

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

BÁRBARA HELOHÁ FALCÃO TEIXEIRA

**CORRELATOS FONÉTICO-ACÚSTICOS DE FRONTEIRAS
PROSÓDICAS NA FALA ESPONTÂNEA**

BELO HORIZONTE

2018

BÁRBARA HELOHÁ FALCÃO TEIXEIRA

**CORRELATOS FONÉTICO-ACÚSTICOS DE FRONTEIRAS
PROSÓDICAS NA FALA ESPONTÂNEA**

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Linguística teórica e descritiva

Área de concentração: Linguística Teórica e Descritiva.

Linha de Pesquisa: Estudos Linguísticos Baseados em Corpora.

Orientador: Prof. Dr. Tommaso Raso.

Co-Orientador: Prof. Dr. Plínio Almeida Barbosa.

Belo Horizonte

Faculdade de Letras da UFMG

2018

Ficha catalográfica elaborada pelos Bibliotecários da Biblioteca FALE/UFMG

T266c Teixeira, Bárbara Helohá Falcão.
 Correlatos fonético-acústicos de fronteiras prosódicas na fala espontânea [manuscrito] / Bárbara Helohá Falcão Teixeira. – 2018.
 174 f., enc. : il., tabs.
 Orientador: Tommaso Raso.
 Coorientador: Plínio Almeida Barbosa.
 Área de concentração: Linguística Teórica e Descritiva.
 Linha de Pesquisa: Estudos Linguísticos Baseados em Córpora.
 Inclui CD contendo exemplos de áudio.
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Faculdade de Letras.
 Bibliografia: f. 159-165.
 Anexos: f. 166-174.

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



FOLHA DE APROVAÇÃO

Correlatos fonético-acústicos de fronteiras prosódicas na fala espontânea

BÁRBARA HELOHÁ FALCÃO TEIXEIRA

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

Aprovada em 20 de fevereiro de 2018, pela banca constituída pelos membros:


Prof(a). Tommaso Raso - Orientador
UFMG


Prof(a). Plinio Almeida Barbosa - Coorientador
UNICAMP


Prof(a). João Antonio de Moraes
UFRJ


Prof(a). Rui Rothe Neves
UFMG

Belo Horizonte, 20 de fevereiro de 2018.

AGRADECIMENTOS

Profs. Tommaso Raso e Plínio Barbosa.

Maryualê, Marcelo e Lucia.

Prof. Heliana Mello.

Colegas do LEEL.

Mãe e pai.

Carlos!

RESUMO

Este trabalho apresenta propostas de modelos de detecção automática de fronteiras prosódicas, baseadas em múltiplos parâmetros fonético-acústicos, para possibilitar a construção de uma ferramenta computacional capaz de segmentar a fala espontânea em unidades entonacionais. Uma amostra de sete trechos de fala espontânea monológica foi segmentada em unidades entonacionais por quatorze segmentadores treinados. As fronteiras prosódicas percebidas pelos segmentadores foram marcadas como fronteiras prosódicas terminais e não terminais. Desenvolveu-se um *script* do *software* de análise acústica Praat para extrair uma série de parâmetros fonético-acústicos ao longo do sinal sonoro. Dois classificadores estatísticos, *Random Forest* (RF) e *Linear Discriminant Analysis* (LDA), foram utilizados para gerar modelos com subconjuntos de parâmetros fonético-acústicos, que poderiam funcionar como preditores de fronteiras prosódicas. A avaliação inicial dos classificadores mostra que ambos os classificadores apresentam sucesso relativo na detecção automática de fronteiras. O classificador LDA apresentou uma maior porcentagem de acerto na previsão de fronteiras prosódicas. Por isso, os modelos obtidos por meio do LDA foram refinados. O modelo final destinado à detecção automáticas de fronteiras terminais apresenta uma convergência de 80% em relação às fronteiras identificadas pelos segmentadores na amostra de fala. Para as fronteiras não terminais, foram obtidos três modelos de classificação. A soma das quantidades de fronteiras identificadas pelos três modelos corresponde à uma convergência de 98% em relação às fronteiras marcadas pelos segmentadores.

Palavras-chave: segmentação da fala; fronteiras prosódicas; fala espontânea.

ABSTRACT

The present study presents proposals for automatic prosodic boundary detection models, based on multiple phonetic-acoustic parameters, in order to enable the construction of a computational tool capable of segmenting spontaneous speech into intonation units. A seven-segment sample of spontaneous monological speech was segmented into intonation units by 14 trained annotators. The prosodic boundaries perceived by the annotators were marked as terminal and non-terminal prosodic boundaries. A Praat acoustic analysis software script was developed to extract a series of phonetic-acoustic parameters along the sound signal. Two statistical classifiers, Random Forest (RF) and Linear Discriminant Analysis (LDA), were applied to generate models comprising subsets of phonetic-acoustic parameters, which could function as prosodic boundary predictors. An initial evaluation of the classifiers indicated that both display relative success in automatic boundary detection, although the LDA classifier presented a higher percentage of accuracy in predicting prosodic boundaries. Therefore, the models obtained through the LDA were refined. The final model for the automatic detection of terminal boundaries presents an 80% convergence in relation to the boundaries identified by the annotators for the speech sample. Concerning the non-terminal boundaries, three classification models were obtained. The sum of the number of boundaries identified by the three models corresponds to a 98% convergence in relation to the boundaries marked by the annotators.

Keywords: speech segmentation; prosodic boundaries; spontaneous speech.

LISTA DE FIGURAS

Figura 1: Exemplo de segmentação do sinal acústico em fones, sílabas, palavras e unidades entonacionais	22
Figura 2: Exemplo de fronteiras prosódicas com e sem pausa silenciosa	29
Figura 3: Exemplos de fronteiras (não terminal e terminal) seguidas por pausa	40
Figura 4: Exemplo de fronteiras não terminais não seguidas por pausas	41
Figura 5: Reset de f0 entre unidades marcadas por fronteiras prosódicas.....	43
Figura 6: Exemplo de segmentação do sinal acústico em unidades VV utilizando transcrição fonética larga com caracteres ASCII	45
Figura 7: De cima para baixo: forma de onda, espectrograma de banda larga e camadas de anotação no Praat para trecho de fala espontânea. Destaca-se exemplo de camadas de anotação do Praat.	53
Figura 8: Janelas de trabalho do script	55
Figura 9: De cima para baixo: forma de onda, espectrograma de banda larga e camadas de anotação no Praat para trecho de fala espontânea. Destacam-se as janelas usadas para análise de fronteira.....	56
Figura 10: De cima para baixo: forma de onda, espectrograma de banda larga e camadas de anotação no Praat para trecho de fala espontânea. Destacam-se as janelas usadas para análise de correlato de fronteira do ponto central.....	58
Figura 11: Trecho (1).....	62
Figura 12: Trecho (2).....	63
Figura 13: Trecho (3).....	64
Figura 14: Medidas de taxa de elocução e articulação para a fronteira terminal destacada no trecho (1) – srl, srr, arr e arl.....	65
Figura 15: Duração bruta das sílabas fonéticas e duração normalizada das unidades VVs... ..	68
Figura 16: Duração normalizada e suavizada das unidades VVs.....	69
Figura 17: Medidas locais de duração normalizada para a fronteira terminal destacada no trecho (1)	70
Figura 18: Medidas globais de duração normalizada para a fronteira terminal destacada no trecho (1) – mzl, m zr, SDzl, SDzr, skzl, skzr.....	71
Figura 19: Medidas de duração normalizada para a fronteira terminal destacada no trecho (1) - mzl	72
Figura 20: Medidas de duração normalizada para a fronteira terminal destacada no trecho (1) - m zr.....	73
Figura 21: Distribuição assimétrica positiva	74
Figura 22: Distribuição assimétrica negativa	74
Figura 23: Medidas globais de duração normalizada do segmento para a fronteira terminal destacada no trecho (1) – skzl.....	75
Figura 24: Medidas globais de duração normalizada do segmento para a fronteira terminal destacada no trecho (1) – skzr	76
Figura 25: Medidas locais de f0 para a fronteira terminal destacada no trecho (2)	78
Figura 26: Medidas globais de f0 para a fronteira terminal destacada no trecho (2) - f0medl, sdf0l, skf0l, f0medr, sdf0r e skf0r.	80
Figura 27: Medidas globais de f0 para a fronteira terminal destacada no trecho (2) - f0medl.....	80

