, porventura, nenhuma das propostas de ordenação atender simultaneamente aos dois critérios de avaliação descritos, a melhor proposta deve aquela que apresentar o melhor equilíbrio entre a quantidade de fronteiras devidamente reconhecidas e a quantidade de erros gerados. Portanto, em ambas as possibilidades de ordenação, o desempenho final é medido com base na porcentagem de posições identificadas corretamente.

Na seção 6.1, são apresentados os dados usados nas fases de desenvolvimento e validação dos submodelos necessários nas propostas de ordenação hierárquica. O submodelo para identificar as fronteiras com pausa percebida é comum às duas propostas de ordenação hierárquica. Em virtude disso, as informações relativas ao submodelo em questão são apresentadas separadamente na seção 6.2. As seções 6.3 e 6.4 são destinadas aos submodelos específicos de cada proposta de ordenação hierárquica e apresentam as simulações das propostas de ordenação hierárquica de submodelos.

6.1. Dados

Para o desenvolvimento e validação de submodelos de identificação de fronteiras com e sem pausa percebida, foram consideradas as fronteiras com e sem pausa percebida das amostras I, II e extra. Na tarefa de percepção de pausa, os segmentadores marcaram as pausas percebidas nos áudios das amostras I e II, isto é, os segmentadores não marcaram as pausas

percebidas nos áudios provenientes da amostra extra. Porém, as fronteiras da amostra extra também foram divididas entre fronteiras com e sem pausa percebida com o objetivo de aumentar os dados disponíveis para realizar o desenvolvimento e a validação dos submodelos. Portanto, embora os anotadores tenham anotado apenas as pausas percebidas nos áudios provenientes das amostras I e II, as fronteiras com e sem pausa percebida das três amostras em discussão foram estabelecidas com base no mesmo critério de duração mínima de 100 ms de pausa física.

Foram usadas as posições contabilizadas pelo BreakDescriptor a partir da janela de análise de 3 unidades. Isso significa que há uma pequena diferença entre os dados usados na tarefa de percepção de pausa e os dados usados no desenvolvimento e validação dos submodelos de identificação de fronteiras com e sem pausa percebida. Na tarefa de percepção de pausa, os anotadores avaliaram todos os áudios das amostras I e II. A tarefa consistia em anotar as pausas percebidas, independentemente da localização da fronteira. Em outras palavras, os anotadores marcaram as pausas percebidas do início ao final dos áudios. Nas fases de desenvolvimento e a validação dos submodelos, as posições usadas não podem ser exatamente todas as fronteiras encontradas do início ao final dos áudios. Como já foi dito, uma das razões que explica esta diferença é o fato de o BreakDescriptor não extrair os parâmetros relativos às fronteiras eventualmente localizadas bem no início ou final do áudio em razão da extensão da janela de análise escolhida. Além disso, a diferença é devida ao acréscimo das posições relativas à amostra extra. Abaixo, na tabela 69, são apresentadas as posições utilizadas para desenvolver e validar os submodelos de identificação de fronteiras com e sem pausa percebida.

Tabela 69: Distribuição de posições usadas para desenvolver e validar os submodelos de identificação de fronteiras

Posições	Total	Amostra	Frequência
Terminal com pausa percebida	144	I	64
		II	36
		Extra	44
Terminal sem pausa percebida	24	I	12
		II	11
		Extra	1
Não-terminal com pausa percebida	357	I	109
		II	128
		Extra	120
Não-terminal sem pausa percebida	411	I	149
		II	174
		Extra	88
Sem nenhuma fronteira	2264	I	1030
		II	795
		Extra	439

As posições apresentadas na tabela 69 foram subdivididas aleatoriamente em amostras diferentes destinadas ao desenvolvimento e à validação dos submodelos. Os critérios de subdivisão dos dados são apresentados em detalhes nas seções 6.2, 6.3 e 6.4.

6.2. Submodelo de identificação de fronteiras com pausa percebida

Foi desenvolvido um submodelo para identificar TB e NTB com pausa percebida. O desenvolvimento do submodelo em questão foi realizado a partir de uma pequena amostra balanceada escolhida aleatoriamente. Para o desenvolvimento deste submodelo, é importante que o balanceamento dos dados seja realizado em função do número de fronteiras terminais com pausa percebida. Caso o balanceamento seja feito em função do número de fronteiras não-terminais com pausa percebida, é impossível selecionar uma quantidade equivalente de fronteiras terminais com pausa percebida, porque os dados não possuem um número tão grande de fronteiras deste tipo.

Para o desenvolvimento do submodelo de identificação de fronteiras com pausa percebida (FRO-compausapercebida), 70% das fronteiras terminais com pausa percebida foram escolhidas aleatoriamente (vide tabela 69). A quantidade de fronteiras usadas no desenvolvimento do submodelo corresponde, então, a 100. O balanceamento das categorias foi feito observando os 100 casos de fronteiras terminais escolhidos aleatoriamente e selecionando

aleatoriamente uma quantidade equivalente de posições de fronteira não-terminal. Portanto, a amostra usada na fase de desenvolvimento do submodelo em discussão é composta 200 posições dentre as quais 100 são TB com pausa percebida e 100 são NTB com pausa percebida. Esta amostra é chamada de Full-FROcompausapercebida.

Abaixo, a tabela 70 apresenta o desempenho alcançado pelo submodelo FRO-compausapercebida na fase de desenvolvimento.

Tabela 70: Desempenho do submodelo FRO-compausapercebida na fase de desenvolvimento

Fase	Amostra	TB c/ pausa percebida ident.	NTB c/ pausa percebida ident.	TB c/ pausa percebida marcada como NTB c/ pausa percebida	NTB c/ pausa percebida marcada como TB c/ pausa percebida
Des	Full-FROcompausa percebida	78% (78)	72% (72)	22% (22)	28% (28)

Posteriormente, para fins de validação do poder de generalização, o submodelo FRO-compausapercebida foi aplicado nas demais posições com pausa percebida, conforme mostra a tabela 71.

Tabela 71: Desempenho do submodelo FRO-compausapercebida na fase de validação

Fase	Amostra	TB c/ pausa percebida ident.	NTB c/ pausa percebida ident.	TB c/ pausa percebida marcada como NTB c/ pausa percebida	NTB c/ pausa percebida marcada como TB c/ pausa percebida
Val	Demais posições com pausa percebida	66% (29)	70% (181)	34% (15)	30% (76)

A seguir, a tabela 72 apresenta a composição do submodelo FRO-compausapercebida.

Tabela 72: Composição do submodelo FRO-compausapercebida

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	df0medd	Diferença de inclinação média de f0 dentro das janelas dos dois lados da fronteira	0,28366281
2º	df0medr1	Inclinação média de f0 na unidade pós-fronteiriça	0,11666346
3º	SDzd	Diferença de variação de duração normalizada média dos dois lados da fronteira	0,10781087
4º	ard	Diferença de taxa de articulação dos dois lados da fronteira	0,05233334
5º	zdloc	Diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça	0,03218949
6º	f0meddloc	Reset de f0	0,02972086

O submodelo FRO-compausapercebida é composto por seis parâmetros fonético-acústicos. Os parâmetros do submodelo em discussão podem ser divididos em três grupos com base no peso relativo. O primeiro grupo composto pela diferença de inclinação média de f0 dentro das janelas dos dois lados da fronteira, que apresentou um peso relativo bem mais alto que o peso dos demais parâmetros. O segundo grupo composto pela inclinação média de f0 na unidade pós-fronteiriça e a diferença de variação de duração normalizada média dos dois lados da fronteira com os pesos de 0,11 e 0,10 respectivamente. O terceiro composto pela diferença de taxa de articulação dos dois lados da fronteira, diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça e o reset de f0 com os pesos variando de 0,05 a 0,02.

6.3. Ordenação hierárquica de submodelos com base em quatro etapas interligadas

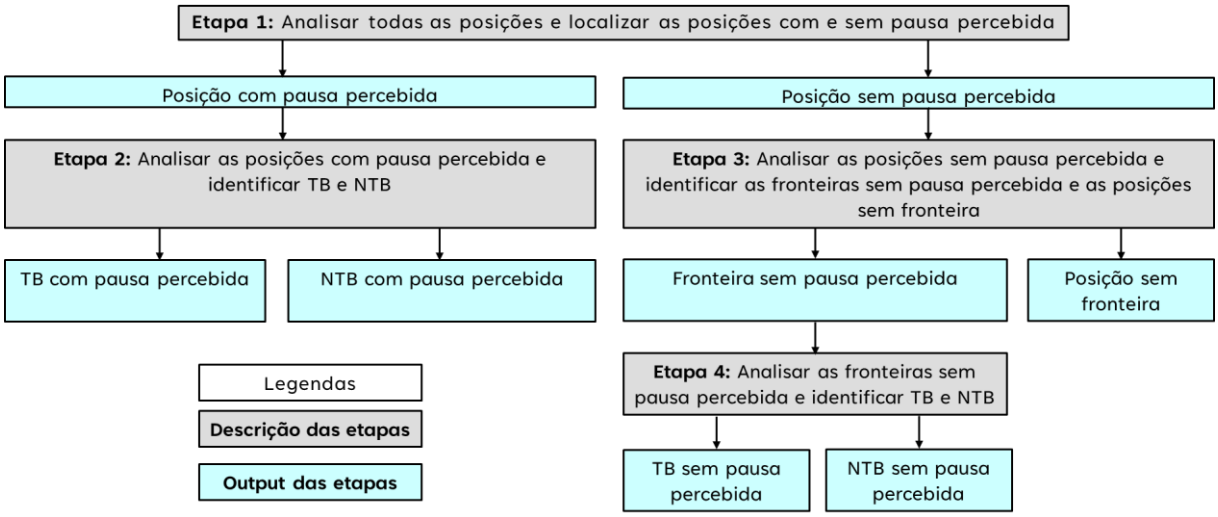
Na primeira etapa da proposta de ordenação hierárquica de submodelos com base em quatro etapas interligadas, a ferramenta de identificação de fronteira deve identificar no sinal sonoro as pausas relevantes perceptualmente para os ouvintes, isto é, as pausas percebidas de pelo menos 100 ms. Encontradas as posições de pausa percebida, na segunda etapa, o submodelo FRO-compausapercebida (vide seção 6.2) deve ser aplicado para informar se as pausas percebidas de pelo menos 100 ms identificadas automaticamente na primeira etapa estabelecem fronteiras terminais ou não-terminais com pausa percebida, de acordo com a percepção da maioria dos ouvintes.

Restam, então, todas as posições sem fronteira, fronteiras terminais e não-terminais sem pausa percebida. Foram desenvolvidos dois novos submodelos para analisar as posições restantes. Estes submodelos são requeridos apenas na proposta de ordenação hierárquica discutida neste momento. Portanto, eles não são observados na proposta de ordenação hierárquica de submodelos com base em três etapas interligadas (vide seção 6.4). Para a terceira etapa, foi desenvolvido um submodelo que tem como propósito identificar as posições sem pausa percebida, dividindo-as entre posições sem fronteira (que são necessariamente sem pausa percebida) e fronteiras sem pausa percebida. Neste caso, o submodelo não determina se as fronteiras sem pausa percebida compreendem TB ou NTB, isto é, ele identifica de modo geral as fronteiras sem pausa percebida. O submodelo a ser aplicado na terceira etapa é chamado de submodelo de identificação de posições sem pausa percebida (POS-sempausapercebida).

Para a quarta e última etapa, foi desenvolvido um submodelo para distinguir TB e NTB sem pausa percebida (FRO-sempausapercebida). Este submodelo opera informando se as

fronteiras sem pausa percebida identificadas pelo submodelo anterior (POS-sempausapercebida) compreendem TB ou NTB sem pausa percebida. Portanto, o último submodelo é responsável por dividir as fronteiras sem pausa em TB e NTB sem pausa. Evidentemente, o adequado funcionamento do submodelo FRO-sempausapercebida depende diretamente do desempenho do submodelo POS-sempausapercebida, pois o submodelo FRO-sempausapercebida deve analisar apenas as fronteiras sem pausa percebida reconhecidas anteriormente. Abaixo, o quadro 7 apresenta um diagrama ilustrativo da proposta de reconhecimento das fronteiras com e sem pausa percebida a partir da ordenação hierárquica de submodelos com base em quatro etapas interligadas.

Quadro 7: Diagrama da proposta de reconhecimento das fronteiras com e sem pausa percebida a partir da ordenação hierárquica de submodelos com base em quatro etapas interligadas



O submodelo necessário na terceira etapa (POS-sempausapercebida) foi desenvolvido a partir de uma pequena amostra composta por posições balanceadas entre si (denominada Full-POSsempausapercebida). Primeiramente, 70% das fronteiras terminais sem pausa percebida (vide tabela 69) foram escolhidas aleatoriamente. Em seguida, a quantidade equivalente de NTB sem pausa percebida foi selecionada aleatoriamente. O total de fronteiras TB e NTB selecionadas aleatoriamente corresponde a 34 fronteiras, sendo que 17 são TB e 17 são NTB. O balanceamento das categorias foi feito observando as 34 fronteiras escolhidas aleatoriamente e selecionando aleatoriamente a quantidade equivalente de posições sem nenhuma fronteira prosódica. Portanto, a amostra usada no desenvolvimento do submodelo em discussão é composta por 17 TB sem pausa percebida, 17 NTB sem pausa percebida e 34 posições sem fronteira prosódica.

Abaixo, a tabela 73 apresenta o desempenho alcançado pelo submodelo POS-sempausapercebida na fase de desenvolvimento.

Tabela 73: Desempenho do submodelo POS-sempausapercebida na fase de desenvolvimento

Fase	Amostra	FRO s/ pausa percebida ident.	POS s/ FRO ident.	FRO s/ pausa percebida marcada como POS s/ FRO	POS s/ FRO marcada como FRO s/ pausa percebida
Des	Full-POSsempausa percebida	79% (27)	82% (28)	21% (7)	18% (6)

Posteriormente, o submodelo POS-sempausapercebida foi aplicado no restante das posições sem pausa percebida para fins de validação, conforme mostra a tabela 74.

Tabela 74: Desempenho do submodelo POS-sempausapercebida na fase de validação

Fase	Amostra	FRO s/ pausa percebida ident.	POS s/ FRO ident.	FRO s/ pausa percebida marcada como POS s/ FRO	POS s/ FRO marcada como FRO s/ pausa percebida
Val.	Restante das posições s/ pausa percebida	59% (235)	56% (1238)	41% (166)	44% (989)

A seguir, a tabela 75 apresenta a composição do submodelo POS-sempausapercebida.

Tabela 75: Composição do submodelo POS-sempausapercebida

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psdur	Duração da pausa percebida	21,443610
2º	zl2	Duração normalizada média na unidade anterior à unidade pré-fronteiriça (penúltima unidade antes da fronteira)	0,343865
3º	zl3	Duração normalizada média na unidade pré-fronteiriça	0,3288884
4º	f0medd	Diferença entre medianas da F0 nas janelas dos dois lados da fronteira	0,1828489
5º	skdf0d	Diferença de variação de cauda da distribuição de F0 nas janelas dos dois lados da fronteira	0,1572054
6º	ard	Diferença de taxa de articulação dos dois lados da fronteira	0,0273170
7º	df0meddloc	Diferença de inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça	0,0264837

O submodelo POS-sempausapercebida é composto por sete parâmetros fonético-acústicos. Os parâmetros deste submodelo podem ser divididos em quatro grupos com base no peso relativo. O primeiro grupo composto pela duração da pausa percebida, que apresentou o maior peso relativo. O segundo grupo composto pela duração normalizada média na penúltima unidade antes da fronteira e a duração normalizada média na unidade pré-fronteiriça com os pesos de 0,34 e 0,32 respectivamente. O terceiro composto pela diferença entre medianas da F0 nas janelas dos dois lados da fronteira e a diferença de variação de cauda da distribuição de F0 nas janelas dos dois lados da fronteira com os pesos de 0,18 0,15 respectivamente. O quarto grupo composto pela diferença de taxa de articulação dos dois lados da fronteira e a diferença de inclinação média de f0 entre as unidades pós-fronteiriça e fronteiriça com o peso de aproximadamente 0,02 em ambos os casos.

O submodelo FRO-sempausapercebida foi desenvolvido a partir de uma pequena amostra composta por posições balanceadas entre si (denominada Full-FROsempausapercebida). O balanceamento das subcategorias foi feito escolhendo aleatoriamente 70% das fronteiras terminais sem pausa percebida (vide tabela 69) e selecionando aleatoriamente a quantidade equivalente de fronteiras não-terminais sem pausa percebida (vide tabela 69). Desta forma, a amostra Full-FROsempausapercebida é composta por 34 posições, sendo que 17 são TB sem pausa percebida e 17 NTB sem pausa percebida.

Abaixo, a tabela 76 apresenta o desempenho alcançado pelo submodelo FRO-sempausapercebida na fase de desenvolvimento.

Tabela 76: Desempenho do submodelo FRO-sempausapercebida na fase de desenvolvimento

Fase	**Amostra**	**TB s/ pausa percebida ident.**	**NTB s/ pausa percebida ident.**	**TB s/ pausa percebida marcada como NTB s/ pausa percebida**	**NTB s/ pausa percebida marcada como TB s/ pausa percebida**
Des	Full-FROsempausa percebida	100% (17)	76% (13)	NA	24% (4)

O submodelo FRO-sempausapercebida foi validado posteriormente. Neste caso, a validação consistiu em aplicar o submodelo nas fronteiras sem pausa percebida remanescentes. A seguir, a tabela 77 apresenta o desempenho do submodelo FRO-sempausapercebida na fase de validação.

Tabela 77: Desempenho do submodelo FRO-sempausapercebida na fase de validação

Fase	Amostra	TB s/ pausa percebida ident.	NTB s/ pausa percebida ident.	TB s/ pausa percebida marcada como NTB s/ pausa percebida	NTB s/ pausa percebida marcada como TB s/ pausa percebida
Val	Restante de fronteiras sem pausa percebida	71% (5)	60% (235)	29% (2)	40% (159)

A seguir, a tabela 78 apresenta a composição do submodelo FRO-sempausapercebida.

Tabela 78: Composição do modelo FRO-sempausapercebida

Rank	Sigla	Parâmetro fonético-acústico	Peso
1º	psdur	Duração da pausa percebida	21,4666062
2º	zl2	Duração normalizada média na unidade anterior à unidade pré-fronteiriça (penúltima unidade antes da fronteira)	0,40909641
3º	sdf0d	Diferença de variação das medianas de f0 dos dois lados da fronteira	0,14855759
4º	skf0d	Diferença das caudas da distribuição de f0 dos dois lados da fronteira	0,09164049
5º	f0meddloc	Reset de f0	0,05998988
6º	zl3	Duração normalizada média na unidade pré-fronteiriça	0,04017814

O submodelo FRO-sempausapercebida é composto por seis parâmetros fonético-acústicos. O parâmetro de maior peso relativo é a duração da pausa percebida com o peso de 21,4. Com o peso 0,4, o segundo parâmetro de maior relevância hierárquica é a duração normalizada média na unidade anterior à unidade pré-fronteiriça (penúltima unidade antes da fronteira). Os demais parâmetros do submodelo em discussão são a diferença de variação das medianas de f0 dos dois lados da fronteira, a diferença das caudas da distribuição de f0 dos dois lados da fronteira, o reset de f0 e a duração normalizada média na unidade pré-fronteiriça com os pesos variando de 0,14 a 0,04.

Em resumo, a ordenação hierárquica de submodelos com base em quatro etapas interligadas é preestabelecida e fixa. Inicialmente, é necessário que as pausas percebidas a partir de 100 ms sejam identificadas automaticamente. Nas posições de pausa percebida identificadas automaticamente, o submodelo FRO-compausapercebida identifica as fronteiras TB e NTB com pausa percebida. Nas posições remanescentes, o submodelo POS-sempausapercebida identifica as fronteiras sem pausa percebida (de modo geral) e as posições sem fronteira. Nas posições de fronteira sem pausa percebida (previamente reconhecidas pelo submodelo POS-sempausapercebida), o submodelo FRO-sem pausapercebida identifica, então, TB e NTB sem pausa percebida.

Para fins de estimação/ simulação de quais seriam os resultados alcançados por uma ferramenta de identificação de fronteira implementada desta forma, os submodelos foram aplicados em todas as posições provenientes das amostras I, II e extra, seguindo a ordenação hierárquica proposta. A aplicação foi realizada através do ambiente para computação estatística R (R Core Team, 2017). Dado que as pausas percebidas de pelo menos 100 ms já foram reconhecidas automaticamente no sinal sonoro, o submodelo FRO-compausapercebida foi aplicado apenas nas posições em discussão com o objetivo de identificar as posições de TB e NTB com pausa percebida. Abaixo, a tabela 79 apresenta o desempenho do submodelo FRO-compausapercebida na fase de simulação.

Tabela 79: Simulação do desempenho do submodelo FRO-compausapercebida

Amostra	**TB c/ pausa percebida ident.**	**NTB c/ pausa percebida ident.**	**TB c/ pausa percebida marcada como NTB c/ pausa percebida**	**NTB c/ pausa percebida marcada como TB c/ pausa percebida**
Todas as posições com pausa percebida	74% (107)	71% (253)	26% (37)	29% (104)

De acordo com a ordenação em discussão, as posições sem pausa percebida devem ser marcadas como fronteira sem pausa percebida ou posição sem fronteira com auxílio do submodelo POS-sempausapercebida. Portanto, na fase de simulação, o submodelo POS-sempausapercebida foi aplicado em todas as posições sem pausa percebida. Abaixo, a tabela 80 apresenta o desempenho do submodelo POS-sempausapercebida na fase de simulação.

Tabela 80: Simulação do desempenho do submodelo FRO-compausapercebida

Amostra	**FRO s/ pausa percebida ident.**	**POS s/ FRO ident.**	**FRO s/ pausa percebida marcada como POS s/ FRO**	**POS s/ FRO marcada como FRO s/ pausa percebida**
Todas as posições s/ pausa percebida	60% (262)	56% (1269)	40% (173)	44% (995)

Em conformidade com a ordenação hierárquica dos submodelos em quatro etapas interligadas, as posições de TB e NTB sem pausa percebida devem ser reconhecidas por último, após a aplicação do submodelo POS-sempausapercebida. A identificação de TB e NTB sem pausa percebida é feita por meio do submodelo FRO-sempausapercebida. Este submodelo opera em dois tipos de posições. Ele opera nas posições de fronteira sem pausa percebida

reconhecidas pelo submodelo anterior, mas também opera nas posições sem nenhuma fronteira, que foram marcadas indevidamente como fronteira sem pausa percebida. Por isso, o último submodelo marca como TB ou NTB sem pausa percebida até mesmo algumas posições sem fronteira prosódica, caso elas tenham sido indevidamente marcadas como fronteira sem pausa percebida pelo submodelo anterior. Evidentemente, as posições sem fronteira marcadas indevidamente como fronteira sem pausa percebida permanecem com a classificação errada. De certo modo, o erro é até ampliado, visto que essas posições passam a ser consideradas TB ou NTB sem pausa percebida pelo submodelo FRO-sempausapercebida. Para facilitar a compreensão dos resultados obtidos na fase de simulação final, a tabela 81 apresenta o desempenho do submodelo FRO-sempausapercebida, quando aplicado nas fronteiras sem pausa percebida identificadas corretamente pelo submodelo anterior.

Tabela 81: Simulação do desempenho do submodelo FRO-sempausapercebida em posição de fronteira sem pausa percebida identificada corretamente pelo submodelo anterior

Amostra	TB s/ pausa percebida ident.	NTB s/ pausa percebida ident.	TB s/ pausa percebida marcada como NTB s/ pausa percebida	NTB s/ pausa percebida marcada como TB s/ pausa percebida
Fronteiras sem pausa percebida corretamente identificadas pelo submodelo anterior	94% (17)	60% (146)	6% (1)	40% (98)

Na tabela 82, é apresentado o desempenho do submodelo FRO-sempausapercebida, quando aplicado nas posições sem fronteira marcadas indevidamente como fronteira sem pausa percebida pelo submodelo anterior.

Tabela 82: Simulação do desempenho do submodelo FRO-sempausapercebida em posições sem fronteira marcadas indevidamente como fronteira sem pausa percebida pelo submodelo anterior

Amostra	Posição sem fronteira marcada como NTB s/ pausa percebida	Posição sem fronteira marcada como TB s/ pausa percebida
Fronteiras sem pausa percebida indevidamente marcadas pelo submodelo anterior	54% (540)	46% (455)

De modo geral, a simulação dos resultados a serem alcançados com a proposta de aplicação hierárquica de submodelos com base em quatro etapas interligadas sugere que houve um certo avanço no que diz respeito à superestimação dos parâmetros de pausa física. Os resultados mostraram que mais de 70% de TB e NTB com pausa percebida foram devidamente

reconhecidas pelo submodelo FRO-compausapercebida. Porém, ainda assim, houve uma certa confusão entre TB e NTB com pausa percebida. Aproximadamente, 26% das posições de TB com pausa percebida foram marcadas como NTB com pausa percebida e 29% das fronteiras NTB com pausa percebida foram marcadas indevidamente como TB.

A identificação automática de TB e NTB sem pausa percebida, por sua vez, parece ser mais complicada. Na fase de simulação dos resultados a serem alcançados através do submodelo POS-sempausapercebida, 60% das fronteiras sem pausa percebida foram identificadas corretamente. Portanto, 40% das fronteiras sem pausa foram marcadas indevidamente como posições sem fronteira nenhuma. Ainda na simulação do submodelo POS-sempausapercebida, apenas 56% das posições sem fronteira foram reconhecidas. Assim, 44% das posições sem fronteira foram indevidamente consideradas fronteiras sem pausa percebida. O submodelo POS-sempausapercebida foi desenvolvido para identificar as posições sem fronteira e as fronteiras sem pausa percebida, isto é, o submodelo em questão não foi desenvolvido para informar se as fronteiras sem pausa percebida eram TB ou NTB. Os resultados do submodelo POS-sempausapercebida sugerem, em geral, que é difícil distinguir as posições sem fronteira das fronteiras sem pausa percebida, pois parte das fronteiras sem pausa percebida foi marcada como posição sem fronteira e vice-versa.

Após a identificação das posições sem fronteira e identificação das fronteiras sem pausa percebida através do submodelo POS-sempausapercebida, o submodelo FRO-sempausapercebida foi aplicado **apenas** nas posições de fronteira sem pausa percebida indicadas previamente pelo submodelo POS-sempausapercebida. Evidentemente, isso significa que o submodelo POS-sempausapercebida influencia diretamente na *performance* do submodelo FRO-sempausapercebida. O nível de interferência pode ser considerado alto por dois motivos. Primeiramente, o submodelo POS-sempausapercebida limita as fronteiras terminais e não-terminais sem pausa percebida que poderiam ser identificadas pelo submodelo FRO-sempausapercebida. Além disso, o submodelo POS-sempausapercebida "força" o submodelo FRO-sempausapercebida a marcar como TB ou NTB sem pausa percebida as posições que foram marcadas indevidamente como fronteira sem pausa percebida pelo submodelo POS-sempausapercebida.

Nas posições que realmente são fronteiras sem pausa percebida e foram devidamente identificadas pelo submodelo POS-sempausapercebida, o submodelo FRO-sempausapercebida apresentou um resultado satisfatório no que diz respeito ao reconhecimento de TB sem pausa percebida. O submodelo em discussão identificou 94% das fronteiras terminais sem pausa percebida que haviam sido marcadas como fronteira sem pausa percebida pelo submodelo POS-

sempausapercebida, isto é, apenas 6% destas posições foram marcadas indevidamente como NTB sem pausa percebida. Dentre as fronteiras não-terminais sem pausa percebida que foram marcadas como fronteira sem percebida pelo submodelo POS-sempausapercebida, o submodelo FRO-sempausapercebida identificou 60% delas como NTB sem pausa percebida e deixou de identificar 40%, marcando-as como TB sem pausa percebida.

Nas posições sem fronteira que foram marcadas como fronteira sem pausa percebida pelo submodelo POS-sempausapercebida, o submodelo FRO-sempausapercebida marcou 54% delas como TB sem pausa percebida e 46% como NTB sem pausa percebida. Portanto, estas posições passaram a ser consideradas TB ou NTB sem pausa percebida pelo último submodelo em razão da marcação errada realizada pelo submodelo POS-sempausapercebida.

A proposta de aplicação de hierárquica de submodelos com base em quatro etapas interligadas falha principalmente ao analisar as posições sem pausa percebida. A falha de maior gravidade é obtida quando o submodelo POS-sempausapercebida marca posições sem fronteira como fronteira sem pausa percebida, isto é, algumas fronteiras sem pausa percebida indicadas pelo submodelo POS-sempausapercebida não são fronteiras prosódicas de fato. Em razão da forma como a proposta de aplicação hierárquica foi organizada, essas falhas fazem com que "as fronteiras sem pausa percebida" que não são fronteiras passem a ser consideradas TB ou NTB com pausa percebida pelo submodelo subsequente, que é o submodelo FRO-sempausapercebida.

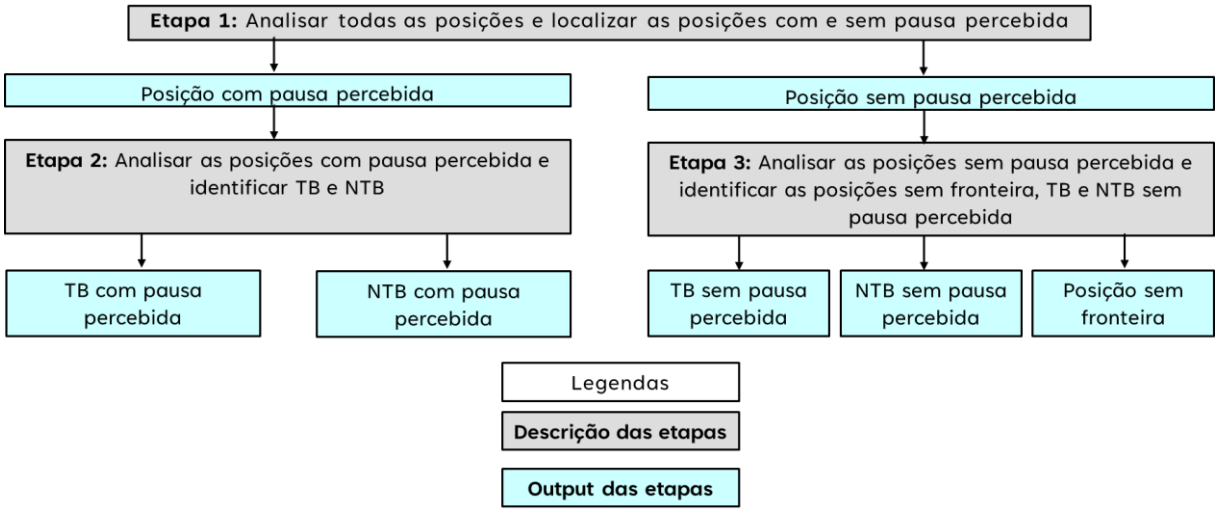
Outro ponto falho da proposta de aplicação hierárquica aqui em discussão está relacionado ao reconhecimento de NTB sem pausa percebida. Entre as fronteiras não-terminais sem pausa percebida que haviam sido reconhecidas como fronteira sem percebida pelo submodelo POS-sempausapercebida, apenas 60% delas foram devidamente marcadas como NTB sem pausa percebida. De modo geral, a implementação de uma ferramenta de identificação automática de TB e NTB, seguindo a proposta de ordenação hierárquica em discussão, requer que o submodelo POS-sempausapercebida apresente melhores resultados e consiga reconhecer com maior facilidade as fronteiras sem pausa percebida e as posições sem nenhuma fronteira. Esta melhora é fundamental para que o submodelo FRO-sempausapercebida possa analisar posições que realmente são fronteiras prosódicas sem pausa percebida.

6.4. Ordenação hierárquica de submodelos com base em três etapas interligadas

Com a ordenação hierárquica de submodelos com base em três etapas interligadas, as pausas percebidas de pelo menos 100 ms devem ser identificadas automaticamente

inicialmente. Em seguida, o submodelo FRO-compausapercebida deve ser aplicado apenas nas posições de pausa percebida com o objetivo de identificar as posições de TB e NTB acompanhadas por pausa percebida. Após o reconhecimento das fronteiras com pausa percebida, um único submodelo deve ser aplicado nas posições remanescentes para distinguir e identificar TB sem pausa percebida, NTB sem pausa percebida e posição sem fronteira. Portanto, neste caso, são necessários apenas dois submodelos. O segundo submodelo a ser aplicado é chamado de GER-sempausapercebida. A vantagem de utilizar este submodelo é eliminar a possibilidade de um submodelo interferir diretamente no desempenho de outro. Porém, a tarefa a ser executada pelo submodelo pode ser mais considerada mais difícil, visto que ele deve reconhecer três categorias ao mesmo tempo. Abaixo, o quadro 8 apresenta um diagrama ilustrativo da proposta de reconhecimento das fronteiras com e sem pausa percebida a partir da ordenação hierárquica de submodelos com base em três etapas interligadas.