Figura 28: Medidas globais de f_0 para a fronteira terminal destacada no trecho (2) – f_0 medr.	81
Figura 29: Medidas globais de f_0 para a fronteira terminal destacada no trecho (2) – skf_0l ...	82
Figura 30: Medidas globais de f_0 para a fronteira terminal destacada no trecho (2) – skf_0r ..	82
Figura 31: Medidas locais de f_0 para a fronteira terminal destacada no trecho (2) – Derivadas discretas	84
Figura 32: Medidas globais das primeiras derivadas das medianas de f_0 para a fronteira terminal destacada no trecho (2) – df_0medl , $sddf_0l$, $skdf_0l$, df_0medr , $sddf_0r$ e $skdf_0r$	86
Figura 33: Medidas globais das primeiras derivadas das medianas de f_0 para a fronteira terminal destacada no trecho (2) - df_0medl	87
Figura 34: Medidas globais das primeiras derivadas das medianas de f_0 para a fronteira terminal destacada no trecho (2) - df_0medr	87
Figura 35: Medidas globais das primeiras derivadas das medianas de f_0 para a fronteira terminal destacada no trecho (2) – $skdf_0l$	88
Figura 36: Medidas globais das primeiras derivadas das medianas de f_0 para a fronteira terminal destacada no trecho (2) – $skdf_0r$	89
Figura 37: Medidas de intensidade - Ênfase espectral para a unidade R1 localizada após a fronteira terminal destacada no trecho (3)	91
Figura 38: Medidas de pausa silenciosa destacadas no trecho (3) – psp e $psdur$	94
Figura 39: Clusters do Modelo 1	136
Figura 40: Clusters do Modelo 2	137
Figura 41: Clusters do Modelo 3	138

LISTA DE TABELAS

Tabela 1: Número de fronteiras terminais e não terminais coincidentes com pausas	42
Tabela 2: Amostra I	50
Tabela 3: Fronteiras prosódicas	60
Tabela 4: Parâmetros fonético-acústicos - Medidas de taxa de elocução (speech rate) e ritmo	95
Tabela 5: Parâmetros fonético-acústicos - Medidas de duração normalizada do segmento	95
Tabela 6: Parâmetros fonético-acústicos - Medidas de frequência fundamental (f0) em semitons (re 1 Hz).....	97
Tabela 7: Parâmetros fonético-acústicos - Medidas de intensidade	100
Tabela 8: Parâmetros fonético-acústicos - Medidas de pausa silenciosa	100
Tabela 9: Medidas usadas na segunda fase de treinamento da avaliação inicial dos classificadores RF e LDA.....	111
Tabela 10: Treinamento inicial – Frequência bruta de unidades VVs escolhidas aleatoriamente	112
Tabela 11: Poder de predição alcançado pelo RF usando 111 medidas	112
Tabela 12: : Poder de predição alcançado pelo LDA usando 111 medidas	113
Tabela 13: Poder de predição alcançado pelo RF usando 28 medidas	114
Tabela 14: Poder de predição alcançado pelo LDA usando 28 medidas.....	115
Tabela 15: Poder de predição alcançado pelo LDA ao longo da eliminação progressiva de medidas – Fronteiras terminais.....	118
Tabela 16: Poder de predição alcançado pelo LDA ao longo da eliminação progressiva de medidas – Fronteiras não terminais	119
Tabela 17: Refinamento do treinamento LDA-TB – Frequência bruta de unidades VVs escolhidas aleatoriamente	121
Tabela 18: Poder de predição alcançado pelo LDA-TB para predizer as categorias presença e ausência de fronteira terminal usando 20 medidas	121
Tabela 19: Parâmetros relevantes para predizer as categorias presença e ausência de fronteira terminal usando o LDA e 20 medidas	122
Tabela 20: Refinamento do treinamento LDA-NTB – Frequência bruta de unidades VVs escolhidas aleatoriamente	123
Tabela 21: Poder de predição alcançado pelo LDA-NTB para predizer as categorias presença e ausência de fronteira não terminal usando 52 medidas	123
Tabela 22: Parâmetros relevantes para predizer a categoria presença de fronteira não terminal usando o LDA e 52 medidas.....	124
Tabela 23: Refinamento do treinamento LDA-NTB – Frequência bruta de unidades VVs escolhidas aleatoriamente	127
Tabela 24: Poder de predição alcançado pelo LDA-NTB para predizer as categorias presença e ausência de fronteira não terminal usando 9 medidas	128
Tabela 25: Parâmetros relevantes para predizer as categorias presença e ausência de fronteira não terminal, excluindo casos especiais de fronteiras não terminais, usando o LDA e 9 medidas	128
Tabela 26: Teste dos modelos LDA-TB e LDA-NTB – Frequência bruta de unidades VVs escolhidas aleatoriamente	129
Tabela 27: Poder de predição alcançado pelo teste final dos modelos LDA-TB e LDA-NTB para predizer as categorias presença e ausência de fronteiras	129

Tabela 28: Classificação de subtipos de fronteiras não terminais - Frequência bruta de unidades VVs avaliadas	131
Tabela 29: Poder de predição alcançado pelo Modelo 1 para predizer as categorias presença e ausência de fronteira não terminal.....	131
Tabela 30: Poder de predição alcançado pelo Modelo 2 para predizer as categorias presença e ausência de fronteira não terminal.....	132
Tabela 31: Poder de predição alcançado pelo Modelo 3 para predizer as categorias presença e ausência de fronteira não terminal.....	133
Tabela 32: Modelo 2 – Medidas relevantes para a classificação de fronteiras não terminais	134
Tabela 33: Modelo 3 – Medidas relevantes para a classificação de fronteiras não terminais	134
Tabela 34: Medidas em comum aos modelos.....	147

LISTA DE ABREVIATURAS E SIGLAS

F0	Frequência fundamental
LDA	<i>Linear Discriminant Analysis</i>
PB	Português do Brasil
RF	<i>Random Forest</i>
SN	Sintagma Nominal
SP	Sintagma Preposicional
SV	Sintagma Verbal

LISTA DE SÍMBOLOS

- (*) início de turno
- (ABC) identificação do informante
- //) fronteira prosódica terminal; fim de enunciado
- /) fronteira prosódica não terminal; fim de unidade entonacional interna ao enunciado
- ([/no]) *retracting*
- (&) início de palavra interrompida
- (&he) tomada de tempo

SUMÁRIO

INTRODUÇÃO	17
I. Justificativa.....	17
II. Objetivos	19
III. Organização do trabalho	20
1. Segmentação da fala em unidades	21
1.1. Propostas de segmentação da fala em unidades relevantes do ponto de vista comunicativo	25
1.1.1. Proposta dialógica	25
1.1.2. Proposta sintática.....	27
1.1.3. Proposta temporal	28
1.1.4. Proposta pragmática	31
1.2. Estudos sobre a segmentação da fala em unidades delimitadas por meio de fronteiras prosódicas	32
1.3. Parâmetros fonético-acústicos frequentemente citados nos estudos de fronteiras prosódicas	39
1.4. Segmentação automática da fala.....	46
2. Metodologia	48
2.1. Apresentação do <i>corpus</i> C-ORAL-BRASIL	48
2.2. Dados	49
2.3. Tratamento dos dados.....	50
2.3.1. Parâmetros fonético-acústicos extraídos pelo <i>BreakDescriptor</i>	61
I. Medidas de taxa de elocução (<i>speech rate</i>) e ritmo	64
II. Medidas de duração normalizada dos segmentos silábicos	66
III. Medidas de frequência fundamental (<i>f0</i>).....	77
IV. Medidas de intensidade	90