Quadro 8: Diagrama da proposta de reconhecimento das fronteiras com e sem pausa percebida a partir da ordenação hierárquica de submodelos com base em três etapas interligadas



O submodelo GER-sempausapercebida foi desenvolvido a partir de uma amostra composta por posições balanceadas entre si (denominada Full-GERsempausapercebida). Primeiramente, 70% das fronteiras terminais sem pausa percebida (vide tabela 69) foram escolhidas aleatoriamente. Em seguida, a quantidade equivalente de NTB sem pausa percebida foi selecionada aleatoriamente. No total, foram selecionadas 34 fronteiras, sendo que 17 são TB e 17 são NTB. O balanceamento das categorias foi feito observando as 34 fronteiras escolhidas aleatoriamente e selecionando aleatoriamente a quantidade equivalente de posições sem nenhuma fronteira prosódica. Portanto, a amostra usada no desenvolvimento do submodelo

GER-sempausapercebida é composta por 17 TB sem pausa percebida, 17 NTB sem pausa percebida e 34 posições sem fronteira prosódica, totalizando 68 posições. A seguir, a tabela 83 apresenta o desempenho alcançado pelo submodelo GER-sempausapercebida na fase de desenvolvimento.

Tabela 83: Desempenho do submodelo GER-sempausapercebida na fase de desenvolvimento

Desempenho do submodelo	Porcentagem
TB s/ pausa percebida ident.	70,6% (12)
NTB s/ pausa percebida ident.	41,2% (7)
Posição sem fronteira ident.	91,2% (31)
TB s/ pausa percebida marcada como NTB s/ pausa percebida	5,9% (1)
TB s/ pausa percebida marcada como posição sem fronteira	23,5% (4)
NTB s/ pausa percebida marcada como TB s/ pausa percebida	17,6% (3)
NTB s/ pausa percebida marcada como posição sem fronteira	41,2% (7)
Posição sem fronteira marcada como TB s/ pausa percebida	5,9% (2)
Posição sem fronteira marcada como NTB s/ pausa percebida	2,9% (1)

Posteriormente, o submodelo GER-sempausapercebida foi aplicado no restante das posições sem pausa percebida para fins de validação, conforme mostra a tabela 84.

Tabela 84: Desempenho do submodelo GER-sempausapercebida na fase de validação

Desempenho do submodelo	Porcentagem
TB s/ pausa percebida ident.	14,2% (1)
NTB s/ pausa percebida ident.	27,4% (108)
Posição sem fronteira ident.	70,8% (1579)
TB s/ pausa percebida marcada como NTB s/ pausa percebida	42,9% (3)
TB s/ pausa percebida marcada como posição sem fronteira	42,9% (3)
NTB s/ pausa percebida marcada como TB s/ pausa percebida	15,5% (61)
NTB s/ pausa percebida marcada como posição sem fronteira	57,1% (225)
Posição sem fronteira marcada como TB s/ pausa percebida	12,7% (283)
Posição sem fronteira marcada como NTB s/ pausa percebida	16,5% (368)

A seguir, a tabela 85 apresenta a composição do submodelo GER-sempausapercebida.

Tabela 85: Composição do submodelo GER-sempausapercebida

Rank	Sigla	Parâmetro fonético-acústico	Peso 1	Peso 2
1º	zr1	Duração normalizada média na unidade pós-fronteiriça	89,59653	57,41074288
2º	zdloc	Diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça	89,9453098	57,05547239
3º	zl0	Duração normalizada média na unidade fronteiriça	89,2998727	57,53918595
4º	psdur	Duração da pausa percebida	20,2545003	29,86331182
5º	df0med0	Inclinação média de f0 na unidade fronteiriça	0,664721	0,70530647
6º	df0medr1	Inclinação média da f0 na 1ª unidade da janela à direita da fronteira	0,7241735	0,57804118
7º	df0meddloc	Diferença de inclinação média da f0 entre as unidades pós-fronteiriça e fronteiriça	0,4734982	0,71257728
8º	zl3	Duração normalizada média na unidade pré-fronteiriça	0,3791097	0,06730178
9º	f0meddloc	Reset de f0	0,3389544	0,13617675
10º	skdf0d	Diferença de variação de cauda da distribuição de f0 nas janelas dos dois lados da fronteira	0,1297375	0,10962916

Os submodelos necessários na proposta de ordenação hierárquica em discussão foram aplicados em todas as posições provenientes das amostras I, II e extra com o objetivo de simular quais seriam os resultados alcançados pela ferramenta de identificação automática de fronteira, caso ela fosse implementada desta forma. A aplicação foi realizada através do ambiente para computação estatística R (R Core Team, 2017).

A proposta de aplicação hierárquica de submodelos com base em três etapas interligadas é semelhante à proposta baseada em quatro etapas até certo ponto. A semelhança é devida ao processo de identificar as pausas percebidas de pelo menos 100 ms e aplicar o submodelo FRO-compausapercebida inicialmente. O desempenho relativo à simulação dos resultados a serem alcançados através do submodelo FRO-compausapercebida é o mesmo apresentado na tabela 79 e, por este motivo, ele não é apresentado novamente. Neste sentido, no que diz respeito à supervalorização dos parâmetros de pausa física, o avanço destacado na seção anterior também é válido.

Dado que o submodelo FRO-compausapercebida já foi usado para reconhecer as fronteiras TB e NTB com pausa percebida em um primeiro momento, as próximas posições a serem reconhecidas são as posições sem fronteira, TB e NTB sem pausa percebida com auxílio

do submodelo GER-sempausapercebida. Abaixo, a tabela 86 apresenta o desempenho do submodelo GER-sempausapercebida na fase de simulação.

Tabela 86: Simulação do desempenho do submodelo GER-sempausapercebida em posição sem fronteira, TB e NTB sem pausa percebida

Desempenho do submodelo	Porcentagem
TB s/ pausa percebida ident.	54,2% (13)
NTB s/ pausa percebida ident.	28% (115)
Posição sem fronteira ident.	71,1% (1610)
TB s/ pausa percebida marcada como NTB s/ pausa percebida	16,6% (4)
TB s/ pausa percebida marcada como posição sem fronteira	29,2% (7)
NTB s/ pausa percebida marcada como TB s/ pausa percebida	15,6% (64)
NTB s/ pausa percebida marcada como posição sem fronteira	56,4% (232)
Posição sem fronteira marcada como TB s/ pausa percebida	12,6% (285)
Posição sem fronteira marcada como NTB s/ pausa percebida	16,3% (369)

Os resultados do submodelo destinado a identificar ao mesmo tempo as posições sem fronteira, TB e NTB sem pausa percebida (GER-sempausapercebida) mostraram que a dificuldade para diferenciar estas posições persiste na fase simulação. O submodelo GER-sempausapercebida identificou corretamente 71,1% das posições sem fronteira, mas marcou respectivamente 12,6% e 16,3% das posições sem fronteira como TB e NTB sem pausa percebida. Aproximadamente 54,2% das fronteiras terminais sem pausa percebida foram devidamente identificadas, mas, em contrapartida, respectivamente 16,6% e 29,2% foram marcadas incorretamente como NTB sem pausa percebida e posição sem fronteira. Entre os três tipos de posições a serem identificadas através do submodelo GER-sempausa, a menor porcentagem de reconhecimento foi relativa às posições de NTB sem pausa percebida. O submodelo em discussão identificou apenas 28% das fronteiras NTB sem pausa percebida, marcando 15,6% delas como TB sem pausa percebida e 56,4% como posição sem fronteira.

Embora algumas posições sem fronteira marcadas pelo submodelo sejam, na verdade, posições de TB e NTB sem pausa percebida, as posições sem fronteira foram as posições identificadas com maior acurácia, pois a porcentagem de reconhecimento foi superior a 70%. As principais limitações do submodelo GER-sempausapercebida estão relacionadas à identificação de TB e NTB sem pausa percebida. Diversas posições de TB sem pausa percebida

foram marcadas indevidamente como NTB sem pausa percebida ou posição sem fronteira. O mesmo ocorre com NTB: diversas NTB sem pausa percebida foram marcadas como TB sem pausa percebida ou posição sem fronteira. Todas estas marcações incorretas podem ser vistas como indicativos de que o submodelo apresenta uma certa dificuldade para diferenciar as posições sem fronteira, TB e NTB sem pausa percebida. Para implementar uma ferramenta de identificação automática de fronteiras, de acordo com a proposta de ordenação hierárquica em discussão, é fundamental que o submodelo GER-sempausapercebida consiga reconhecer e diferenciar com maior facilidade as três posições que ele analisa.

6.5. Resumo do capítulo

Neste capítulo, com base em métodos diferentes de ordenação hierárquica de submodelos, duas formas diferentes de identificar automaticamente fronteiras com e sem pausa percebida foram avaliados. Os submodelos necessários foram desenvolvidos a partir de parâmetros extraídos de janelas de análise de três unidades e amostras compostas por categorias balanceadas entre si. Os submodelos desenvolvidos foram validados em dados sem nenhum tipo de balanceamento de categorias e, posteriormente, foi realizada uma simulação de resultados a serem alcançados em caso de implementação.

Foram destacados os resultados alcançados pelos submodelos nas três fases relatadas (desenvolvimento, validação e simulação) e os parâmetros que compõem os submodelos desenvolvidos. O submodelo de identificação de fronteiras com pausa percebida, comum às duas propostas de ordenação hierárquica de submodelos, apresentou resultados satisfatórios de modo geral, reduzindo a importância da pausa física. Os submodelos de identificação de posições sem fronteira e fronteiras sem pausa percebida, necessários em cada proposta de ordenação hierárquica de submodelos, apresentam limitações evidentes.

7. Considerações finais

Neste trabalho, foram investigadas formas de tornar mais rápido o processo de compilação de *corpora* de fala segmentados em unidades entonacionais bem como as combinações de parâmetros fonético-acústicos envolvidos na percepção das fronteiras prosódicas. Tanto o estudo sobre a identificação automática das fronteiras, quanto o estudo sobre as combinações de parâmetros envolvidos na percepção das fronteiras, foram conduzidos

à luz dos modelos e submodelos apresentados e discutidos. Inicialmente, os modelos desenvolvidos em Teixeira (2018) foram aplicados nas amostras II e total para fins de revalidação. As revalidações dos modelos em discussão mostraram que o desempenho se mantém relativamente estável em amostras diferentes daquelas em que os modelos foram desenvolvidos originalmente, o que permite supor que a generalização foi alcançada.

Em um segundo momento, foram desenvolvidos novos modelos para identificar TB e NTB a partir de diferentes estratégias. As estratégias empregadas consistiam em desenvolver novos modelos de identificação de TB e NTB a partir de amostras balanceadas e parâmetros extraídos de janelas de análise compostas por 10, 7 e 3 unidades. Em todos os três casos, os modelos desenvolvidos foram validados em amostras não balanceadas depois.

A análise dos resultados alcançados pelos diversos modelos TB e NTB permite realizar uma série de observações sobre o processo de identificação automática das fronteiras prosódicas em discussão. Os modelos TB, em sua maioria, reconheceram pelo menos 75% das fronteiras terminais. Porém, é importante ressaltar que a maior porcentagem de reconhecimento de TB é explicada pelo fato de aproximadamente 80% delas serem acompanhadas por pausa física. Os modelos TB raramente identificaram as fronteiras TB sem pausa física. Os falsos alarmes gerados pelos modelos TB ocorreram principalmente em posição de NTB com pausa física.

Os modelos NTB, por sua vez, não apresentaram resultados muito satisfatórios, seja pela pequena quantidade de fronteiras identificadas ou pela grande quantidade falsos alarmes, e, frequentemente, pelos dois motivos. A maioria das fronteiras NTB identificada era acompanhada por pausa física. Portanto, frequentemente, os modelos NTB não foram capazes de reconhecer as fronteiras NTB sem pausa física. Os falsos alarmes dos modelos NTB foram obtidos principalmente em posição de TB com pausa física.

As exceções são os modelos NTB-b1, NTB-b2 e NTB-nbfull. Os modelos NTB-b1 e NTB-b2 se diferem dos demais modelos NTB por terem identificado mais de 68% das posições de NTB, mas não se diferem no que diz respeito aos falsos alarmes, pois, em ambos os casos, foram obtidos diversos falsos alarmes em posição sem fronteira e em posição de TB com pausa física. O modelo NTB-nbfull, por sua vez, se difere dos demais modelos NTB por ter obtido apenas 0,3% de falsos alarmes, mas, em contrapartida, não se difere no que diz respeito à capacidade de reconhecer NTB, pois identificou apenas 40,5% das fronteiras não-terminais.

Nos modelos TB e NTB diversos, pelo menos um parâmetro relativo à pausa física apresenta elevado peso relativo. De modo geral, os resultados sugerem que os diversos modelos TB e NTB superestimam a importância dos parâmetros relativos às pausas físicas como recurso de identificação das fronteiras. Neste sentido, o fato de os modelos TB e NTB apresentarem

problemas semelhantes é explicado pela superestimação dos parâmetros de pausa física. Evidentemente, é justificável a inclusão dos parâmetros de pausa física nos modelos diversos, visto que a pausa física é frequentemente associada à percepção das fronteiras prosódicas em diversos trabalhos da área. Porém, os parâmetros de pausa física parecem impedir que as fronteiras terminais e não-terminais se diferenciem entre si, o que é um problema para a implementação da ferramenta de identificação automática de fronteiras.

Os modelos TB e NTB aqui apresentados não necessariamente mostram os parâmetros que são percebidos pelos segmentadores quando eles percebem as fronteiras prosódicas. Portanto, os modelos não necessariamente capturam como ocorre a percepção das fronteiras e não necessariamente indicam quais parâmetros pesam mais quando uma fronteira é percebida pelos ouvintes. Os modelos são ferramentas que visam reproduzir a segmentação realizada pelos segmentadores. Os elementos usados pelos segmentadores para identificar as fronteiras são diferentes dos elementos observados pelos modelos. Os modelos devem ser considerados um recurso que permite inferir sobre as combinações de parâmetros fonético-acústicos que podem levar à percepção das fronteiras, isto é, eles auxiliam a compreender os parâmetros que podem estar envolvidos na percepção das fronteiras.

Além disso, apesar dos modelos TB e NTB terem as suas limitações, eles indicam as combinações de parâmetros fonético-acústicos relevantes para distinguir determinadas posições de outras. Neste sentido, os modelos apresentados não indicam diretamente quais parâmetros caracterizam um determinado tipo de fronteira prosódica. No âmbito dos modelos TB desenvolvidos, é possível inferir as combinações de parâmetros importantes para distinguir as posições de TB das posições de ausência de TB (que podem ser NTB ou posição sem nenhuma fronteira). No caso dos modelos NTB, é possível inferir as combinações de parâmetros relevantes para distinguir NTB das posições sem fronteira.

Os parâmetros de duração normalizada assumiram um peso relativo mais alto nos modelos NTB, enquanto nos modelos TB, eles quase não foram incluídos e quando foram, o peso relativo foi sempre mais baixo. Neste sentido, os parâmetros de duração normalizada são mais importantes para diferenciar NTB das posições sem fronteira. Os parâmetros de f0 tendem a ser mais importantes em modelos de identificação das fronteiras TB em razão do elevado peso relativo. Nos modelos NTB, eles raramente foram incluídos e quando foram, o peso relativo foi baixo. Uma exceção é o modelo NTB-2, que é composto principalmente por parâmetros de f0. No entanto, este modelo identificou apenas 25% das fronteiras não-terminais não identificadas anteriormente pelo modelo NTB-1. Portanto, de modo geral, os parâmetros de f0 são mais relevantes para diferenciar as posições de TB das posições de ausência de TB.

Os parâmetros de taxa de articulação e elocução foram incluídos em modelos TB e NTB, mas apenas em alguns modelos NTB, o peso relativo destes parâmetros foi alto. Neste sentido, os parâmetros de taxa de articulação e elocução foram relevantes para diferenciar as posições de NTB das posições sem nenhuma fronteira, a depender do modelo NTB. Os parâmetros de intensidade não foram relevantes em nenhum modelo de identificação de fronteiras. Portanto, eles não podem ser considerados relevantes para distinguir as fronteiras TB das posições de ausência de TB e nem para distinguir as fronteiras NTB das posições sem fronteira prosódica.

Na literatura da área, alguns parâmetros são frequentemente associados à percepção das fronteiras prosódicas em diversas línguas. Estes parâmetros são a pausa física, o reset de f0, a mudança no contorno entoacional de f0, o alongamento em posições anteriores às fronteiras, o aumento na taxa de elocução e alguns parâmetros relativos à intensidade. Considerando os diversos modelos TB desenvolvidos, a combinação de parâmetros "básica" para diferenciar as fronteiras TB das posições de ausência de TB é composta pela pausa física, o reset de f0 e a inclinação média de f0 em unidades próximas ao final das palavras fonológicas, que se relaciona à mudança no contorno entonacional de f0. Considerando os diversos modelos NTB, a combinação de parâmetros "básica" para diferenciar as fronteiras NTB das posições sem fronteira abrange a pausa física, parâmetros de duração normalizada média em unidades próximas ao final das palavras fonológicas, e, em alguns casos, a diferença nas taxas de elocução e/ ou articulação. De certa forma, com exceção dos parâmetros de intensidade, os parâmetros relevantes para diferenciar as oposições de posições avaliadas neste trabalho são compatíveis com os parâmetros citados na literatura da área.

Os modelos TB e NTB aqui apresentados podem ser considerados a primeira tentativa de identificar automaticamente estes tipos de fronteiras, apesar das evidentes limitações. Os resultados sugerem, de modo geral, que apenas dois modelos (um para identificar TB e outro para NTB) não são suficientes para solucionar os problemas causados pela superestimação dos parâmetros de pausa física. Em outras palavras, um único modelo TB não parece ser capaz de parar de obter falsos alarmes em posições de NTB com pausa física e, ao mesmo tempo, passar a reconhecer as fronteiras terminais com e sem pausa física. O mesmo se aplica à identificação das fronteiras NTB. Um único modelo NTB não parece ser capaz de reconhecer as fronteiras não-terminais com e sem pausa física e, ao mesmo tempo, diminuir drasticamente os falsos alarmes em posição sem fronteira e em posição de TB com pausa física.

O desempenho dos modelos é medido com relação à percepção de fronteira de tipo terminal e não-terminal pela maioria dos segmentadores. Deste modo, para os modelos serem considerados de alta *performance*, eles precisam apresentar uma segmentação bastante

compatível com a segmentação perceptual feita pelos segmentadores. Em razão do problema de superestimação dos parâmetros de pausa física, foi necessário elaborar uma estratégia diferente para detectar automaticamente as fronteiras prosódicas. Uma forma de permanecer observando a percepção dos segmentadores e buscar reavaliar os modelos desenvolvidos é diferenciar as fronteiras a partir de outro critério também de base perceptual. O mais relevante nesse sentido é a percepção auditiva de pausa.

Os objetivos da análise sobre a percepção de pausa são: i) contribuir com a melhor compreensão sobre o fenômeno; ii) buscar um refinamento interno às macrocategorias de fronteiras terminais e não-terminais em correspondência com a percepção humana; iii) desenvolver submodelos, que possam ser aplicados aos dados de forma hierárquica, para identificar automaticamente subgrupos de fronteiras terminais e não-terminais percebidas; iv) buscar solucionar o problema da superestimação dos parâmetros de pausa física nos modelos de identificação automática das fronteiras. Foi realizado um teste de percepção auditiva de pausa por um grupo de 14 anotadores. A tarefa de percepção de pausa foi muito útil e forneceu dados para investigar os quatro aspectos citados.

Quanto ao fenômeno de percepção auditiva de pausa, este trabalho contribuiu nas discussões acerca da duração mínima de pausa física que pode ser percebida como pausa pelos ouvintes, evitando confusões com os períodos de silêncio relativo às consoantes oclusivas. A duração mínima aqui proposta é 100 ms. O trabalho foi além e apresentou outros elementos que podem explicar a percepção de pausa dos ouvintes.

Os resultados mostraram que os fatores são diferentes a depender da fronteira associada à pausa física. As variáveis significativas exclusivas do modelo destinado à percepção de pausa em posição de TB com pausa física são a duração da coda antes da pausa física, a presença de coda [w], /R/ ou /S/ antes da pausa física e a presença do reset de f0. A única variável significativa exclusiva do modelo destinado à percepção de pausa em posição de disfluência com pausa física é a taxa de articulação, que apresenta um coeficiente negativo e indica uma correlação inversamente proporcional entre a taxa de articulação e a saliência da pausa percebida. O modelo de percepção de pausa em posição de NTB com pausa física não possui nenhuma variável exclusiva.

Por isso, são estimadas as variáveis relevantes para a percepção de pausa em posição de NTB com pausa física em comparação com os modelos de percepção de pausa em posição de TB e disfluência com pausa física. Entre NTB e TB, os parâmetros distintivos são a duração da coda antes da pausa física, a duração da consoante após a pausa física, a presença de coda [w], /R/ ou /S/ antes da pausa física, o reset de f0, a duração da vogal antes da pausa física e a

presença de consoante vozeada após a pausa física. Entre NTB e disfluência, os parâmetros distintivos são a taxa de articulação e a duração da consoante após a pausa física.

O refinamento interno às macrocategorias de fronteiras foi feito com base na duração mínima de pausa percebida que foi proposta neste trabalho. As fronteiras foram subdivididas em: i) TB com pausa percebida; ii) TB sem pausa percebida; iii) NTB com pausa percebida; iv) NTB sem pausa percebida. Foram desenvolvidos uma série de submodelos para identificar automaticamente os quatro subgrupos de fronteiras prosódicas. A ideia consistia em desenvolver submodelos para identificar fronteiras com e sem pausa percebida para que eles fossem aplicados aos dados seguindo uma ordenação hierárquica predeterminada. Foram avaliados dois métodos de diferentes de ordenação hierárquica de submodelos com o objetivo de compreender qual método funcionaria melhor na ferramenta de identificação de fronteiras.

A primeira proposta é denominada ordenação hierárquica com base em quatro etapas interligadas. Com a ordenação hierárquica de submodelos em discussão, inicialmente, é necessário que as pausas percebidas a partir de 100 ms sejam identificadas automaticamente. Nas posições de pausa percebida identificadas automaticamente, o submodelo FRO-compausapercebida deve ser aplicado para identificar as fronteiras TB e NTB com pausa percebida. Nas posições remanescentes, o submodelo POS-sempausapercebida é aplicado em seguida para identificar as posições sem fronteira e as fronteiras sem pausa percebida, mas, neste momento, o submodelo não informa se as fronteiras sem pausa percebida identificadas compreendem TB ou NTB. Por último, o submodelo FRO-sem pausapercebida é aplicado nas posições de fronteira sem pausa percebida previamente reconhecidas pelo submodelo POS-sempausapercebida com objetivo de identificar se estas posições são TB ou NTB sem pausa percebida.

A segunda proposta é chamada de ordenação hierárquica de submodelos com base em três etapas interligadas. Neste caso, as pausas percebidas de pelo menos 100 ms também devem ser identificadas primeiramente. Em seguida, o submodelo FRO-compausapercebida é aplicado apenas nas posições de pausa percebida reconhecidas com o objetivo de identificar as posições de TB e NTB acompanhadas por pausa percebida. Após o reconhecimento das fronteiras com pausa percebida, um único submodelo deve ser aplicado nas posições remanescentes para identificar TB sem pausa percebida, NTB sem pausa percebida e posição sem fronteira.

O processo inicial de reconhecimento das fronteiras com pausa percebida é comum às propostas de ordenação hierárquica. As propostas se diferem quanto à análise das posições sem fronteira e posições de fronteira sem pausa percebida. Na primeira proposta, o reconhecimento de TB e NTB sem pausa percebida depende diretamente do desempenho do submodelo POS-

sempausapercebida, o que pode ser considerado uma desvantagem. Na segunda proposta, o reconhecimento de TB e NTB sem pausa percebida não depende de outro submodelo, mas, neste caso, o submodelo precisa identificar ao mesmo tempo as posições sem fronteira, TB e NTB sem pausa percebida, o que pode ser considerado uma tarefa de classificação mais complexa.

Na fase de simulação, o submodelo de identificação de fronteiras com pausa percebida (FRO-compausapercebida), comum às propostas de ordenação hierárquica, reconheceu mais de 70% das fronteiras TB e NTB com pausa percebida. Também na fase simulação, o desempenho do submodelo POS-sempausapercebida, necessário apenas na proposta de ordenação hierárquica de submodelos com base em quatro etapas, pode ser considerado intermediário. O submodelo POS-sempausapercebida identificou 60% das fronteiras sem pausa percebida e 56% das posições sem fronteira foram identificadas. Portanto, 40% das fronteiras sem pausa percebida deixaram de ser reconhecidas e 44% das fronteiras sem pausa percebida marcadas não são "autênticas" fronteiras sem pausa percebida.

Ainda com relação à fase de simulação, nas posições que realmente são fronteiras sem pausa percebida e haviam sido devidamente identificadas pelo submodelo POS-sempausapercebida, o submodelo FRO-sempausapercebida apresentou um desempenho intermediário, identificando 94% das fronteiras terminais sem pausa percebida e 60% das fronteiras não-terminais sem pausa percebida. Nas posições que foram marcadas indevidamente como fronteiras sem pausa percebida pelo submodelo POS-sempausapercebida, o submodelo FRO-sempausapercebida acabou marcando estas posições como TB ou NTB sem pausa percebida. Estas marcações incorretas têm um impacto muito grande na *performance* geral da proposta de ordenação hierárquica de submodelos com base em quatro etapas interligadas, pois não é possível confiar que as fronteiras TB e NTB sem pausa percebida marcadas ao término das predições são realmente TB e NTB sem pausa percebida.

Na fase de desenvolvimento, o submodelo de identificação de fronteiras sem pausa percebida (FRO-sempausapercebida) identificou 100% das posições de TB sem pausa percebida e 76% das posições de NTB sem pausa percebida. Na fase de validação, há uma queda nas porcentagens de reconhecimento. Respectivamente 71% e 60% das fronteiras TB e NTB sem pausa percebida foram identificadas. Por se tratar de uma diminuição de 29% na porcentagem de reconhecimento de TB sem pausa percebida, os resultados sugerem que a quantidade de dados usados na fase de desenvolvimento do submodelo FRO-sempausapercebida é insuficiente para desenvolver um submodelo com alto poder de generalização.

A constatação de que a quantidade de dados, possivelmente, foi insuficiente para o desenvolvimento do submodelo FRO-sempausapercebida não anula a simulação de resultados a serem alcançados por meio da proposta de ordenação hierárquica de submodelos com base em quatro etapas interligadas. Neste sentido, a capacidade do submodelo FRO-sempausapercebida reconhecer TB e NTB sem pausa percebida pode ser considerada intermediária, caso o submodelo em questão esteja analisando fronteiras sem pausa percebida devidamente reconhecidas pelo submodelo anterior denominado POS-sempausapercebida.

Com o processo de retreinamento do submodelo FRO-sempausapercebida a partir de uma maior quantidade de dados, o resultado esperado é o submodelo em questão apresentar um maior poder de generalização e passar a reconhecer melhor ainda as fronteiras TB e NTB sem pausa percebida marcadas corretamente como fronteira sem pausa percebida pelo submodelo POS-sempausapercebida. Obviamente, mesmo com o processo de retreinamento do submodelo FRO-sempausapercebida a partir de uma maior quantidade de dados, o submodelo FRO-sempausapercebida permanece dependendo diretamente do desempenho do submodelo POS-sempausapercebida. Desta forma, para a proposta de ordenação hierárquica de submodelos com base em quatro interligadas funcionar como um todo, é importante que o submodelo POS-sempausapercebida também seja retreinado a partir de uma maior quantidade de dados com o objetivo de melhorar as suas predições e garantir que o novo submodelo FRO-sempausapercebida realmente analise posições de fronteira sem pausa percebida.

Na fase de simulação, o submodelo GER-sempausapercebida, usado apenas na proposta de ordenação hierárquica de submodelos com base em três etapas com o objetivo de identificar ao mesmo tempo posições sem fronteira, TB e NTB sem pausa percebida, reconheceu bem apenas as posições sem fronteira prosódica (70% de reconhecimento). A capacidade deste submodelo identificar TB e NTB sem pausa percebida foi baixa. Respectivamente 54,2% e 28% das posições de TB e NTB sem pausa percebida foram devidamente reconhecidas. Adicionalmente, os resultados mostraram que diversas TB e NTB sem pausa marcadas pelo submodelo GER-sempausapercebida compreendem falsos alarmes, isto é, são marcações erradas. Por isso, o submodelo destinado à identificação de TB e NTB sem pausa percebida na proposta de ordenação hierárquica de submodelos com base em três etapas interligadas também precisa de ajustes.

Em geral, o submodelo designado para reconhecer TB e NTB com pausa percebida apresentou resultados satisfatórios, reduziu os efeitos indesejados causados pelos parâmetros de pausa física e representa um avanço com relação aos modelos diversos para identificar TB e NTB. A identificação automática das fronteiras TB e NTB sem pausa percebida precisa ser

aperfeiçoada, pois os submodelos designados para identificar TB e NTB sem pausa percebida nas duas propostas de ordenação hierárquica tiveram dificuldade para diferenciar TB sem pausa percebida de NTB sem pausa percebida.

Embora os submodelos desenvolvidos tenham as suas limitações, eles também auxiliam a compreender as combinações de parâmetros fonético-acústicos relevantes para distinguir determinadas posições de outras. No caso do submodelo de identificação das fronteiras com pausa percebida (FRO-compausapercebida), os parâmetros mais relevantes para diferenciar TB e NTB com pausa percebida são a diferença de inclinação média de f0 dentro das janelas dos dois lados da fronteira, a inclinação média de f0 na unidade pós-fronteiriça e a diferença de variação de duração normalizada média dos dois lados da fronteira. No âmbito do submodelo de identificação de fronteiras sem pausa percebida (FRO-sempausapercebida), os parâmetros relevantes para diferenciar TB e NTB sem pausa percebida são a duração da pausa percebida, a duração normalizada média na penúltima unidade antes da fronteira e a diferença de variação das medianas de f0 dos dois lados da fronteira. No âmbito do submodelo de identificação de posição sem fronteira, TB e NTB sem pausa percebida (GER-sempausapercebida), os parâmetros relevantes para diferenciar estas três posições são a duração normalizada média na unidade pós-fronteiriça, a diferença da duração normalizada entre as unidades pós-fronteiriça e fronteiriça, a duração normalizada média na unidade fronteiriça e a duração da pausa percebida.

Neste trabalho, a posição dos parâmetros que podem estar envolvidos na percepção das fronteiras foi analisada através dos diversos modelos e submodelos desenvolvidos. Considerando todos os modelos e submodelos, os parâmetros observados para diferenciar uma determinada posição de outra se concentram frequentemente na unidade fronteiriça e nas unidades imediatamente adjacentes às fronteiras. Neste sentido, os parâmetros que podem estar envolvidos na percepção das fronteiras ocorrem principalmente nas posições em que os segmentadores marcaram a percepção das fronteiras e em posições próximas às posições de atribuição de percepção das fronteiras.

O submodelo de identificação de fronteiras com pausa percebida pode ser testado no estado atual. Porém, neste momento, não é possível determinar qual método de ordenação hierárquica de submodelos apresenta resultados melhores de modo geral. Naturalmente, a escolha do método de ordenação hierárquica está sujeita ao aperfeiçoamento dos submodelos necessários para identificar as fronteiras sem pausa percebida. Por isso, as perspectivas futuras deste trabalho são aumentar os dados e treinar novamente, conforme as diferentes propostas de ordenação hierárquica, os submodelos necessários para identificar automaticamente as fronteiras sem pausa percebida.

Tanto o teste a ser realizado antes da implementação, quanto a implementação em si da ferramenta de identificação de fronteiras prosódicas, serão feitos através do software Praat. Nos áudios, duas formas diferentes de identificar automaticamente as fronteiras são possíveis. As predições podem ser feitas diretamente nos áudios sem fornecer nenhum tipo de tratamento aos áudios e as predições podem ser feitas através da análise dos áudios em conjunto com a segmentação em unidades V-V. Evidentemente, no caso da última predição, é necessário o trabalho prévio de segmentação das unidades V-V. Este trabalho prévio pode ser feito rapidamente através do uso de *scripts* de segmentação do sinal sonoro em *onsets* de vogais consecutivos como o BeatExtractor (Barbosa, 2004) e o VowelDetector (Arantes, 2021). A forma de realizar as predições será escolhida com base nos resultados obtidos nos testes iniciais.

REFERÊNCIAS

AGNELLO, J. *A study of intra- and inter-phrasal pauses and their relationship to the rate of speech*. Dissertação, Ohio State University, 1963.

AMIR, N.; SILVER-VAROD, V. e IZRE´EL, S. Characteristics of intonation unit boundaries in spontaneous spoken Hebrew – perception and acoustic correlates. *ISCA Archive: Speech Prosody*, Nara, 2004, p.677-680.

ANANTHAKRISHNAN, S., NARAYANAN, S. An automatic prosody recognizer using a coupled multi-stream acoustic model and a syntactic-prosodic language model. *Proc. Int. Conf. Acoust., Speech, Signal Process*, 2005. p. 269–272.