V. Medidas de pausa	93
2.3.2 Visão geral dos parâmetros fonético-acústicos calculados pelo <i>BreakDescriptor</i>	94
2.4. Análise estatística	100
2.4.1 Avaliação Inicial.....	104
2.4.2. Refinamento do classificador	104
2.4.3. Refinamento do modelo de fronteiras não terminais.....	106
3. Análises estatísticas e discussão dos resultados	108
3.1. Resultados da avaliação inicial dos classificadores <i>Random Forest</i> e <i>Linear Discriminant Analysis</i>	110
3.1.1. Poder de predição da fase I de avaliação inicial dos classificadores estatísticos	112
3.1.2. Poder de predição da fase II de avaliação inicial dos classificadores estatísticos.....	114
3.1.3. Discussão da avaliação inicial de classificadores estatísticos	116
3.2. Refinamento do classificador LDA	117
3.2.1. Resultados obtidos por meio da eliminação progressiva de medidas com baixo peso no modelo	118
3.2.2. Discussão dos resultados obtidos por meio da eliminação progressiva de medidas com baixo peso.....	120
3.2.3. Resultados obtidos por meio da eliminação de medidas não apontadas na literatura como relevantes para o estabelecimento de fronteiras prosódicas	120
3.2.4. Discussão dos resultados obtidos por meio da eliminação de medidas não apontadas na literatura como relevantes para o estabelecimento de fronteiras prosódicas.....	126
3.3. Refinamento do classificador de fronteiras não terminais.....	127
3.4. Teste no subconjunto de dados	129
3.5. Classificação de subtipos de fronteiras não terminais	130
3.6. Análise de <i>cluster</i> hierárquica de fronteiras não terminais.....	135
4. Discussão sobre os modelos finais de classificação de fronteiras prosódicas	139
I. Fronteiras terminais.....	139

II. Fronteiras não terminais	141
III. Parâmetros importantes para prever fronteiras prosódicas	145
IV. Diferença entre os parâmetros para estabelecer fronteiras terminais e não terminais.	150
a) Medidas de pausa	150
b) Medidas de duração normalizada e suavizada do segmento silábico.....	150
c) Medidas de frequência fundamental	151
d) Medidas de <i>speech rate</i> e ritmo.....	152
e) Medidas de intensidade	152
5. Conclusões e perspectivas futuras.....	154
REFERÊNCIAS	159
ANEXOS	166

INTRODUÇÃO

É comumente reconhecido que a fala é realizada e percebida em pequenas unidades prosódicas, geralmente formadas por algumas palavras, marcadas por fronteiras prosódicas (Schubiger, 1958; Chafe, 1980; Schuetze-Coburn, 1994; Ladd, 2008). As unidades podem ser analisadas funcionalmente segundo perspectivas teóricas diferentes: sintáticas (Cooper e Paccia Cooper, 1980; Selkirk, 2005), pragmáticas (Halliday, 1965; Cresti, 2000; Szczepek Reed, 2012) e cognitivas (Chafe, 1994; Croft, 1995; Bybee, 2010). As fronteiras, contudo, podem ser estudadas *per se*, independentemente da perspectiva teórica segundo a qual a unidade em que a fala é segmentada (Barth-Weingarten, 2016).

Este trabalho tem como foco investigar os parâmetros fonético-acústicos associados às fronteiras prosódicas em dados de fala espontânea monológica em português brasileiro (PB). A análise aqui desenvolvida utilizou uma metodologia de base acústica em conjunto com a percepção humana de fronteiras. Um trabalho como esse tem dois propósitos principais. O primeiro é contribuir ao melhor entendimento teórico sobre os parâmetros fonético-acústicos que orientam a segmentação da fala em unidades, o segundo é possibilitar o desenvolvimento de uma ferramenta computacional para detectar automaticamente ou, pelo menos, semi-automaticamente, fronteiras prosódicas ao longo do fluxo da fala. Futuramente, uma ferramenta computacional desse tipo poderá auxiliar na compilação de *corpora* de fala espontânea, porque poderá tornar 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.

I. Justificativa

É de suma importância segmentar o fluxo da fala em pequenos fragmentos determinados por fronteiras, tanto por motivos cognitivos, quanto por motivos linguísticos e respiratórios. Uma das causas que possivelmente explicaria a segmentação da fala em pequenos fragmentos delimitados por fronteiras é a capacidade da memória de trabalho do ser humano. Assim, a segmentação do fluxo da fala em pequenos agrupamentos marcados por fronteiras é

uma necessidade do processamento linguístico devido ao limite da memória de trabalho humana (Cowan, 1998).

Outra causa que pode explicar a segmentação é a produção de grupos respiratórios. Neste caso, a segmentação ocorreria nos pontos de inspiração humana. Deste modo, a fala seria segmentada em fragmentos produzidos em uma única respiração em virtude da capacidade do pulmão humano (Stetson, 1928).

A segmentação da fala em pequenos fragmentos determinados por meio de fronteiras também é justificada por razões linguísticas. Uma das razões linguísticas que possivelmente explicaria essa segmentação é a necessidade de captar adequadamente o domínio das relações linguísticas na fala.

Estruturas simples em Português do Brasil (doravante PB) mostram a importância de segmentar a fala em unidades:

Exemplo (1)

Não comprei a maçã.

O exemplo (1) pode ser segmentado de várias formas, como por exemplo:

- a) Não comprei a maçã /¹
- b) Não // comprei a maçã //

As duas possibilidades de segmentação carregam consigo possibilidades de interpretações distintas. O enunciado pode ser interpretado como um enunciado em que o falante não comprou a maçã (opção A) ou como um enunciado em que o falante comprou a maçã (opção B). A ambiguidade de interpretação é desfeita pela segmentação prosódica, pois é

¹ Para uma melhor compreensão dos exemplos: a barra simples (/) indica uma fronteira prosódica não terminal e a barra dupla (//) indica uma fronteira prosódica terminal. Para uma melhor compreensão dos dois tipos de fronteira prosódica, veja-se a seção 1.1.4 deste trabalho.

ela quem diz como as unidades segmentadas devem ser interpretadas para captar adequadamente o real domínio das relações linguísticas.

Uma fronteira prosódica é uma quebra perceptível que marca o agrupamento das palavras em um enunciado (Zhang, 2012; Blaauw, 1994 entre outros). Muitos trabalhos da área adotam as fronteiras de natureza prosódica como critério de orientação para segmentar a fala em unidades relevantes para a comunicação oral entre falantes (Cresti, 2000; Cooper, Paccia e Lapointe, 1978; Frazier, Clifton e Carlson, 2004; Swerts, Collier, Terken, 1994 e outros). Este trabalho é justificado pela necessidade de investigar a correlação entre as fronteiras prosódicas e os parâmetros fonético-acústicos que veiculam sua percepção.

II. Objetivos

O objetivo principal deste trabalho é compreender os parâmetros fonético-acústicos responsáveis pela segmentação da fala espontânea em unidades delimitadas por meio de fronteiras prosódicas. Um trabalho como este tem objetivos teóricos e práticos.

Os objetivos teóricos são:

- Descrever e investigar os parâmetros fonético-acústicos que orientam a produção e a percepção de fronteiras prosódicas;
- Descrever e investigar os parâmetros fonético-acústicos que distinguem diversos tipos de fronteiras prosódicas;
- Verificar as hierarquias dos parâmetros fonético-acústicos para a marcação dos tipos fronteiras prosódicas.

Os objetivos práticos são:

- Buscar e investigar o desempenho dos modelos de classificação estatística construídos por meio das combinações de múltiplos parâmetros fonético-acústicos capazes de

identificar, com a menor margem de erro, fronteiras de tipos diferentes em dados de fala espontânea em consonância com o julgamento de falantes (segmentadores treinados);

- Possibilitar a implementação de uma ferramenta computacional para segmentação, pelo menos parcial, do fluxo da fala em unidades marcadas por fronteiras prosódicas.