ANANTHAKRISHNAN, S., NARAYANAN, S. Automatic prosodic event detection using acoustic, lexical, and syntactic evidence. *IEEE Trans Audio, Speech Lang Process*, 2008, p. 216-228.

ARANTES, P. VowelDetector (versão 2.0) [Programa de computador], 2021. Disponível em: https://github.com/parantes/vowel-detector . Acesso em: 25/05/22.

ARVANITI, A.; BALTAZANI, M. Greek ToBI: A System For The Annotation Of Greek Speech Corpora. In: *Proceedings of the Second International Conference on Language Resources and Evaluation*. 2000, p. 555-562.

AUER, P. Zum Segmentierungsproblem in der gesprochenen Sprache, 2010. Disponível em: http://www.inlist.uni-bayreuth.de/issues/49/InList49.pdf . Acesso em: 01/02/21.

AURAN, C., BOUZON, C.; HIRST, D. The AixMARSEC project: an evolutive database of spoken English. In *Proceedings of Speech Prosody 2004*, Nara, 2004, p. 561–564.

AVANZI, M., LACHERET-DUJOUR, A., VICTORRI, B. ANALOR: A Tool for Semi-Automatic Annotation of French Prosodic Structure: ANALOR, In *Proc. speech prosody*. 2008, p.119-122.

BARBOSA, P. *Caractérisation et génération automatique de la structuration rythmique du français*. Tese de Doutorado, Institut National Polytecnique de Grenoble, 1994.

BARBOSA, P. BeatExtractor for Praat. [Programa de computador]. 2004. Disponível em: https://github.com/pabarbosa/prosody-scripts/tree/master/BeatExtractor . Acesso em: 15/07/2022.

BARBOSA, P. *Incursões em torno do ritmo da fala*. Campinas: Pontes, 2006.

BARBOSA, P. Prominence-and boundary-related acoustic correlations in Brazilian Portuguese read and spontaneous speech. In *Proc. speech prosody*. 2008, p. 257–260.

BARBOSA, P. BreakDescriptor (Versão 2.0) [Programa de computador], 2019. Disponível em: https://github.com/pabarbosa/prosody-scripts/tree/master/BreakDescriptor . Acesso em: 15/07/2022.

BARIK, C. Cross-linguistic study of temporal characteristics of different speech materials. *Language and Speech 20*, 1977, p. 16- 126.

BARNWELL, B. *The perceptual basis of intonation units*. Manuscrito, 2011.

BARTH-WEINGARTEN, D. *Intonation Units Revised: Cesuras in talk-in-interaction.* Philadelphia: John Benjamins Publishing Company, 2016.

BELZ, M.; TROUVAIN, J. Are 'silent' pauses always silent? *Proc. International Congress of Phonetic Sciences (ICPhS)*, Melbourne, 2019.

BIRKNER, K. Konstruktionen zur Personenattribuierung. In: GÜNTHNER, S., WOLFGANG, I. *Konstruktionen in der Interaktion*. Berlim: Mouton de Gruyter, 2006, p. 205-238.

BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer*, [Programa de computador], 2015. Disponível em: https://www.fon.hum.uva.nl/praat/ . Acesso em: 16/01/2015.

BOERSMA, M. Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America*, 117(6), 2005, p. 3890-3901.

BOLINGER, D. Around the edges of language. In: BOLINGER, D. (ed.)., *Intonation: Selected Readings*. Harmondsworth. Penguin: 1972, p. 19-29.

BOOMER, S. Hesitation and grammaticai encoding. *Language and Speech* 8, 1965, p.148-158.

BRAZIL, D. *The communicative value of intonation*. Cambridge: CUP, 1985.

BRAZIL, D.; COULTHARD, M.; JOHNS, C. *Discourse Intonation and Language Teaching*. Londres: Longman, 1980.

BRENNAN, R.; PREDIGER, D. Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, v.41, 1981, p. 687−699.

BRIERLEY, C.; ATWELL, E. An approach for detecting prosodic phrase boundaries in spoken English. *ACM Crossroads journal*, vol. 14, 2007, p. 1-11.

BUHMANN, J.; CASPERS, J.; SWERTS, M. Annotation of prominent words, prosodic boundaries and segmental lengthening by no-expert transcribers in the spoken Dutch corpus. *Proceedings of LREC 2002*, 2002. p. 779–85.

BUTCHER, A. Aspects of the speech pause: *Phonetic correlates and communicative functions.* Aipuk (Arbeitsberichte Institut für Phonetik Kiel), Dissertação, 1981.

BYRD, D.; SATZMAN, E. The elastic phrase: Modeling the dynamics of boundary adjacent lengthening. *Journal of Phonetics* 31, 2003, p.149–180.

CAMPBELL, N. Automatic detection of prosodic boundaries in speech. *Speech communication*. North-Holland: Elsevier Science Publishers, 1993, p. 343-354.

CAMPIONE, E.; VÉRONIS, J. A Large-Scale Multilingual Study of Silent Pause Duration, *Proceedings of the Speech Prosody 2002 conference*, 2002, p. 199–202.

CARLETTA, J. Assessing agreement on classification tasks: The Kappa statistic. *Computational Linguistics*, 22(2), 1996, p. 249–254.

CARLSON, R., HIRSCHBERG, J., SWERTS, M. Cues to upcoming Swedish prosodic boundaries: Subjective judgment studies and acoustic correlates. *Speech Communication*, 2005, p. 326–333.

CARLSON, K., How prosody influences sentence comprehension. *Lang. Linguist. Compass* 3, 2009, p.1188–1200.

CHAFE, W. *Discourse, consciousness anda time: The Flow and dsiplacement of Conscious Experience in Speaking and writing.* Chicago: University of Chicago Press, 1994.

CHENG, W.; GREAVES, C.; WARREN, M. The creation of prosodically transcribed intercultural corpus: The Hong Kong Corpus of Spoken English (prosodic), *ICAME Journal*, vol. 29, 2005, p. 47-68.

CHO, T., MCQUEEN, M. Phonological versus phonetic cues in native and nonnative listening: Korean and Dutch listeners' perception of Dutch and English Consonants. *Journal of the Acoustical Society of America*, 119(5), 2006, p. 3085-3096.

COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* (20), p, 37-46, 1960.

COLE, J.; MO, Y.; BAEK, S. The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, 25(7), 2010, p. 1141–1177.

COUPER-KUHLEN, E. *An Introduction to English Prosody*. Tübingen: Niemeyer, 1986.

COUPER-KUHLEN, E. Prosodic Cues of Discourse Units, *Encyclopedia of Language & Linguistics*, 2006, p. 178–182.

COWAN, N. *Attention and memory: An integrated framework*. Oxford: Oxford University Press, 1998.

COWAN, N. Exploring the possible and necessary in working memory development. *Monographs Society Res Child*, 81, 2016, p. 149–158.

CRESTI, E.; MONEGLIA, M. (Orgs.). *C-ORAL-ROM: Integrated reference corpora for spoken Romance Languages.* Amsterdam/Philadelphia: John Benjamins, 2005.

CRESTI, E. *Corpus di Italiano parlato*. v. 1. Firenze: Accademia della Crusca, 2000.

CROFT, W. Intonation Units and Grammatical Structure. *Linguistics* 33 (5), 1995, p. 839-882, 1995.

CROFT, W. Intonation units and grammatical structure in Wardaman in crosslinguistic perspective. *Australian Journal of Linguistics* 27, 2007, p. 1-39.

CRUTTENDEN, A. *Intonation*. 2ed. Cambridge: CUP, 1997.

CRYSTAL, D. Intonation and Linguistic Theory, *The Nordic Languages and Modern Linguistics 2: Proceedings of the Second International Conference of Nordic and General Linguistics*, no 2, 1975, p. 267–303.

CRYSTAL, D. *Prosodic Systems and Intonation in English*. Cambridge: CUP, 1969.

CUTLER, A., DAHAN, D. VAN DONSELAAR, W. Prosody in the comprehension of spoken language: a literature review, *Lang. Speech*, vol. 40, no 2, 1997, p. 141–201.

DEGAND, L.; SIMON, A. On identifying basic discourse units in speech: theoretical and empirical issues. *Discours*, 2009, p. 1–17.

DILLEY, L.; SHATTUCK-HUFNAGEL, S.; OSTENDORF, M. Glottalization of word initial vowels as a function of prosodic structure. *Journal of Phonetcs*, 24(4), 1996, p. 423–444

DRURY, E.; BAUM, R.; VALERIOTE, H.; STEINHAUSER, K. Punctuation and implicit prosody in silent reading: An ERP study investigating English garden-path sentences. *Frontiers in Psychology*, 7, 2016, p. 1-12.

DU BOIS, J.; CUMMING, S.; SHUETZE-COBURN, S.; PAOLINO, D. *Discourse Transcription*. Santa Barbara: Department of Linguistics, University of California, 1992.

DU BOIS, J.; CHAFE, W.; MEYER, C.; THOMPSON, S.; ENGLEBRETSON, R.; MARTEY, N. *Santa Barbara corpus of spoken American English*, Parts 1-4. Philadelphia: Linguistic Data Consortium, 2000-2005.

DUEZ, D. Perception of silent pauses in continuous speech. *Lang Speech*, v. 28, 1985, p. 377-389.

ESCUDERO, P. *Linguistic Perception and Second Language Acquisition: Explaining the attainment of optimal phonological categorization.* Tese de Doutorado. Universidade de Utrecht, 2005.

ESCUDERO, P.; BENDERS, T.; LIPSKI, C. Native, nonnative and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, 37, 2009, p.452-465.

FLEISS, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, Vol. 76, No. 5, p. 378–382, 1971.

FON, J.; JOHNSON K.; CHEN, S. Durational patterning at syntactic and discourse boundaries in Mandarin spontaneous speech. *Lang Speech*, 2011, p. 5–32.

FORD, E. Contingency and Units in Interaction. *Discourse Studies* 6 (1), 2004, p.27-52.

FRAZIER, L.; CARLSON, K. CLIFTON, C. Prosodic phrasing is central to language comprehension. Trends in Cognitive Sciences, nº 10, 2006, p. 244–249.

FROTA, S. The intonational phonology of European Portuguese. In S. Jun (ed.). *Prosodic Typology II*. Oxford: Oxford University Press, 2014, p. 6-42

FUCHS, S.; KRIVOKAPIĆ, J.; JANNEDY, S. Prosodic boundaries in German: Final lengthening in spontaneous speech. *The Journal of the Acoustical Society of America*, 127(3), 2010, 1851. https://doi.org/10.1121/1.3384378

FUCH S., KRIVOKAPIĆ, J., JANNEDY, S. Prosodic boundaries in German: Final lengthening in spontaneous speech. *J Acoust Soc Am*. 2010.

GALVES, C. *Ensaios sobre as gramáticas do português*. Campinas: Editora da Unicamp, 2001.

GLUSHKO, A.; STEINHAUER, K.; DEPRIEST, J.; KOELSCH, S. Neurophysiological correlates of musical and prosodic phrasing: Shared processing mechanisms and effects of musical expertise. *PLoS ONE*, 11(5), 2016, p. 1-27.

GOLDMAN, J. EasyAlign: an automatic phonetic alignment tool under Praat. *Proceedings of InterSpeech*. Firenze: 2011.

GOLDMAN-EISLER, F. *Psycholinguistics: Experiments in spontaneous speech*. Academic Press: Londres, 1968.

GORDON, M.; LADEFOGED, P. Phonation types: A cross-linguistic overview. *Journal of Phonetics*, 29, 2001, p. 383–406.

GOTTFRIED, L.; BEDDOR, S. Perception of temporal and spectral information in French vowels. *Language and Speech*, 31, p. 57-75.

GRICE, M.; REYELT, M.; BENZMÜLLER, R.; MAYER, J.; BATLINER, A.; Consistency in Transcription and Labelling of German Intonation with GToBI In*: Proc. ICSLP 96 Philadelphia*, 1996.

GROSS, D.; ALLEN, J.; TRANM, R. The TRAINS 91 dialogues. *Technical Report TN92-1*, Department of Computer Science, Universidade de Rochester, 1993.

GUSSENHOVEN, C. *The phonology of tone and intonation*. Nova York: Cambridge University Press, 2004.

HALFORD, K. The discourse function of intonation. In: HALFORD, K.; PILCH, H. (eds.). *Intonation*. Tübingen: Nara, 1994, p. 69-88.

HALFORD, K. Talk Units. *The structure of spoken Canadian English*. Tübingen: Nara, 1996.

HALLIDAY, M. *Intonation and grammar in British English*. The Hague: Mouton, 1967.

HALLIDAY, M. *An introduction to functional grammar*. Londres: Arnold, 1985.

HANSON, M.; STEVENS, N.; KUO, J.; CHEN, Y.; SLIFKA, J. Towards model of phonation. *Journal of Phonetics*, 29, 2001, p. 451–480.

Hargreaves, A.; Starkweather, A. Collection of temporal data with duration tabulator. *J. Exp. Anal. Behav*. 2: 179, 1959.

HARNSBERGER, D. The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis. *Journal of Phonetic*s, 29, 2001, p. 303-327.

HARRIS, O., UMEDA, N., BOURNE, J. Boundary perception in fluent speech. *Journal of Phonetics*, 1981. p. 1–18.

HELDNER, M. Detection thresholds for gaps, overlaps, and no-gap-no-overlaps. *The Journal of the Acoustical Society of America*, 130(1), 2011, p.508–513.

HENDERSON, A.; GOLDMAN-EIDER, F.; SKARBEK, A. Sequential temporal patterns in spontaneous speech. *Language and Speech 9*, 1966, p. 207-216.

HERMES, J. Stylization of pitch contours. In SUDHOFF, D.; LENERTOVÁ, R.; MEYER, S.; PAPPERT, P.; AUGURZKY, I.; MLEINEK, R.; RICHTER, N.; Schliesser, J. (Eds.), *Methods in empirical prosody research*. Berlin: De Gruyter, p. 29-61.

HIEKE, E.; KOWAL, S.; O'CONNELL, C. The trouble with "articulatory pauses". *Language and Speech* 26, 1983, p. 203-214.

HIRSCHMAN, L.; BATES, M.; DAHL, D.; FISHER, W., GAROFOLO, J., PALLETT, D.; HUNICKE-SMITH, K.; PRICE, P.; RUDNICKY, A.; TZOUKERMAN, E. In MARCUS, M.; KAUFMANN, M. (ed.), Multi-Site Data Collection for a Spoken Language Corpus, *Proc. Fifth Speech and Natural Language Workshop*, Nova York: Arden House, 1993.

HOFHUIS, J.; GUSSENHOVEN, C.; RIETVELD, T. Final lengthening at prosodic boundaries in Dutch. In ELENIUS, E.; BRANDERUD, P . (Eds.), *Proceedings of the XIII international congress of phonetic sciences*, Vol. 1, 1995, p. 154–157.

HOLT, L.; LOTTO, J.; KLUENDER, R. Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America*, 109, 2001, p.764-774.

HWANG, H.; STEINHAUER, K. Phrase length matters: The interplay between implicit prosody and syntax in Korean ‘garden path’ sentences. *Journal of Cognitive Neuroscience*, 23, 2011, p.3555–3575.

IZRE'EL, S.; BOIS, J.; ARIEL, M. *The corpus of spoken Israeli Hebrew (CoSIH)*, 2009.

IZRE’EL, S. Intonation Units and the Structure of Spontaneous Spoken Language: A View from Hebrew. In: AURAN, C. et al. *Proceedings of the IDP05 International Symposium on Discourse-Prosody Interfaces*. Aix-en-Provence; 2011. p. 1–20.

IZRE'EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. In search of a baisc unit of spoken language: Segmenting speech. In IZRE’EL S.; MELLO, H.; PANUNZI, A.; RASO, T. (eds). *In search of basic units of spoken language: A corpus-driven approach.* Amsterdam: John Benjamins, 2020.

JUN, S.; FOUGERON, C. A phonological model of French intonation. In BOTINIS, A. (ed.) *Intonation: Analysis, Modeling and Technology*, Dordrecht: Kluwer Academic Publishers, 2000, p. 209-242.

JUN, S.; LEE, S.; KIM, K.; LEE, Y. Labeler agreement in transcribing korean intonation with K-ToBI. In *Interspeech'10*, 2010, p. 211–214.

JUSCZYK, W. From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21, 1993, p. 3-28.

KENDALL, M. *Rank correlation methods*. Londres: Griffin, 1948.

KENDALL, T. *Speech rate, pause, and linguistic variation: an examination through the Sociolinguistic Archive and Analysis Project*. Tese de Doutorado, NC: Duke University, 2009.

KIBRIK, A.; PODLESSKAYA, I. *Rasskazy o snovidenijah: korpusnoe issledovanie ustnogo russkogo diskursa [Night dream stories: A corpus study of spoken Russian discourse]*. Moscow: Languages of Slavonic Culture, 2009.

KJELGAARD, M.; SPEER, R. Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40, 1999, p. 153–194.

KLUENDER, R.; LOTTO, J.; HOLT, L.; BLOEDEL, B. Role of experience in language-specific functional mappings for vowel sounds as inferred from human, nonhuman, and computational models. *Journal of the Acoustical Society of America*, 104, 1998, p. 3568-3582.

KOMPE, R.; KIESSLING, A.; NIEMANN, H.; NOTH, E.; BATLINER, A.; SCHACHTT, S.; RULAND, T.; BLOCK, H. Improving parsing of spontaneous speech with the help of prosodic boundaries. *IEEE Int Conf Acoust Speech, Signal Process*, 1997, p. 811-814.

KOWAL, S.; O'CONNELL, C. Some temporal aspects of stories told while or after watching a film. *Bulletin of the Psychunomic Society*, v. 25, 1987, p. 364 366.

KOWAL, S.; O'CONNELL, C.; SABIN, E. Development of temporal patterning and vocal hesitation in spontaneous narratives. *Journal of Psycholinguistic Research*, v. 4, 1975, p. 195-207.

KRALJIC, T.; BRENNAN, E. Prosodic disambiguation of syntactic structure: For the speaker or for the addressee?. *Cognitive Psychology*, v. 50, 2005, p. 194–231.

KRIVOKAPIĆ, J. Prosodic planning: Effects of phrasal length and complexity on pause duration. *J Phon*. 2007, p. 162–179.

LACHERET, A.; OBIN, N.; AVANZI, M. Design and evaluation of shared prosodic annotation for spontaneous french speech: from expert knowledge to non-expert annotation. In *Proceedings of the 4th linguistic annotation workshop*, 2010, p. 265–273.

LADD, R. Declination reset and the hierarchical organization of utterances. *Journal of the Acoustical Society of America.* v.84, 1988, p. 530–544.

LADD, R. Declination: a review and some hypotheses, *Phonology Yearbook* 1, 1984, p.53-74.

LADD, R. *Intonational phonology*. Cambridge: CUP, 1996.

LAY, H.; PAIVIO, A. The effects of rask difficulty and anxiety on hesitation in speech. Canadian *Journal of Behavioral Science*, 1969, p. 25-37.

LEVIN, H.; SILVERMAN, I. Hesitation phenomena in children's speech. *Language and Speech*, 8, 1965, p. 67-85.

LEVIN, H.; SILVERMAN, I.; FORD, L. Hesitation in children's speech during explanation and description. *Journal of Verbal Leaming and Verbal Behavior* , v. 6, 1967, p.560-564.

LI, A.; LIN, M.; CHEN, X.; ZU, Y.; SUN, G.; HUA, W.; YIN, Z.; YAN, J. Speech corpus of Chinese discourse and the phonetic research, In *ICSLP-2000*, vol.4, 2000 p.13-18.

LISKER, L. Closure duration and the intervocalic voiced voiceless distinction in English. *Language*, v. 33, 1957, p. 42–49.

LUDUSAN, B.; DUPOUX, E. Towards Low Resource Prosodic Boundary Detection. In *Proceedings of International Workshop on Spoken Language Technologies for Under-resourced Languages* (SLTU'14), 2014, p. 231-237.

MARCUS, S. Acoustic Determinants of Perceptual-Center (P-Center) Location. *Perception and Psychophysics*, 30(3). 1981, p. 247-256.

MARTIN, J. On judging pauses in simultaneous speech. *Journal of Verbal Learning and Verbal Behavior* , v.9, 1970, p. 75-78.

MARTIN, J.; STRANGE, W. The perception of hesitation in spontaneous speech. *Perception and Psychophysics*, 1968, p. 427-438.

MARTIN, P. Les problèmes de l'intonation: Recherches et applications. *Langue Francaise*, v. 19, 1973, p. 4–32.

MASCHLER. Y. *Metalanguage in Interaction: Hebrew Discourse Markers*. Amsterdam: John Benjamins, 2009.

MCGUIRE, G. *Phonetic Category Learning*. Tese de Doutorado. The Ohio State University, 2007.

MELLO, H.; RASO, T. (Ed.) *C-ORAL – Brasil I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG, 2012.

MELLO, H.; RASO, T.; MITTMANN, M.; VALE, H.; CÔRTES, P. Transcrição e segmentação prosódica do corpus C-ORAL-BRASIL: critérios de implementação e validação. In: MELLO, H.; RASO, T. (Ed.) *C-ORAL – Brasil I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG, 2012, p.

MELLO, H.; RASO, T.; FERRARI, L. (Ed.) *C-ORAL – Brasil II: Corpus de referência do português brasileiro falado informal*, em preparação.

MERTENS, P. SIMON, A. Towards Automatic Detection of Prosodic Boundaries in Spoken French. In: MERTENS, P. SIMON, A. *Proceedings of the Discourse-Prosody Interface Conference*, 2013, p. 81-87.

MERTENS, P. Intonational grouping, boundaries and syntactic structure in French. In HOUSE, D.; TOUATI, P. (eds). *Proceedings. ESCA Workshop on Prosody*, 1993, p. 156-159.

METTOUCHI, A.; CHANARD, C. From Fieldwork to Annotated Corpora: the CorpAfroAs Project. *Faits de Langue-Les Cahiers*. n. 2, 2010, p. 255-265,

MO, Y.; COLE, J. Perception of Prosodic Boundaries in Spontaneous Speech with and Without Silent Pauses. *The Journal of the Acoustical Society of America* , v. 127, n. 3, 2010.

MO, Y., COLE, J., LEE. E-K. Naïve listeners' prominence and boundary perception. In: BARBOSA, P. A., MADUREIRA S., REIS C., Speech Prosody. Campinas: ISCA, 2008, p. 739-742.

MO, Y. Duration and intensity as perceptual cues for naïve listeners' prominence and boundary perception. *Proceedings of Speech Prosody*. Campinas: Brasil, 2008, p.739- 742.

MONEGLIA, M.; CRESTI, E. L'intonazione e i criteri di transcrizione del parlato adulto e infantile. In: BORTOLINI, U.; PIZZUTO, E. (Ed.). *Il Progetto CHILDES Italia*. Pisa: Del Cerro, 1997. p. 57-90.

MONEGLIA, M.; FABBRI, M.; QUAZZA, S.; ANDREA; PANIZZA, A.; DANIELI, M; GARRIDO, J. M.; SWERTS, M. Evaluation of Consensus on the Annotation of Terminal and Non-Terminal Prosodic Breaks in the C-ORAL-ROM corpus. In: CRESTI, E.; MONEGLIA, M. (Ed.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins, 2005. p. 257-276.

MUKHERJEE, J. *Form and Function of Parasyntactic Presentation Structures. A Corpus-based Study of Talk Units in Spoken English*. Atlanta: Rodopi, 2001.

NEBOT, A. Sobre prosodia, marcadores del discurso y unidades del discurso en español: Evidencias de un corpus oral espontáneo. *Onomazein*, 2013, p. 201–213.

NESPOR, M.; VOGEL, I. *Prosodic phonology*. Dordrecht: Foris, 1986.

NI, CJ., ZHANG, AY., LIU, WJ., XU, B. Automatic prosodic break detection and feature analysis. *J. Comput. Sci. Tchol*., 27, 2012, p. 1184-1196.

NICKELS, S.; OPITZ, B.; STEINHAUER, K. ERPS show that classroom-instructed late second language learners rely on the same prosodic cues in syntactic parsing as native speakers. *Neuroscience Letters*, 557, 2013, p.107–111.

OLIVEIRA Jr., M., *Prosodic Features in Spontaneous Narratives*. Tese de Doutorado, Simon Fraser University, Vancouver, 2000.

ONO, T.; THOMPSON, S. What can conversation tell us about syntax? In DAVIS, P. (Ed.), *Alternative linguistics: Descriptive and theoretical modes*. Netherlands: John Benjamins, 1995, p. 213-271.

OOSTDIJK, N. The Spoken Dutch Corpus. Overview and first evaluation. In *Proc of LREC-2000*, Vol. 2, 2000, p. 887-894.

OSTENDORF, M., VEILLEUX, N. A hierarchical stochastic model for automatic prediction of prosodic boundary location. *Comput Linguist*., 1994, p. 27–54.

OSTENDORF, M.; PRICE, P.; SHATTUCK-HUFNAGEL, S. The Boston University Radio News Corpus, *Boston University Technical Report*, No. ECS-95-001,1995.

PALMER, H. *English intonation, with systematic exercises*. Cambridge: Heffer, 1922.

PARK, J. Cognitive and interactional motivations for the intonation unit. *Studies in Language*, 26, 2002, p. 637–680.

PAUKER, E.; ITZHAK, I.; BAUM, S.; STEINHAUER, K. Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the boundary deletion hypothesis. *Journal of Cognitive Neuroscience*, 23(10), 2011, p. 2731–2751.

PAUL, D.; BAKER, J. The Design for the Wall Street Journal-based CSR Corpus, *Proceedings DARPA Speech and Natural Language Workshop*, 1992, p. 357-362.

PETERS, B.; KOHLER, K.; WESENER, T. Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In: KOHLER, K.; KLEBER, F.; PETERS, B. (eds.). *Prosodic Structures in German Spontaneous Speech (AIPUK 35)*, Kiel: IPDS, 2005, p.143-184.

PIERREHUMBERT, J. *Phonetics and phonology of English intonation*. Dissertação. Massachusetts Institute of Technology, 1980.

PIJPER, J.; SANDERMAN, A On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America*, 96(4), 1994, p. 2037 2047.

PIKE, L. *The intonation of American English*. Ann Arbor: University of Michigan Press, 1945.

PITRELLI, F.; BECKMAN, E; HIRSCHBERG, J. Evaluation of Prosodic Transcription Labelling Reliability in the ToBI Framework, *Proceedings of the 1994 International Conference on Spoken Language Processing* (PLACE), 1994, p. 123-126.

POMPINO-MARSCHALL, B. On the Psychoacoustic Nature of the P-Center Phenomenon. *Journal of Phonetics* (17), 1989, p. 175-192.

PRICE, M. OSTENDORF, S., SHATTUCK-HUFNAGEL, FONG, C. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 90(6), 1991, p. 2956-2970.

R DEVELOPMENT CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna/ Austria: R Foundation for Statistical Computing, 2017.

RASO, T. ; MITTMANN, M. M. ; MENDES, A. C. O. O papel da pausa na segmentação prosódica de corpora de fala. *Revista de Estudos da Linguagem*, v. 23, 2015, p. 883-922.

RASO, T.; TEIXEIRA, B.; BARBOSA, P. Modelling automatic detection of prosodic boundaries for Brazilian Portuguese spontaneous speech. *Journal of Speech Sciences*, vol. 9, 2020, p. 105-128.

RASO, T. O corpus C-ORAL-BRASIL. In MELLO, H.; RASO, T. (Ed.) C-ORAL – Brasil I: *Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG, 2012, P. 55-90.

REICHEL, D.; MÁDY, K. Comparing parameterizations of pitch register and its discontinuities at prosodic boundaries for Hungarian, *INTERSPEECH-2014*, 2014, p. 111-115.

REYELT, M.; GRICE, M.; Benzmüller, R.; Mayer, J.; Batliner, A.Prosodische Etikettierung des Deutschen mit ToBI. In: GIBBON, D. (ed.) *Natural Language and Speech Technology*. Berlin: Mouton de Gruyter, 1996.

ROSS, K.; ONTENDORF, M. Prediction of abstract prosodic labels for speech synthesis. *Computer Speech and Language*, 1996, p. 155-185.

ROSSI, M. L'intonation: Le système du français : description et modélisation. *Collection L'essentiel français.* Gap: Ophrys, 1999.

RUDER, K. Duration of silent interval as a perceptual cue of speech pauses. *Perceptual and Motor Skills*, 36, 1973, p. 47-57.

SABIN, E. *Pause and Rate Phenomena in Adult Narratives*. Saint Louis University: Science Research, 1976.

SABIN, E.; CLEMMER, E. J.; O'CONNELL, D. C.; KOWAL, S. A pausological approach to speech development. In: SIEGMAN, A.; FELDSTEIN, S. *Of Time and Speech*. Hillsdale: Lawrence Erlbaum, 1979. p.35-55.

SCHAFER, J.; SPEER, R.; WARREN, P. Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In: TANENHAUS, M.; TRUESWELL, J. (Eds.) *Approaches to Studying World Situated Language Use: Psycholinguistic, Linguistic and Computational Perspectives on Bridging the Product and Action Tradition.* Cambridge: MIT Press, 2005, p. 209-225.

SCHEGLOFF, E. Reflections on Studying Prosody in Talk-in-interaction. *Language and Speech* 41 (3-4), 1998, p. 235-263.

SCHUETZE-COBURN, S., SHAPLEY, M., WEBER., G. Units of intonation in discourse: a comparison of acoustic and auditory analyses. *Lang Speech*. 1991, p. 207–234.

SELKIRK, E. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge: MIT Press, 1984.

SELKIRK, E. Sentence prosody: Intonation, Stress and Phrasing. In: GOLDSMITH, J. (Ed.). *The Handbook of Phonological Theory*. Oxford: Blackwell, 1995. p. 550-569.

SELKIRK, E. Comments on intonational phrasing in English. In: FROTA, S.; VIGARIO, M.; FREITAS, M. (eds.) Prosodies. *With special reference to Iberian languages*. Berlin: Mouton de Gruyter, 2005, p. 11-58.

SELMINI, A. M.; VIOLARO, F. Segmentação Automática de Fala para o Português Brasileiro. In: *XXVI Simpósio Brasileiro de Telecomunicações (Anais do XXVI Simpósio Brasileiro de Telecomunicações)*, Rio de Janeiro, 2008.

SELTING, M. The Construction of Units in Conversational Talk. *Language in Society*, 29 (4), 2000, p. 477–517.

SHRIBERG, E., STOLCKE A., HAKKANI-TUR, G. Prosody-based automatic segmentation of speech into sentence and topics. *Speech Commun*, 2000, 32, p. 127–154.

SHRIBERG, E. Spontaneous speech: How people really talk and why engineers should care, *9th Eur. Conf. Speech Commun. Technol*., 2005, p. 1781–1784.

SHRIBERG, E.; STOLCKE, A.; HAKKANI-TÜR, D.; TÜR, G. Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, 32(1–2), 2000, p. 127–154.

SIEGMAN, A.; POPE, B. Ambiguity and verbal fluency in the TAT. *Journal of Consulting Psychology*, v. 30, 1966, p. 239-245.

SILBER-VAROD, V. *The SpeeCHain perspective: Form and function of prosodic boundary tones in spontaneous spoken Hebrew*. Saarbrücken: LAP Lambert Academic Publishing, 2013.

SILVERMAN, K.; BECKMAN, M., PITRELLI, J. et al. TOBI: A standard for labeling English prosody. *International Conference on Speech and Language Processing* (ICSLP), v. 2, 1992. p. 867-870.

SIMON, A.; CHRISTODOULIDES, G. Perception of prosodic boundaries by naïve listeners in French. *Proc. Speech Prosody*, 2016, p. 1158-1162

SNEDEKER, J., TRUESWEEL, J. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48, 2003, p. 103–130.

SOSA, J. La entonación del español. Madrid: Cátedra, 1999.

SPEER, S.; KJELGAARD, M.; DOBROTH, K. The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 1996, p.249-271.

STEINHAUER, K.; FRIEDERICI, A. Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, 30(3), 2001, p. 267–295.

STEINHAUER, K. Electrophysiological correlates of prosody and punctuation. *Brain and Language*, 86, 2003, p. 142–164.

STELMA, J.; CAMERON, L. Intonation Units in Spoken Interaction: Developing Transcription Skills. *Text & Talk* 27 (3), 2007, p. 361–393.

STETSON, R. Motor phonetics: *A Study of Speech Movements in Action*. Oberlin: Springer Netherlands, 1928.

STROM, V. Detection of accents, phrase boundaries and sentence modality in German with prosodic features. *Fourth European Conference on Speech Communication and Technology*, Madrid: ISCA; 2006. p. 2039–2042.

SVENDSEN, T.; SOONG, F. On the automatic segmentation of speech signals. *Ieee international conference on acoustics, speech, and signal processing - icassp '87*, Dallas, 1987. Conferences. Dallas, US: Institute of Electrical & Electronics Engineers (IEEE), 1987. p. 77-80.

SWERTS M., GELUYKENS, R. Prosody as a marker of information flow in spoken discourse. *Lang Speech.* 1994, v. 37, p. 21–43.