III. Organização do trabalho

O trabalho foi organizado em cinco capítulos. O primeiro capítulo se propõe a apresentar o problema de pesquisa central do trabalho, que é a importância da segmentação da fala em unidades marcadas por meio de fronteiras prosódicas. Neste capítulo, são apresentados alguns estudos sobre a segmentação da fala em unidades e os parâmetros fonético-acústicos tradicionalmente investigados nas pesquisas sobre a marcação de fronteiras prosódicas na fala. Também são citadas ferramentas computacionais já disponíveis para a segmentação automática do fluxo da fala.

No capítulo 2 é descrita a metodologia empregada, os dados analisados, o tratamento dos dados e as análises estatísticas desenvolvidas. No capítulo 3, são apresentados detalhadamente os resultados obtidos por meio das análises estatísticas. No capítulo 4, é apresentada a discussão final sobre os modelos finais de classificação de fronteiras prosódicas. No capítulo 5, são apresentadas as conclusões e as perspectivas futuras do trabalho aqui desenvolvido.

1. Segmentação da fala em unidades

Embora fala e escrita sejam duas modalidades das línguas, tais modalidades são muito diferentes entre si. O meio da fala são ondas sonoras produzidas pelo aparelho fonador, que são decodificadas pelo aparelho auditivo humano. O meio da escrita são símbolos gráficos, que são decodificados pela visão humana. O canal da fala é o ar, pelo qual as ondas sonoras se propagam. Já o canal da escrita é um suporte relativamente estável, como pedra, madeira, papel etc. Para uma discussão mais ampla sobre as diferenças entre fala e escrita, veja-se Raso (2013). Para os fins deste trabalho, é importante citar que as diferenças diamétricas entre fala e escrita resultam em modos distintos de organização da informação linguística. Por isso, fala e escrita não devem ser analisadas seguindo critérios semelhantes.

As análises da fala devem ser baseadas por meio da avaliação do sinal sonoro, o que inclui a informação prosódica ao longo do fluxo da fala. Naturalmente, conforme seja o interesse da pesquisa linguística, há várias formas de segmentar o fluxo da fala, como, por exemplo, a segmentação em fones, sílabas, palavras e unidades entonacionais mostrada na figura (1).

Exemplo (2) – bfammn24 – Áudio 1²

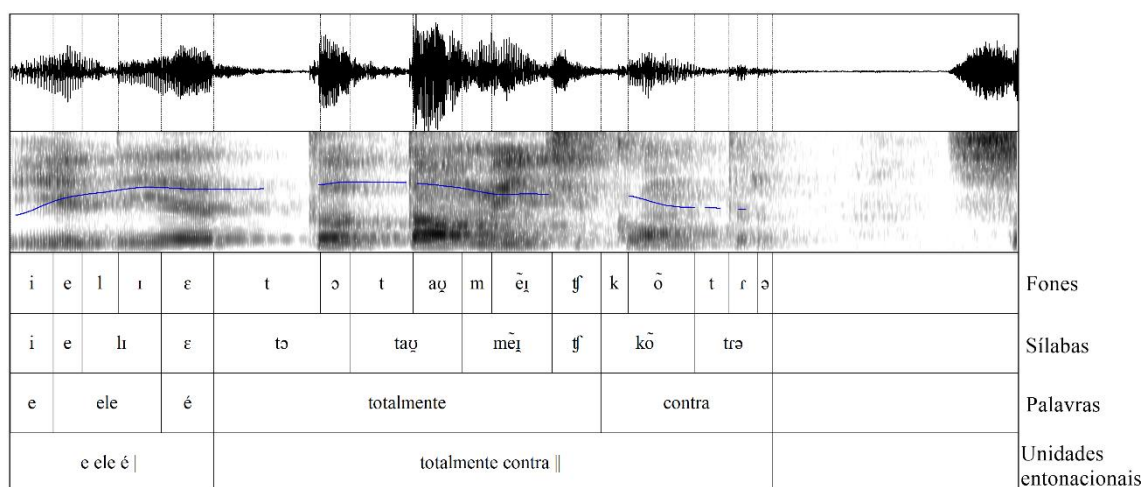


Figura 1: Exemplo de segmentação do sinal acústico em fones, sílabas, palavras e unidades entonacionais – Trecho: e ele é totalmente contra

De fato, a segmentação da fala em fones, sílabas ou palavras é muito importante para alguns estudos. Entretanto, os fones, as sílabas e as palavras mostram-se insuficientes para compreender aspectos da comunicação oral entre falantes. Os enunciados de fala somente são interpretados adequadamente diante da segmentação prosódica.

Exemplo (3)

Pedro vai para o Rio até amanhã.

Algumas das possibilidades de segmentação e interpretação do exemplo (3) são:

² Todos os áudios estão disponíveis no link:
https://drive.google.com/drive/folders/1uXWhU0JXsFcvcv9uwFPBeNO11_lcHefq

- a) Pedro! (*Chamamento*). Vai para o Rio! (*Ordem*) Até amanhã! (*Despedida*)
- b) Pedro vai para o Rio (*Asserção*). Até amanhã. (*Asserção*)
- c) Pedro vai para o Rio até amanhã. (*Asserção*)
- d) Pedro, vai para o Rio até amanhã! (*Ordem*)

O exemplo (3) e as suas diversas possibilidades de segmentação evidenciam que, de acordo com a segmentação do enunciado, as relações linguísticas mudam completamente. Caso o sintagma “*Pedro*” se configure como um chamamento e o sintagma “*vai para o Rio*” como uma ordem, *Pedro* não é sujeito do sintagma “*vai para o Rio*.” Assim como, se o sintagma “*até amanhã*” é uma despedida, este sintagma não se configura como adjunto adverbial do sintagma “*vai para o Rio*”. É fundamental que a segmentação da fala em unidades discretas seja adequada às especificidades da fala, pois, a interpretação dos enunciados produzidos oralmente só é acessível através do sinal acústico.

Algumas das funções da prosódia são segmentar o fluxo da fala em unidades, atrair a atenção do ouvinte para um ponto específico da mensagem, mostrar a estrutura sintática proferida pelo falante, retirar ambiguidades de sentenças com mais de um sentido, distinguir asserções de questões (entre outros atos de fala), expressar atitudes e emoções do locutor, auxiliar na identificação dos falantes, contribuir no reconhecimento de gêneros discursivos diferentes etc (Fónagy, 2003).

Barbosa (2012) distingue três funções principais da prosódia:

- I. Funções dialógicas e não dialógicas, que assinalam respectivamente a marcação de turno e a modalidade de um enunciado (se é um trecho interrogativo ou assertivo, por exemplo);
- II. Funções de proeminência, que assinalam a saliência de um constituinte prosódico em relação a outro;
- III. Funções demarcativas, que assinalam as fronteiras de constituintes prosódicos, como por exemplo, as sílabas, as palavras fonológicas e os agrupamentos da fala em unidades entonacionais.

Como exemplo da função dialógica, tem-se dois turnos de fala.

Exemplo (4) – bfamd108 – Áudio 2

*BRU: [107] com quem //³

*AND: [108] é com Tita //

Como exemplo da função de proeminência, tem-se a oposição entre dois enunciados com foco em posições diferentes.

Exemplo (5)

- O vestido da NOIVA é verde.
- O vestido da noiva é VERDE.

Como exemplo da função demarcativa, tem-se a oposição entre dois enunciados compostos por quatro palavras iguais, muito distintos entre si, de acordo com a posição das fronteiras prosódicas que assinalam os agrupamentos em unidades relevantes para compreensão da fala.

Exemplo (6)

- Não está chovendo //
- Não // está chovendo //

³ Para uma melhor compreensão dos exemplos retirados do *corpus* C-ORAL-BRASIL, segundo os critérios de transcrição adotados (Mello e Raso, 2009): a) a sigla formada por uma sequência de seis letras e dois números como "bfamd108" indica a identificação dos textos no *corpus*; b) a sigla formada pelo símbolo asterisco (*) seguido por três letras maiúsculas como "*BRU" indica o falante do turno dialógico; c) o número entre colchetes como "[245]" identifica o enunciado no texto de origem; d) a barra simples (/) indica uma fronteira prosódica não terminal e a barra dupla (//) indica uma fronteira prosódica terminal; e) o colchete seguido por barra simples e um número como [/3] indica *retracting* (falha para executar determinado número de palavras).