SWERTS, M., GELUYKENS, R., TERKEN, J. Prosodic correlates of discourse units in spontaneous speech. P*roceedings of the International Conference on Spoken Language Processing*. 1992, p. 421–424.

SWERTS, M., COLLIER, R. TERKEN, J. Prosodic predictors of discourse finality in spontaneous monologues. *Speech Communication*, v. 15, 1994, p. 79–90.

SWERTS, M. *Prosodic Features of Discourse Units*. 1994. Thesis (PhD) – Technische Universiteit Eindhoven, 1994.

SWERTS, M. Prosodic features at discourse boundaries of different strength. *J Acoust Soc Am*, vol. 101, 1997, p. 514–521.

SYRDAL, A.; MCGORY, A. Inter-transcriber reliability of ToBI prosodic labeling, *Proceedings of the International Conference on Spoken Language Processing*, 2000, p. 235-238.

SZASZÁK, G., BEKE, A. Exploiting Prosody for Syntactic Analysis in Automatic Speech Understanding. *J Lang Model*, 2012.

SZCZEPEK REED, B. Turn-final intonation in English. In: COUPER-KUHLEN, E., FORD, C. (eds), *Sound Patterns in Interaction*. Amsterdam: Benjamins, 2004, p. 97-117.

SZCZEPEK REED, B. Intonation Phrases in Natural Conversation: A Participants Category? In BARTH-WEINGARTEN, D.; REBER, E.; SELTING, M. *Prosody in Interaction*. Amsterdam: Benjamins, 2010, p. 191–212.

TAKEUCHI, H., SUBRAMANIAM, L., ROY, S., PUNJANI, D., NASUKAWA, T. Sentence boundary detection in conversational speech transcripts using noisily labeled examples. International. *Journal on Document Analysis and Recognition*. 2007. p. 147–155.

TAYLOR, I. Content and structure in sentence production. *Journal of Verbal Learning and Verbal Behavior*, v. 8, 1969, p. 170- 175.

TEIXEIRA, B. *Correlatos fonético-acústicos de fronteiras prosódicas na fala espontânea*, Dissertação de mestrado. Universidade Federal de Minas Gerais, Belo Horizonte, 2018.

'T HART, J., COLLIER, R., COHEN, A. *A perceptual study on intonation: An experimental approach to speech melody*. Cambridge: CUP, 1990.

THORSEN, N. Intonation and text in standard Danish. *Journal of the Acoustical Society of America*, 77(3), 1985, p. 1205–1216.

THORSEN, N. Sentence intonation in textual context. *Journal of the Acoustical Society of America*, 80(4), 1986, p. 1041–1047.

TSENG C-Y., CHANG C-H. Pause or no pause?: Prosodic phrase boundaries revisited. *Tsinghua Sci Technol*, 2008, p. 500–509.

TSENG, C-Y., FU, B. Duration, Intensity and Pause Predictions in Relation to Prosody Organization. *Delta*, Lisboa: ISCA, 2005, p. 5–8.

TSENG, C-Y., PIN, SH., LEE, Y., WANG, HM., CHEN, YC. Fluent speech prosody: Framework and modeling. *Speech Communication*, 2005, p. 284–309.

TYLER, J. Prosodic correlates of discourse boundaries and hierarchy in discourse production. *Lingua*, vol. 133, 2013, p. 101–126.

VAISSIÈRE, J. Language-independent prosodic features. Prosody: Models and Measurements, *Springer Verlag*, 1983, p.53-65.

VAN HEMERT, J. P. Automatic segmentation of speech. *IEEE Transactions On Signal Processing, Piscataway*, v. 39, n. 4, Institute of Electrical & Electronics Engineers (IEEE), 1991, p. 1008-1012.

VENDITTI, JJ. *The ToBI model of Japanese intonation. ICPhS satellite workshop on Intonation: Models and ToBI Labeling*. San Francisco, California, 1999.

VERZEANO, M.; FINESINGER, J. An automatic analyzer for the study of speech in interaction and in free association. *Science*, 110, 1949, p. 45-46.

WALKER, M.; TRIMBOLI, C. Smooth transitions in conversational interactions, *J. Soc. Psychol.* 117, 1982, p. 305–306.

WARREN, P.; GRABE, E.; NOLAN, F. Prosody, phonology and closure ambiguities. *Language and Cognitive Processes*, v. 10, 1995

WATSON, G.; GIBSON, E. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19, 2004, p. 713- 755.

WATSON, G.; GIBSON, E. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19, 2004, p. 713- 755.

WHEELDON, L., LAHIRI, A. Prosodic units in speech production. *J Mem Lang*. 1997, 37, p. 356–381.

WHITE, L., WIGET, L., RAUCH, O., MATTYS, L. Segmentation Cues in Spontaneous and Read Speech. *Proc 5th Conf Speech Prosody*. 2010, p. 1–4.

WIGHTMAN, C. ToBI or not ToBI, *Proceedings of Speech Prosody 2002*, 2002, p. 25–29.

WIGHTMAN, C.; OSTENDORF, M. Automatic labeling of prosodic patterns. *IEEE Transactions on speech and audio processing*, 1994, p. 469-481.

WIGHTMAN, C., SHATTUCK-HUFNAGEL, S., OSTENDORF, M., PRICE, P. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91, p. 1707-1717, 1992.

XIE, L. Discovering salient prosodic cues and their interactions for automatic story segmentation in Mandarin broadcast news. *Multimedia Systems*, 2008, p. 237–253.

XU, Y. In Defense of Lab Speech. *Journal of Phonetics*, v. 38, n. 3, 2010, p. 329-336.

XU, Y. Speech Prosody: A Methodological Review. *Journal of Speech Sciences*, 1(1), 2011, p. 85-115.

ZELLERS, M.; POST, B. Aperiodicity at topic structure boundaries. *Proceedings of speech prosody: Fifth conference*, 2010, p. 451–480.

ZELLNER, B. Pauses and the temporal structure of speech. In: KELLER, E. (Org.) *Fundamentals of speech synthesis and speech recognition.* Chichester: John Wiley, 1994. p. 41-62.

APÊNDICES

Apêndice A: Métricas dos modelos de identificação automática de fronteiras terminais

Modelo	Fase	Accuracy	Sensitivity	Specificity
TB-nb	Des	0,93	0,94	0,84
	Val	0,94	0,96	0,75
	Rev-1	0,90	0,91	0,70
	Rev-2	0,92	0,93	0,75
TB-nbfull	Des	0,96	0,97	0,73
TB-b1	Des	0,87	0,91	0,83
	Val-1	0,92	0,92	0,82
	Val-2	0,87	0,88	0,74
	Val-3	0,89	0,90	0,79
TB-b2	Des	0,87	0,91	0,83
	Val-1	0,87	0,88	0,82
	Val-2	0,91	0,92	0,81
	Val-3	0,90	0,90	0,81
TB-bfull	Des	0,84	0,87	0,82
	Val-1	0,90	0,90	0,82
	Val-2	0,91	0,92	0,83
	Val-3	0,88	0,88	0,78
TB-b7	Des	0,90	0,94	0,86
	Val-1	0,91	0,92	0,82
	Val-2	0,90	0,91	0,70
	Val-3	0,91	0,92	0,78
TB-b3	Des	0,83	0,87	0,80
	Val-1	0,92	0,93	0,82
	Val-2	0,90	0,91	0,72
	Val-3	0,91	0,92	0,78

Apêndice B: Métricas dos modelos de identificação de fronteiras não-terminais

Modelo	Fase	Accuracy	Sensitivity	Specificity
NTB-1	Des	0,72	0,78	0,68
	Rev-1	0,70	0,72	0,66
	Rev-2	0,72	0,74	0,66
NTB-2	Des	0,62	0,78	0,48
	Rev-1	0,53	0,53	0,56
	Rev-2	0,60	0,61	0,50
NTB-3	Des	0,83	0,88	0,69
	Rev-1	0,83	0,86	0,23
	Rev-2	0,81	0,83	0,28
NTB-nbfull	Des	0,86	0,99	0,41
NTB-b1	Des	0,74	0,76	0,72
	Val-1a	0,74	0,74	0,72
	Val-1b	0,67	0,65	0,71
	Val-1c	0,71	0,70	0,72
	Val-2a	0,71	0,70	0,73
	Val-2b	0,64	0,62	0,71
	Val-2c	0,68	0,67	0,72
NTB-b2	Des	0,76	0,78	0,73
	Val-1a	0,65	0,61	0,79
	Val-1b	0,73	0,72	0,73
	Val-1c	0,68	0,66	0,76
	Val-2a	0,61	0,57	0,79
	Val-2b	0,70	0,69	0,73
	Val-2c	0,65	0,62	0,76
NTB-bfull	Des	0,74	0,93	0,55
	Val-1a	0,85	0,93	0,52
	Val-1b	0,82	0,91	0,58
	Val-1c	0,84	0,92	0,55
	Val-2a	0,82	0,89	0,52
	Val-2b	0,80	0,88	0,57
	Val-2c	0,80	0,88	0,55
NTB-b7	Des	0,73	0,97	0,50
	Val-1a	0,85	0,94	0,50
	Val-1b	0,83	0,97	0,46
	Val-1c	0,84	0,95	0,47
	Val-2a	0,82	0,89	0,50
	Val-2b	0,80	0,92	0,46
	Val-2c	0,80	0,90	0,47
NTB-b3	Des	0,74	0,93	0,54
	Val-1a	0,82	0,88	0,56
	Val-1b	0,81	0,92	0,52
	Val-1c	0,82	0,90	0,54
	Val-2a	0,79	0,84	0,57
	Val-2b	0,79	0,88	0,52
	Val-2c	0,79	0,86	0,54

Apêndice C: Métricas do submodelo comum às diferentes propostas de ordenação hierárquica de submodelos (FRO-compausapercebida)

Submodelo	Fase	Accuracy	Sensitivity	Specificity
FRO-compausapercebida	Des	0,75	0,72	0,78
	Val	0,70	0,70	0,66

Apêndice D: Métricas dos submodelos exclusivos da proposta de ordenação hierárquica de submodelos com base em quatro etapas interligadas (submodelos POS-sempausapercebida e FRO-sempausapercebida)

Submodelo	Fase	Accuracy	Sensitivity	Specificity
POS-sempausapercebida	Des	0,80	0,82	0,79
	Val	0,56	0,56	0,59
FRO-sempausapercebida	Des	0,88	0,76	1
	Val	0,60	0,60	0,71

Apêndice E: Métricas do submodelo exclusivo da proposta de ordenação hierárquica de submodelos com base em três etapas interligadas (submodelo GER-sempausapercebida)

Submodelo	Fase	Accuracy	Sensitivity / Specificity		
			TB sem pausa percebida	NTB sem pausa percebida	Posição sem fronteira
GER-sempausapercebida	Des	0,75	0,71 / 0,92	0,47 / 0,96	0,91 / 0,68
	Val	0,64	0,14 / 0,87	0,27 / 0,83	0,70 / 0,43