A função demarcativa revela a sua importância para assinalar as fronteiras de qualquer constituinte prosódico. A prática de segmentar o fluxo da fala em unidades marcadas por fronteiras prosódicas para estudar a organização da fala tem sido consistentemente explorada no âmbito de compilação de grandes *corpora* orais de português brasileiro, inglês, italiano, francês, espanhol, hebraico, português europeu etc. Podemos citar o Santa Barbara Corpus of Spoken American English (Du Bois *et al.*, 2005), o C-ORAL-ROM (Cresti e Moneglia, 2005), o C-ORAL-BRASIL (Raso e Mello, 2012), o Corpus of Spoken Israeli Hebrew (Izre'el *et al.*, 2009) e o corpus of Spoken AfroAsiatic Languages (Mettouchi e Chanard, 2010).

1.1. Propostas de segmentação da fala em unidades relevantes do ponto de vista comunicativo

Cresti e Gramigni (2004) apresentam as principais propostas de segmentação da fala em unidades:

- A proposta dialógica;
- A proposta sintática;
- A proposta temporal;
- A proposta pragmática.

1.1.1. Proposta dialógica

A perspectiva dialógica adota o turno dialógico de uma interação oral como unidade de referência para a segmentação da fala. Assim, no exemplo abaixo retirado do *corpus* C-ORAL-BRASIL (Raso e Mello, 2012), as unidades de referência seriam duas.

Exemplo (7) – bfamd116 – Áudio 3

*CRI: [21] a Maria vai vim hoje //

*ASI: [22] a Maria não //

E no exemplo (8), também retirado do *corpus* C-ORAL-BRASIL, teríamos uma única unidade.

Exemplo (8) – bfammn08 – Áudio 4

*HMB: [4] bom // eu ã sugiro jantar / sem a pessoa me falar que que ela gosta de comer // porque / cada um / tem um [/1] uma afinidade com um tipo de coisa / gosta / de um tipo de comida / então eu nunca sugiro antes de ter uma [/1] um feedback da pessoa // ela tem que falar / assim / de que que ela gosta // se ela gosta de carne / se ela gosta de peixe / se ela gosta de frango / se ea gosta de verdura / e quem ã fala nada / pode ser qualquer coisa / normalmente / ã dá certo // porque / aí / eu vou sugerir de acordo com [/1] com [/1] com o que eu penso //

Os exemplos (7) e (8) mostram que os turnos, apesar de serem unidades naturais da fala, são unidades que podem ser muito distintas entre si, podendo ser muito longos ou extremamente curtos, compostos por estruturas complexas ou até mesmo por uma interjeição. A alta variabilidade dos turnos dialógicos indica que eles devem ser vistos como unidades de outro tipo. Por isso, eles não seriam muito úteis para segmentar a fala em unidades relevantes do ponto de vista comunicativo.

1.1.2. Proposta sintática

A perspectiva sintática adota a sentença falada como unidade de referência para a segmentação da fala. A predicação pode ser definida como uma estrutura sintática que possui uma predicação verbal relativa a um SN sujeito, hierarquicamente subordinado ao SV (Harris, 1962). Entretanto, ao analisar as predicações verbais, Cresti e Gramigni (2004) mostram que estas estruturas correspondem a menos de 5% dos enunciados do C-ORAL-ROM italiano, se formulada com SN e SV pleno.

A sentença falada também pode ser definida como a máxima projeção do SV (Chomsky, 1970). Todavia, a máxima projeção do SV não inclui os enunciados de fala sem formas verbais, além daqueles enunciados em que o verbo não é núcleo ou não possui função verbal. No *corpus* C-ORAL-ROM (Cresti e Moneglia, 2005), os resultados mostram que, na fala espontânea, 38,1% do total de enunciados sequer possui verbo em sua estrutura. Boa parte dos enunciados são formados somente por SNs, SPs, adjetivos ou interjeições. No *corpus* C-ORAL-BRASIL, 22,1% dos enunciados de textos monológicos e 29,5% dos enunciados de textos dialógicos também não possuem formas verbais.

Exemplo (9) – bfamdl05 – Áudio 5

*CES: [223] então / até final de novembro //

*ENC: [224] isso //

Exemplo (10) – bfamdl04 – Áudio 6

*SIL: [6] capaz //

*KAT: [7] ahn //

Em diversas línguas, os enunciados sem verbos constituem cerca de um terço do total (Cresti, 2005; Biber *et al.*, 1999). Raso e Mittmann (2012) observam que, no *corpus* de PB C-ORAL-BRASIL, o número de enunciados sem verbo e enunciados em que o verbo não é núcleo ou não possui função verbal supera 50%.

O conceito de sentença falada nas duas acepções parece ser inadequado para segmentar a fala em unidades relevantes do ponto de vista comunicativo. Os princípios de organização da fala e da escrita são muito diferentes. Por isso, os critérios sintáticos, comuns na segmentação da língua escrita, incluindo a estrutura por sentença e as regras de pontuação, não são capazes de segmentar a fala em unidades relevantes para a comunicação oral entre falantes.

1.1.3. Proposta temporal

A perspectiva temporal adota como unidade de segmentação da fala um trecho de fala presente entre duas pausas do mesmo falante. Frequentemente, distinguem-se “pausa silenciosa” e “pausa preenchida” (Zellner, 1994; Soriano, 2006; Merlo e Barbosa, 2012). A pausa silenciosa é composta por uma interrupção do sinal acústico, ou seja, um silêncio no fluxo da fala. As pausas preenchidas são disfluências compostas por algum material vocalizado (Fletcher, 2010). A pausa preenchida é considerada uma interrupção cognitiva que não interrompe o sinal acústico, porque este tipo de pausa é realizado por meio de alguma vocalização.

Alguns trabalhos adotam a pausa silenciosa como critério de segmentação da fala em unidades (Edwards e Lampert, 1993; Cruttenden, 1997). Alguns *corpora* orais disponíveis consideram uma duração de pausa silenciosa específica para segmentar as unidades relevantes do ponto de vista comunicativo. Assim, nestes *corpora*, uma pausa silenciosa de determinada duração assinala o começo de uma unidade e a próxima pausa silenciosa desta mesma duração assinala o término da unidade.

O *Dutch Corpus* (Buhmann *et al.*, 2002) considerou uma pausa silenciosa igual ou superior a 200 ms como critério de segmentação da fala em unidades. *Corpora* japoneses

(Maruyama, 2012, 2015) consideraram uma pausa silenciosa igual ou superior a 100 ms. Naturalmente, a decisão sobre quanto considerar uma pausa suficientemente longa para constituir critério de segmentação é arbitrária. Knowles (1991) adota pelo menos 250 ms de silêncio, Roelof de Pijper e Sandeman (1994) adotam pelo menos 100 ms, enquanto Avanzi, Lacheret-Dujour e Victorri (2008) definem em 300 ms a pausa mínima para o Francês e Batliner *et al.* (1998) adotam pausas silenciosas superiores a 500 ms.

Uma avaliação a respeito das pausas silenciosas revela que unidades relevantes para a comunicação falada nem sempre são marcadas por pausas silenciosas. O trecho abaixo, também retirado do *corpus* C-ORAL-BRASIL (em preparação), ilustra por meio da área sombreada a ausência de pausas silenciosas depois de fronteira.

Exemplo (11) – bmidmasc01 – Áudio 7

*ROB: mas / para um aprendizado novo / esse cérebro / mais jovem / tá mais capaz //

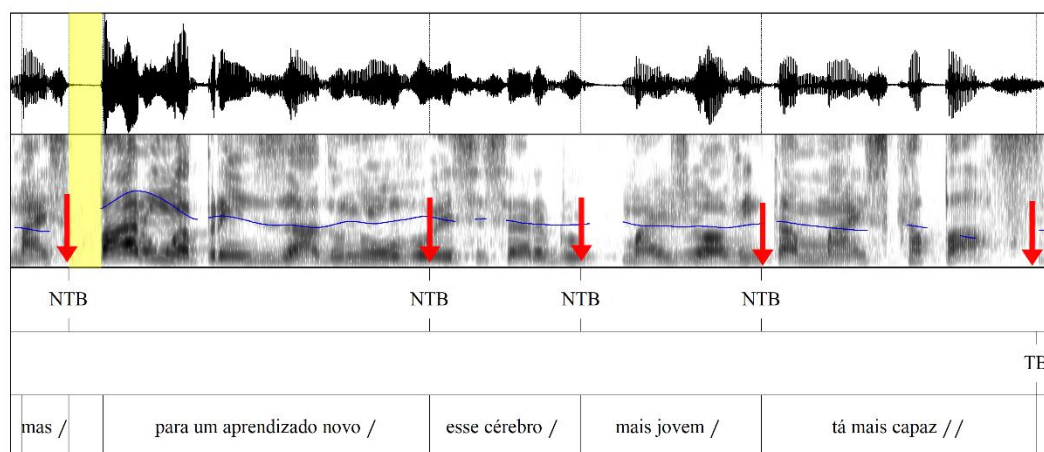


Figura 2: Exemplo de fronteiras prosódicas com e sem pausa silenciosa

A figura (2) mostra uma unidade delimitada por meio de uma fronteira terminal e segmentado internamente por meio de quatro fronteiras não terminais. Observa-se que apenas

a fronteira não terminal localizada após a palavra “mas” é seguida por pausa silenciosa de 150 ms. As fronteiras prosódicas localizadas após as palavras “novo”, “cérebro”, “jovem” e “capaz” não são seguidas por pausa silenciosa, pois, não há nenhuma interrupção do sinal sonoro logo após o final das palavras.

A pausa silenciosa é um parâmetro fonético-acústico tradicionalmente apontado como relevante para a marcação de fronteiras prosódicas no fluxo da fala (Swerts *et al.*, 1992; Swerts, 1997; Liu *et al.*, 2006; Schuetze-Coburn, 1991; Tyler, 2013). Entretanto, observou-se que nem sempre fronteiras prosódicas são marcadas por meio de pausas silenciosas. De acordo com Shriberg (2005, p. 1), “pauses are neither necessary nor sufficient indicators of boundaries.” Raso *et al.* (2015) avaliaram a coincidência entre a presença de fronteira prosódica e a presença de pausa silenciosa após a fronteira. Os resultados obtidos neste trabalho mostram que: a) 62% das fronteiras não terminais não apresentam pausas silenciosas; b) 33% das fronteiras terminais também não apresentam pausas silenciosas.

Evidentemente, é impossível negar o papel crucial que a pausa silenciosa assume na tarefa de detecção automática de fronteiras prosódicas, dada a maior facilidade para desenvolver uma ferramenta computacional capaz de reconhecer um fenômeno dessa natureza. Assim, em geral, as pausas silenciosas podem ser vistas como marcadores de fronteiras prosódicas, porque naturalmente todas as pausas assinalam fronteiras. Todavia, elas não devem ser vistas como o único parâmetro marcador de fronteira, visto que uma quantidade considerável de fronteiras prosódicas perceptivelmente relevantes não é marcada por meio de pausas silenciosas.

Há um outro problema relacionado à segmentação do fluxo da fala baseando-se exclusivamente em pausas silenciosas. As pausas não são indicadores confiáveis para discriminar o tipo de fronteira prosódica (Tseng e Chang, 2008; Raso *et al.*, 2015). Em outras palavras, para determinar se uma fronteira é percebida como conclusiva ou não conclusiva, é necessário recorrer a outros parâmetros fonético-acústicos. Deste modo, para fornecer a uma ferramenta computacional instruções de como detectar fronteiras prosódicas automaticamente, também é necessário investigar os demais parâmetros fonético-acústicos que orientam a produção e a percepção dos dois tipos de fronteiras prosódicas.

1.1.4. Proposta pragmática

A proposta pragmática argumenta que o enunciado é a unidade de referência da fala, porque ele possui características prosódicas e pragmáticas tais que conferem a ele autonomia comunicativa, ou seja, o estatuto de ato de fala (Austin, 1962). O enunciado é delimitado por meio de fronteiras prosódicas percebidas pelos falantes como conclusivas (Crystal, 1975). Há dois tipos de fronteiras prosódicas perceptivamente relevantes. Além das fronteiras que veiculam a percepção de conclusão, existem fronteiras que veiculam a percepção de continuidade. Chamaremos as primeiras de fronteiras terminais (TB) e as segundas, de fronteiras não terminais (NTB).

As fronteiras terminais segmentam o fluxo da fala em unidades autônomas pragmaticamente e prosodicamente. As fronteiras não terminais têm como função comunicar o caráter não conclusivo de uma unidade interna ao enunciado e sinalizar a continuidade da unidade em questão. Deste modo, as fronteiras terminais delimitam o enunciado e as fronteiras não terminais delimitam as unidades entonacionais internas do enunciado.

Os enunciados constituídos por uma única unidade são classificados como simples, e aqueles que são segmentados em mais unidades, como complexos (Cresti, 2000). Os enunciados denominados complexos são constituídos por pelo menos duas unidades, das quais somente a última é marcada por fronteira terminal.

Adotou-se como sistema padrão de anotação de fronteiras o uso de uma barra dupla (//) para marcar fronteira terminal e uma barra simples (/) para marcar fronteira não terminal. Com este padrão de anotação, a fins de comparação, o exemplo (2) poderia ser segmentado de várias formas:

- a) Pedro // (*Chamamento*) Vai para o Rio // (*Ordem*) Até amanhã // (*Despedida*)
- b) Pedro vai para o Rio // (*Asserção*) Até amanhã // (*Asserção*)
- c) Pedro vai para o Rio até amanhã // (*Asserção*)
- d) Pedro / vai para o Rio até amanhã // (*Ordem*)

Em uma segmentação do tipo (a), teríamos três unidades finalizadas por uma fronteira terminal, que indicam três enunciados e, portanto, três atos de fala – um chamamento, uma ordem e uma despedida. Em uma segmentação do tipo (b), teríamos duas unidades finalizadas por duas fronteiras terminais, que indicam cada uma delas uma asserção. Em uma segmentação do tipo (c), teríamos uma unidade marcada por uma fronteira terminal, que indica uma asserção. Em uma segmentação do tipo (d), teríamos duas unidades, mas um único enunciado. Das duas unidades, a primeira é delimitada por uma fronteira não terminal e, portanto, não é autônoma; a segunda é delimitada por uma fronteira terminal.

1.2. Estudos sobre a segmentação da fala em unidades delimitadas por meio de fronteiras prosódicas

Diversos trabalhos já apresentaram contribuições para o estudo da segmentação da fala em pequenas unidades, geralmente formada por algumas palavras, delimitadas por meio de fronteiras prosódicas. Cruttenden (1997) aponta uma série de critérios internos e externos para o estabelecimento de fronteiras prosódicas no fluxo da fala. Os critérios externos incluem o alongamento final da unidade, a pausa silenciosa e a mudança do nível de *pitch* ou da direção do *pitch*. Como critério interno, inclui-se a presença de uma sílaba proeminente, chamada de núcleo, com um movimento de *pitch* significativo. Contudo, a lista de parâmetros fonético-acústicos que estabelecem fronteiras de natureza prosódica foi definida principalmente com base em textos monológicos ou frases isoladas eliciadas em laboratório. De fato, a aplicação desses critérios em textos de fala espontânea é por vezes difícil.