ANEXOS

Script para Praat BreakDescriptor (Barbosa, 2019).

```
# BreakDescriptor-2.psc

# Script implemented by Plinio A. Barbosa (IEL/Univ. of Campinas, Brazil) for computing

# prosody descriptors around breaks from coupled audio/TG files for syllable-size in 10-VV half windows

# The TextGrid, Audio and Reference-statistics (xy.TableOfReal, where xy = BP, EP, F, G, or BE) files need

# to be in the same directory!!!

# Work in collaboration with Tommaso Raso, Maryualê Mittmann and colleagues

# Copyright (C) 2016  Barbosa, P. A.

#

#    This program is free software; you can redistribute it and/or modify

#    it under the terms of the GNU General Public License as published by

#    the Free Software Foundation; version 2 of the License.

#    This program is distributed in the hope that it will be useful,

#    but WITHOUT ANY WARRANTY; without even the implied warranty of

#    MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the

#    GNU General Public License for more details.

#

# Date: Jul 2016, Jan 2019:

form File acquisition

 word FileOut OutPutParameters.txt

 word AudiofileExtension *.wav

 integer VVTier 1

 integer NTBTier 2

 integer TBTier 3

 integer RetrctTier 4

 integer IntrupTier 5

 integer PauseTier 6
```

```
 integer F0Thresholdleft 75
 integer F0Thresholdright 300
 positive Smthf0Thr 5
 positive F0step 0.05
 positive Njudges 19
 positive Criterion 0.7
 integer AnalysisWindow 10
 positive Spectralemphasisthreshold 400
 choice Reference: 1
   button BP
   button EP
   button G
   button F
   button BE
   button S
endform
# Reads the reference file with the triplets (segment, mean,
standard-deviation) from the
# reference speaker. The variable nseg contains the total number of
segments in the file
Create Strings as file list... list 'audiofileExtension$'
numberOfFiles = Get number of strings
if !numberOfFiles
     exit There are no sound files in the folder!
endif
Read from file... 'reference$'.TableOfReal
nseg = Get number of rows
filedelete 'fileOut$'
fileappend 'fileOut$' audiofile boundtime intnb tb ntb tbN ntbN rtN
intN bound type psp psdur srl srr srd arl arr ard mzl mzr mzd zdloc
SDzl SDzr SDzd skzl skzr skzd prl prr prd f0medl f0medr f0medd
f0meddloc sdf0l sdf0r sdf0d skf0l skf0r skf0d df0medl df0medr
df0medd df0meddloc sddf0l sddf0r sddf0d skdf0l skdf0r skdf0d f0rl
f0rr f0rd emphl emphr emphd emphdloc zl1 zr1 f0medl1 f0medr1
df0medl1 df0medr1 zl0 f0med0 df0med0
if analysisWindow == 1
```

```
 fileappend 'fileOut$' 'newline$'
else
for t from 2 to analysisWindow
 fileappend 'fileOut$' zl't' zr't' f0medl't' f0medr't' df0medl't'
df0medr't'
endfor
fileappend 'fileOut$' 'newline$'
endif
for ifile from 1 to numberOfFiles
 select Strings list
 audiofile$ = Get string... ifile
 Read from file... 'audiofile$'
 filename$ = selected$("Sound")
 Filter (pass Hann band)... 0 'spectralemphasisthreshold' 100
 filenamefiltered$ = selected$("Sound")
 select Sound 'filename$'
 To Pitch... 0.0 'f0Thresholdleft' 'f0Thresholdright'
 Smooth... 'smthf0Thr'
 Interpolate
 To Matrix
 To Sound (slice)... 1
 Rename... temp
 To PointProcess (extrema)... 1 yes no None
#
# Reads TextGrid
 arq$ = filename$ + ".TextGrid"
 Read from file... 'arq$'
 ntiers = Get number of tiers
 if ntiers < 6
     pauseTier = 4
 endif
 begin = Get starting time
 end = Get finishing time
```

```
 totaldur = end - begin
 nselected = Get number of intervals... 'vVTier'
 npotbreaks = Get number of points... 'tBTier'
#
############################ df0, VVwise
 tl = Get starting point... 'vVTier' 2
 tr = Get end point... 'vVTier' 2
 select Pitch 'filename$'
 #Down to PitchTier
 k = nselected -1
 df0'k'  =0
 f0ant = Get quantile... 'tl' 'tr' 0.5 semitones re 1 Hz
 f0med1 = f0ant
 df0med1 = 0
 for i from 3 to nselected -1
  select TextGrid 'filename$'
  tl = Get starting point... 'vVTier' 'i'
  tr = Get end point... 'vVTier' 'i'
  select Pitch 'filename$'
  f0current = Get quantile... 'tl' 'tr' 0.5 semitones re 1 Hz
# Computes f0 discrete derivative
  j = i-1
  f0med'j' = f0current
  df0med'j' = f0current - f0ant
  f0ant = f0current
 endfor
###################
# Reading TG and computing normalized duration in smoothed z-scores
#
 select TextGrid 'filename$'
 initialtime = Get starting point... 'vVTier' 2
 kk = 1
 nselected = nselected - 2
```

```
 for i from 1 to nselected
  adv = i + 1
  nome$ = Get label of interval... 'vVTier' 'adv'
  itime = Get starting point... 'vVTier' 'adv'
  ftime = Get end point... 'vVTier' 'adv'
  dur = ftime - itime
  dur = dur*1000
  tint = Get starting point... 'vVTier' 'adv'
  call zscorecomp 'nome$' 'dur' 'tint'
  dur'i' = dur
  z'i' = z
  nome'i'$ = nome$
  select TextGrid 'filename$'
  adv = i + 1
 endfor
# for i from 1 to nselected
 smz1 = (2*z1 + z2)/3
 deriv1 = smz1
 smz2 = (2*z2 + z1)/3
 deriv2 = smz2 - smz1
 i = 3
 if smz1 < smz2
  minsmz = smz1
  maxsmz = smz2
 else
  minsmz = smz2
  maxsmz = smz1
 endif
 while i <= (nselected-2)
  del1 = i - 1
  del2 = i - 2
  adv1 = i + 1
  adv2 = i + 2
```

```
 smz'i' = (5*z'i' + 3*z'del1' + 3*z'adv1' + z'del2' + 1*z'adv2')/13
 deriv'i' = smz'i' - smz'del1'
 if smz'i' < minsmz
  minsmz = smz'i'
 endif
 if smz'i' > maxsmz
  maxsmz = smz'i'
 endif
 i = i + 1
endwhile
tp1 = nselected -1
tp2 = nselected -2
smz'tp1' = (3*z'tp1'+ z'tp2' + z'nselected')/5
deriv'tp1' = smz'tp1' - smz'tp2'
 if smz'tp1' < minsmz
  minsmz = smz'tp1'
 endif
 if smz'tp1' > maxsmz
  maxsmz = smz'tp1'
 endif
smz'nselected' = (2*z'nselected' + z'tp1')/3
deriv'nselected' = smz'nselected' - smz'tp1'
 if smz'nselected' < minsmz
  minsmz = smz'nselected'
 endif
 if smz'nselected' > maxsmz
  maxsmz = smz'nselected'
 endif
tpp = nselected + 2
time = initialtime
boundcount = 0
# loop for 2
for i from 1 to nselected
```

```
  tempsmz = smz'i'
  tpnome$ = nome'i'$
  adv1 = i + 1
  btime'i' = 0
  time = time + dur'i'/1000
  time'i' = time
  if i <> nselected
   adv1 = i + 1
   if (deriv'i' >= 0) and (deriv'adv1' < 0)
     boundary'i' = 1
     boundcount = boundcount + 1
     btime'i' = time
     bctime'boundcount' = time
   else
     boundary'i' = 0
   endif
  else
   del1 = i -1
   if smz'i' > smz'del1'
      boundary'i' = 1
      boundcount = boundcount + 1
      btime'i' = time
      bctime'boundcount' = time
   else
     boundary'i' = 0
   endif
  endif
endfor
# end loop for 2
# Computing parameters in the two analysisWindow-VV-long half
windows
endVV = nselected- analysisWindow + 1
for x from 1 to npotbreaks
```

```
  select TextGrid 'filename$'
  timeVV = Get time of point... 'tBTier' 'x'
  indexVV = Get interval at time... 'VVTier' 'timeVV'
  iVV = indexVV - 1
  if iVV > analysisWindow and iVV < endVV
# Agreement judges
    tbtime = Get time of point... 'tBTier' 'x'
    tb20 = tbtime + 0.02
    tb$ = Get label of point... 'tBTier' 'x'
    tb = number(tb$)
    ntb$ = Get label of point... 'nTBTier' 'x'
    ntb = number(ntb$)
if ntiers > 5
    rt$ = Get label of point... 'retrctTier' 'x'
    rt = number(rt$)
    int$ = Get label of point... 'intrupTier' 'x'
    int = number(int$)
endif
    indpause = Get interval at time... 'pauseTier' 'tb20'
    pause$ = Get label of interval... 'pauseTier' 'indpause'
    if pause$ == "p" or pause$ == "P" or pause$ == "p " or pause$ ==
"P "
           prespause = 1
        durpause = Get end point... 'pauseTier' 'indpause'
           startpause = Get starting point... 'pauseTier' 'indpause'
           durpause = durpause - startpause
    else
           prespause = 0
        durpause = 0
    endif
# computing SR/AR/mz/f0med in the left half-window
    indleft = iVV - analysisWindow
    lefttime = Get starting point... 'VVTier' 'indleft'
```

```
    indright = iVV
    righttime = Get starting point... 'VVTier' 'indright'
    ileftlastwindowunit = iVV-1
    leftlastwindowunit = Get starting point... 'VVTier'
'ileftlastwindowunit'
    srl = analysisWindow/(righttime - lefttime)
    sdur = 0
    sz = 0
    prl = 0
    f0medl = 0
    df0medl = 0
    nunstr = 10
    for l from indleft to iVV-1
        if boundary'l' == 0
           sdur = sdur + dur'l'
     else
           nunstr = nunstr - 1
                prl = prl  + 1
     endif
        sz = sz + smz'l'
        f0medl = f0medl + f0med'l'
        df0medl = df0medl + df0med'l'
        ll = l - indleft + 1
        zl'll' = smz'l'
        f0medl'll' = f0med'l'
        df0medl'll' = df0med'l'
    endfor
    zl11=smz'l'
    f0medl11 = f0med'l'
    df0medl11 =  df0med'l'
    arl = 1000*nunstr/sdur
    mzl = sz/analysisWindow
    f0medl = f0medl/analysisWindow
```

```
    df0medl = df0medl/analysisWindow
    prl = 1000*prl/sdur
# computing left z SD/f0med sD/df0med sD
    sdzl= 0
    sdf0l = 0
    sddf0l = 0
    for l from indleft to iVV-1
     sdzl = sdzl + (smz'l'  - mzl)*(smz'l'  - mzl)
        sdf0l =  sdf0l + (f0med'l'  - f0medl)*(f0med'l'  - f0medl)
        sddf0l =  sddf0l + (df0med'l'  - df0medl)*(df0med'l'  -
df0medl)
    endfor
    sdzl = sqrt(sdzl/(analysisWindow-1))
    sdf0l =  sqrt(sdf0l/(analysisWindow-1))
    sddf0l =  sqrt(sddf0l/(analysisWindow-1))
# computing left z skew/f0med skew/df0med skew
   skzl = 0
   skf0l = 0
   skdf0l = 0
   for l from indleft to iVV-1
       skzl = skzl + ((smz'l'  - mzl)/sdzl)^3
       skf0l = skf0l + ((f0med'l'   - f0medl)/sdf0l)^3
       skdf0l = skdf0l + ((df0med'l'   - df0medl)/sddf0l)^3
   endfor
   skzl = (analysisWindow/((analysisWindow-1)*(analysisWindow-
2)))*skzl
   skf0l = (analysisWindow/((analysisWindow-1)*(analysisWindow-
2)))*skf0l
   skdf0l = (analysisWindow/((analysisWindow-1)*(analysisWindow-
2)))*skdf0l
# Computing left f0rate
   select PointProcess temp
   f0rl = Get mean period... 'lefttime' 'righttime' 0.3 3 5
   f0rl = 1/f0rl
```

```
# Computing Left Spectral Emphasis
   select Sound 'filename$'
   allenergy = Get energy... 'lefttime' 'righttime'
   allenergylast = Get energy... 'leftlastwindowunit' 'righttime'
   select Sound 'filenamefiltered$'
   filtenergy = Get energy... 'lefttime' 'righttime'
   filtenergylast = Get energy... 'leftlastwindowunit' 'righttime'
   emphl = allenergy - filtenergy
   emphlastl = allenergylast - filtenergylast
# Computing SR/AR/mz in the right half-window
    select TextGrid 'filename$'
    indleft = iVV
    lefttime = Get end point... 'VVTier' 'indleft'
    indright = iVV + analysisWindow
    righttime = Get end point... 'VVTier' 'indright'
    irightlastwindowunit = iVV+1
    rightlastwindowunit = Get end point... 'VVTier'
'irightlastwindowunit'
    srr = analysisWindow/(righttime - lefttime)
    srd = srr - srl
    sdur = 0
    sz = 0
    f0medr = 0
    df0medr = 0
    prr = 0
    nunstr = analysisWindow
    for l from iVV+1 to indright
        if boundary'l' == 0
           sdur = sdur + dur'l'
     else
           nunstr = nunstr - 1
           prr = prr  + 1
     endif
```

```
        sz = sz + smz'l'
        f0medr = f0medr + f0med'l'
        df0medr = df0medr + df0med'l'
        ll = l - iVV
        zr'll' = smz'l'
        f0medr'll' = f0med'l'
      df0medr'll' = df0med'l'
    endfor
    arr = 1000*nunstr/sdur
    mzr = sz/analysisWindow
    f0medr = f0medr/analysisWindow
    df0medr = df0medr/analysisWindow
    prr = 1000*prr/sdur
#computing right z SD/f0med sD/df0med sD
    sdzr= 0
    sdf0r = 0
    sddf0r = 0
    for l from iVV+1 to indright
     sdzr = sdzr + (smz'l'  - mzr)*(smz'l'  - mzr)
     sdf0r =  sdf0r + (f0med'l'  - f0medr)*(f0med'l'  - f0medr)
        sddf0r =  sddf0r + (df0med'l'  - df0medr)*(df0med'l'  -
df0medr)
    endfor
    sdzr = sqrt(sdzr/(analysisWindow-1))
    sdf0r =  sqrt(sdf0r/(analysisWindow-1))
    sddf0r =  sqrt(sddf0r/(analysisWindow-1))
# computing right z skew/f0med skew/df0med skew
   skzr = 0
   skf0r = 0
   skdf0r = 0
   for l from  iVV+1 to indright
       skzr = skzr + ((smz'l'  - mzr)/sdzr)^3
       skf0r = skf0r + ((f0med'l'   - f0medr)/sdf0r)^3
```

```
        skdf0r = skdf0r + ((df0med'l'   - df0medr)/sddf0r)^3
   endfor
   skzr = (analysisWindow/((analysisWindow-1)*(analysisWindow-
2)))*skzr
   skf0r = (analysisWindow/((analysisWindow-1)*(analysisWindow-
2)))*skf0r
   skdf0r = (analysisWindow/((analysisWindow-1)*(analysisWindow-
2)))*skdf0r
   ard = arr - arl
   mzd = mzr - mzl
   zdloc = zr1 - zl11
   sdzd = sdzr - sdzl
   skzd = skzr - skzl
   prd = prr - prl
   f0medd = f0medr - f0medl
   df0medd = df0medr - df0medl
   f0meddloc = f0medr1 - f0medl11
   df0meddloc = df0medr1 - df0medl11
   sdf0d = sdf0r - sdf0l
   sddf0d = sddf0r - sddf0l
   skf0d = skf0r - skf0l
   skdf0d = skf0r - skdf0l
# Computing right f0rate
   select PointProcess temp
   f0rr = Get mean period... 'lefttime' 'righttime' 0.3 3 5
   f0rr = 1/f0rr
   f0rd = f0rr - f0rl
# Computing right Spectral Emphasis
   select Sound 'filename$'
   allenergy = Get energy... 'lefttime' 'righttime'
   allenergyfirst = Get energy... 'lefttime' 'rightlastwindowunit'
   select Sound 'filenamefiltered$'
   filtenergy = Get energy... 'lefttime' 'righttime'
   filtenergyfirst = Get energy... 'lefttime' 'rightlastwindowunit'
```

```
    emphr = allenergy - filtenergy
    emphfirstright = allenergyfirst - filtenergyfirst
    emphd = emphr - emphl
    emphdloc = emphfirstright - emphlastl
        type$ = "NA"
        if tb >= njudges*criterion
            tb$ = "YES"
           type$ = "TB"
        else
           tb$ = "NO"
        endif
        if ntb >= njudges*criterion
               ntb$ = "YES"
            type$ = "NTB"
        else
            ntb$ = "NO"
        endif
if ntiers > 5
         if rt >= njudges*criterion
               rt$ = "YES"
            type$ = "RET"
        else
            rt$ = "NO"
        endif
        if int >= njudges*criterion
               int$ = "YES"
            type$ = "INT"
        else
            int$ = "NO"
        endif
else
    rt$ = "NO"
     int$ = "NO"
```

```
endif
   if ntb$ = "NO" and tb$ = "NO" and rt$ = "NO" and int$ = "NO"
       bound$ = "NO"
   else
       bound$ = "YES"
   endif
   fileappend 'fileOut$' 'filename$' 'tbtime:3' 'x' 'tb$' 'ntb$'
'tb' 'ntb' 'rt' 'int' 'bound$' 'type$' 'prespause' 'durpause:3'
'srl:2' 'srr:2' 'srd:2' 'arl:2' 'arr:2' 'ard:2' 'mzl:2' 'mzr:2'
'mzd:2' 'zdloc:2' 'sdzl:2' 'sdzr:2' 'sdzd:2' 'skzl:2' 'skzr:2'
'skzd:2' 'prl:2' 'prr:2' 'prd:2' 'f0medl:0' 'f0medr:0' 'f0medd:0'
'f0meddloc:0' 'sdf0l:2' 'sdf0r:2' 'sdf0d:2' 'skf0l:2' 'skf0r:2'
'skf0d:2' 'df0medl:2' 'df0medr:2' 'df0medd:2' 'df0meddloc:2'
'sddf0l:2' 'sddf0r:2' 'sddf0d:2' 'skdf0l:2' 'skdf0r:2' 'skdf0d:2'
'f0rl:2' 'f0rr:2' 'f0rd:2' 'emphl:5' 'emphr:5' 'emphd:5'
'emphdloc:5' 'zl1:2' 'zr1:2' 'f0medl1:0' 'f0medr1:0' 'df0medl1:2'
'df0medr1:2' 'zl11:2' 'f0medl11:0' 'df0medl11:0'
   if analysisWindow == 1
          fileappend 'fileOut$' 'newline$'
   else
          for t from 2 to analysisWindow
                 zl = zl't'
                 zr = zr't'
                 f0medl = f0medl't'
                 f0medr = f0medr't'
                 df0medl = df0medl't'
                 df0medr = df0medr't'
             fileappend 'fileOut$' 'zl:2' 'zr:2' 'f0medl:0'
'f0medr:0' 'df0medl:2' 'df0medr:2'
          endfor
          fileappend 'fileOut$' 'newline$'
   endif
endif
endfor
# Loop:  for x from 1 to npotbreaks
endfor
# All files
```

```
select all
Remove
##
procedure zscorecomp nome$ dur tint
 sizeunit = length (nome$)
 sumofmeans = 0
 sumofvar = 0
 cpt = 1
 while cpt <= sizeunit
  nb = 1
  terminate = 0
  seg$ = mid$(nome$,cpt,1)
  if cpt < sizeunit
#    if phoneticAlphabet$ = "Other"
     if reference$ = "BP" or reference$ = "EP"
      if mid$(nome$,cpt+1,1) == "h"  or mid$(nome$,cpt+1,1) == "N"
         nb = nb + 1
         seg$ = seg$ + mid$(nome$,cpt+1,1)
      endif
      if (cpt+nb <= sizeunit)
       tp$ = mid$(nome$,cpt,1)
       call isvowel 'tp$'
       if ((mid$(nome$,cpt+nb,1) = "I")  or  (mid$(nome$,cpt+nb,1)
= "U"))  and truevowel
         seg$ = seg$ + mid$(nome$,cpt+nb,1)
         nb= nb+1
       endif
      endif
     endif
     if reference$ = "F"
       if mid$(nome$,cpt+1,1) == "h"  or mid$(nome$,cpt+1,1) == "N"
or mid$(nome$,cpt+1,1) == "x"
         nb = nb + 1
```

```
          seg$ = seg$ + mid$(nome$,cpt+1,1)
       endif
      endif
      endif
     else
       if mid$(nome$,cpt+1,1) == "~"
          nb = nb + 1
          seg$ = seg$ + mid$(nome$,cpt+1,1)
       endif
       if (cpt+nb <= sizeunit)
        tp$ = mid$(nome$,cpt,1)
        call isvowel 'tp$'
        if ((mid$(nome$,cpt+nb,1) = "j")  or  (mid$(nome$,cpt+nb,1)
= "w"))  and truevowel
          seg$ = seg$ + mid$(nome$,cpt+nb,1)
          nb= nb+1
        endif
       endif
#    endif
  endif
  j = 1
  select all
  tableID = selected ("TableOfReal")
  select 'tableID'
  while (j <= nseg) and  not terminate
     label$ =  Get row label... 'j'
     if seg$ = label$
          terminate = 1
          mean = Get value... 'j' 1
          sd      = Get value... 'j' 2
          sumofmeans = mean + sumofmeans
          sumofvar= sd*sd + sumofvar
     endif
```

```
     j = j+1
  endwhile
  if not terminate
   exit Didn't find phone 'seg$' at 'tint'. Pls check the file
TableOfReal
  endif
  cpt= cpt+nb
 endwhile
z = (dur - sumofmeans)/sqrt(sumofvar)
endproc
procedure isvowel temp$
 truevowel = 0
 if temp$ = "i" or temp$ = "e"  or temp$ = "a"  or temp$ = "o"  or
temp$ = "u" or temp$ = "I" or temp$ = "E"
    ...or temp$ = "A"  or temp$ = "y" or temp$ = "O"  or temp$ = "U"
or temp$ = "6"  or temp$ = "@"
    ...or temp$ = "2" or temp$ = "9" or temp$ = "Y"
    truevowel = 1
 endif
endproc
```