Tendo em vista a dificuldade de aplicação desses critérios na fala espontânea, Cruttenden (1997) recomenda a adoção de critérios gramaticais, argumentando que as fronteiras prosódicas muitas vezes coocorrem com as fronteiras dos constituintes sintáticos. Porém, muitos dados de fala espontânea mostram que, frequentemente, não há uma relação biunívoca entre fronteira prosódica e fronteira sintática dos constituintes (Steedman, 1990; Hirst, 1993; Shattuck-Hufnagel e Turk, 1996). A adoção de critérios gramaticais para identificar fronteiras prosódicas mostra-se como uma alternativa pouco interessante, pois as fronteiras prosódicas

deixam de ser estudadas sob a perspectiva acústica e passam a ser estudadas sob uma perspectiva teórica pré-definida.

Du Bois *et al.* (1992) argumentam que uma unidade prototípica delimitada por meio de fronteiras de natureza prosódica tem como características o *reset* do nível base do *pitch* no começo da unidade, a pausa no começo da unidade e o alongamento de uma ou mais sílabas no final da unidade. Entretanto, foi observado que algumas fronteiras prosódicas apresentam apenas parte das características prototípicas citadas.

Os autores, portanto, defendem que há dois graus distintos de fronteiras prosódicas. O primeiro deles compreende as fronteiras “plenas”, que apresentam todas as características prototípicas. O segundo compreende as fronteiras “parciais”, que apresentam apenas algumas das características prototípicas. Este último tipo muitas vezes geraria dúvidas, pois um segmentador treinado pode perceber a fronteira, enquanto outro também treinado pode não a perceber. Du Bois (2008) complementa a lista de pistas acústicas incluindo a *creaky voice*.

Os trabalhos citados até agora têm em comum o foco em um sistema binário dividido entre a percepção e a não percepção de fronteiras ao longo do fluxo da fala. É notório que, em vários casos, não é claro onde a fronteira se localiza e nem se há a marcação de fronteira. Isto se deve ao fato de que a percepção de fronteiras ao longo do sinal acústico da fala não constitui uma decisão categórica (Bolinger, 1972; Birkner, 2006; Barth-Weingarten, 2016). Se algumas fronteiras prosódicas são muito salientes perceptualmente e são percebidas por (quase) todas as pessoas, outras têm um acordo bem menor quanto à sua percepção e ao seu tipo.

Conforme veremos abaixo, a percepção humana não é uma instância categorizada de forma binária. O exemplo (12) segmentado por 14 segmentadores treinados obteve um acordo parcial quanto à percepção de fronteiras prosódicas.

Exemplo (12) – bfammn11 – Áudio 8

*LAU: e com os cadernos todos / encadernados de azul / tudo / todos esticadinhos /
tudo muito bonito / assim //

Em (12), após a palavra “todos” foi percebida uma fronteira não terminal por 13 segmentadores, após a palavra “tudo” foi percebida uma fronteira não terminal por 14 segmentadores, após a palavra “esticadinho” foi percebida uma fronteira não terminal por 14 segmentadores, após a palavra “bonito” foi percebida uma fronteira não terminal por 11 segmentadores, após a palavra assim foi percebida uma fronteira terminal por 14 segmentadores. Após a palavra “azul” observou-se que enquanto sete segmentadores perceberam fronteira não terminal, cinco perceberam fronteira terminal. Dois segmentadores não perceberam nenhum tipo de fronteira.

Assim, dentre as seis fronteiras, três obtiveram total acordo. Duas obtiveram um acordo alto e uma obteve um desacordo substancial. O desacordo obtido sugere que não é claro para os segmentadores se a posição é de fronteira terminal ou de fronteira não terminal.

Para casos de desacordo entre segmentadores, muitos segmentadores acabam tomando decisões com base em razões teóricas de ordem sintática, pragmática ou cognitiva para justificar uma determinada segmentação da fala em unidades relevantes para a comunicação falada. Entretanto, tal atitude causa um efeito de circularidade (Brown *et al.*, 1980; Peters *et al.*, 2005). A solução para casos de desacordo, quanto à presença de fronteira e ao tipo de fronteira, não deve ser determinada por razões teóricas, mas sim por razões perceptivas. É necessário que a análise das fronteiras prosódicas seja autônoma, isto é, independente da análise funcional das unidades que elas geram.

Com base nos trabalhos da área, parece evidente que as fronteiras prosódicas não são sempre da mesma natureza. Uma revisão da literatura indica que os tipos de fronteiras são associados à percepção de conclusão ou de continuação do enunciado (Pike, 1945; Pierrehumbert, 1980; Schegloff, 1998; Szczepek Reed, 2004). Por isso, seria esperado dois conjuntos de parâmetros fonético-acústicos, um necessário para o estabelecimento da fronteira conclusiva, outro necessário para o estabelecimento da fronteira não conclusiva. Todavia, a distinção não parece ser tão simples.

Por exemplo, se considerarmos a fronteira que sinaliza não conclusão do enunciado, normalmente, o movimento ascendente da frequência fundamental (f_0) é apontado como um parâmetro fonético-acústico relevante para a marcação deste tipo de fronteira; entretanto, vários casos de fronteiras não conclusivas não são realizados com um perfil de f_0 ascendente. Deste modo, observamos que existem parâmetros fonético-acústicos diferentes para fronteiras classificadas tradicionalmente da mesma natureza.

Exemplos de fronteiras não terminais com configurações de parâmetros fonético-acústicos diferentes do perfil de f_0 ascendente nos servem como justificativa para pesquisarmos a hipótese de que a gama de tipos de fronteiras prosódicas é superior a dois. Pelo menos quanto às fronteiras não terminais, parece provável que seja oportuno buscar mais de um conjunto de parâmetros fonético-acústicos para explicar as diversas fronteiras. A produção e a percepção de fronteiras ao longo do fluxo da fala são atividades muito refinadas, que possivelmente vão além de uma categorização binária dividida entre fronteira terminal e não terminal.

Tendo em vista o que foi exposto acima, um estudo que busca estudar as fronteiras por si só procura evitar os efeitos de circularidade promovidos pelas teorias com foco na função linguística da unidade, associando a percepção humana de fronteiras aos parâmetros fonético-acústicos. Contudo, conforme veremos, o estabelecimento de uma correlação deste tipo é extremamente complexo.

A complexidade emerge por várias causas. Inicialmente, emerge da quantidade de parâmetros fonético-acústicos combinados no fluxo da fala, que podem contribuir em maior ou menor grau com a percepção de fronteira. No trabalho de Amir, Silver-Varod e Izre'el (2004), foram utilizados quatro parâmetros fonético-acústicos para analisar o estabelecimento de fronteiras: 1) A fala mais rápida no início da unidade marcada por fronteiras (*initial rush*); 2) A fala mais lenta no final da unidade marcada por fronteiras (alongamento da sílaba final); 3) O *reset* da frequência fundamental entre unidades; e 4) a pausa silenciosa após a fronteira.

A partir do hebraico, os autores apresentam uma proposta de parâmetros mais relevantes para o estabelecimento e percepção de fronteiras. Os resultados revelam que o alongamento de sílaba final foi encontrado em 81% dos dados, sendo a pista mais relevante para a percepção, seguida pelo *reset* de f_0 , que ocorreu em 70% (75% desse total era de tons terminais descendentes e 24% de ascendentes), e pelas pausas silenciosas em 55% das fronteiras prosódicas percebidas. A taxa de elocução mais elevada no início da unidade, segundo os autores, teve uma participação mais modesta, tendo sido encontrada em 34% dos casos. Os resultados também mostram que 24% das fronteiras avaliadas apresentaram todos os quatro parâmetros fonético-acústicos avaliados.

Cruttenden (1997), em seu trabalho sobre entonação em inglês, defende que há dois tipos de contornos entonacionais diante de fronteira prosódica. De acordo com este autor, o contorno entonacional ascendente indica que há mais informação para vir após a fronteira prosódica; já o contorno entonacional descendente indica que a informação está completa e, portanto, não há mais nenhuma informação a ser acrescentada.

Mo (2008), em seu estudo sobre inglês, argumenta que a duração é um parâmetro fonético-acústico importante para a percepção de fronteiras prosódicas ao longo da fala espontânea, comprovando que os falantes são sensíveis à percepção desse fenômeno quando solicitados a indicar pontos de ocorrência de fronteiras prosódicas. Cole, Mo e Baek (2008) também investigam os parâmetros fonético-acústicos associados às fronteiras prosódicas em dados de língua inglesa. Os resultados obtidos por meio de testes estatísticos indicam que a duração, a intensidade e a pausa silenciosa são parâmetros significativos para a percepção de fronteiras.

Shen (1992), em um estudo sobre mandarim, mostra que tanto a pausa quanto o alongamento fronteiro são parâmetros fonético-acústicos relevantes para a produção de fronteiras prosódicas. Para a percepção de fronteiras prosódicas, entretanto, de acordo com os testes perceptivos realizados por Shen (1992), a pausa é o parâmetro mais relevante. Outros estudos sobre fronteiras prosódicas em língua inglesa também argumentam que a duração e a presença de pausa silenciosa são parâmetros fonético-acústicos relevantes para o estabelecimento de fronteiras prosódicas (Wightman *et al.*, 1992; Campbell, 1993; Ni *et al.*, 2012).

Em um estudo sobre inglês e chinês, Zhang (2012) mostra por meio de testes perceptivos que falantes nativos de inglês e chinês usam pausa silenciosa, alongamento (pré-fronteiro ou fronteiro) e o *reset* de f0 como parâmetros importantes para perceber fronteiras prosódicas. Contudo, a importância destes três parâmetros para falantes de inglês e chinês é diferente. Para os falantes nativos de inglês, a pausa possui maior importância que o alongamento (pré-fronteiro ou fronteiro) e o *reset* de f0. Para os falantes nativos de chinês, o *reset* de f0 possui maior relevância que a pausa e o alongamento em posições pré-fronteiriças ou fronteiriças.

Para o francês, Mertens e Simon (2013) mostram que o parâmetro fonético-acústico mais importante para o estabelecimento de fronteiras prosódicas conclusivas são as pausas silenciosas (85,8%) e, em seguida, alongamentos nas posições adjacentes às fronteiras (11%). Para o estabelecimento de fronteiras não conclusivas, os autores mostram que os parâmetros mais relevantes são o alongamento nas posições adjacentes às fronteiras (24,1%) e o *pitch* da sílaba fronteira (24,1%).

Para o holandês, Swerts *et al.* (1994) mostram, a partir de testes perceptivos conduzidos por meio da fala eliciada, que o tipo de tom e a duração no final da unidade entonacional são parâmetros relevantes para prever o estabelecimento de fronteiras

prosódicas. A fala eliciada no trabalho de Swerts *et al.* (1994) é obtida por meio de descrição de rotas. O falante é apresentado a um mapa em que são marcados um ponto de origem e um ponto de destino. A tarefa do falante é fornecer orientações de como chegar ao ponto de destino partindo do ponto de origem. Os autores estabelecem três tipos de fronteiras. O primeiro tipo de fronteira indica que ainda há muita informação por vir, o segundo tipo indica que há um pouco mais de informação por vir e o terceiro tipo indica que a informação está completa. Deste modo, neste trabalho é explorado um aspecto não visto pelos demais autores, pois, é abandonada a bipartição entre fronteiras prosódicas que indicam continuidade ou finalidade.

Blaauw (1994) estuda detalhadamente os parâmetros associados às fronteiras prosódicas em alemão. A autora analisa a produção de padrões diferentes de realização de fronteiras prosódicas em duas modalidades de fala: fala espontânea e fala lida. Para a análise dos dados, são estabelecidos quatro padrões fonético-acústicos de realização de fronteiras prosódicas. O primeiro é caracterizado pelo alongamento pré-fronteiriço, o segundo é caracterizado simultaneamente pelo alongamento pré-fronteiriço e pela pausa silenciosa, o terceiro é caracterizado simultaneamente pelo alongamento pré-fronteiriço e por um movimento de *pitch*, o quarto é caracterizado pela presença dos três parâmetros estudados, ou seja, alongamento pré-fronteiriço, movimento de *pitch* e pausa. Blaauw (1994) argumenta que, em ambas as modalidades de fala, a maioria das fronteiras prosódicas (84%) foi realizada com o padrão 4, incluindo então, todos os parâmetros fonético-acústicos analisados. Especificamente, em relação a fala espontânea, de acordo com a autora, o *pitch* ascendente é predominante em fronteiras prosódicas.

Evidentemente, os parâmetros fonético-acústicos envolvidos no estabelecimento de fronteira prosódica podem variar conforme a língua. Apesar de observarmos algumas variações no que diz respeito aos parâmetros fonético-acústicos citados na literatura da área, é bastante claro que a maioria deles é recorrente em diversos estudos sobre diversas línguas. Ainda a respeito dos parâmetros citados na literatura, Auer (2010) mostra, por outro lado, que não há acordo sobre o peso de cada um deles para o estabelecimento de fronteira no plano fonético-acústico. Os estudos da área não mostram com clareza quais são os parâmetros fonético-acústicos que distinguem os dois tipos de fronteiras frequentemente citados (terminais e não terminais) e o papel desempenhado por cada parâmetro para estabelecer as fronteiras. Seriam estes parâmetros hierarquizados, combinados ou alternativos?

A complexidade de um estudo sobre fronteiras prosódicas independentemente da análise funcional das unidades ocorre também, como já foi dito, porque algumas fronteiras são

muito salientes e percebidas facilmente por quase todas as pessoas, enquanto outras não. Para este caso, uma possível hipótese é: as fronteiras que são percebidas por um número maior de pessoas são aquelas que apresentam os parâmetros fonético-acústicos mais prototípicos e usuais na língua em questão.

O estudo de fronteiras prosódicas por si só implica em uma série de dificuldades metodológicas, como por exemplo a dificuldade de encontrar nos dados trechos de fala absolutamente comparáveis para investigar, as limitações que os dados eliciados oferecem e a busca por textos de fala espontânea com alta qualidade acústica. Por fim, um estudo sobre fronteiras prosódicas *per se* é extremamente complexo, porque é difícil investigar um aspecto tão refinado como a percepção humana.

Na tentativa de realizar a análise das fronteiras prosódicas *per se*, Auer (2010) introduz o termo “cesura” para fazer referência à análise independente das fronteiras. Contudo, parece que isso de fato não ocorre, porque a análise proposta por ele contempla as dimensões sintática, prosódica (fonético-acústica) e pragmática das fronteiras.

Uma perspectiva de estudo que visa analisar as fronteiras *per se* de uma forma mais refinada é apresentada por Barth-Weingarten (2016). Em Barth-Weingarten, também se adota o termo “cesuras”. Na visão dessa autora, a perspectiva das “cesuras” promove o estudo das fronteiras *per se*, porque são abandonadas as noções de unidades e fronteiras prosódicas de um determinado tipo, privilegiando a noção de descontinuidades ao longo do fluxo do sinal sonoro. Barth-Weingarten (2016) argumenta que a percepção humana da fala ocorre por meio de um sistema gradiente, abandonando assim o sistema binário dividido entre fronteiras conclusivas e não conclusivas. As cesuras são definidas como o efeito das mudanças dos parâmetros prosódico-fonéticos que resultam na percepção de descontinuidades no fluxo da fala (Barth-Weingarten, p. 60, 2016). Nesta visão, as mudanças dos parâmetros prosódico-fonéticos ao longo do fluxo da fala são o foco de análise, pois são estas mudanças que possibilitam uma percepção mais refinada e gradiente da linguagem oral.

O desacordo perante a percepção de fronteiras pode ser visto como indicativo de que as fronteiras não são objetos concretos, mas sim instâncias parcialmente estáveis, que são produzidas e percebidas pelos falantes através da variação de muitos parâmetros fonético-acústicos, conforme Barth-Weingarten argumenta. O trabalho aqui desenvolvido tem como objetivo investigar a percepção de fronteiras prosódicas ao longo do fluxo da fala à luz de uma abordagem acústica. O estudo parte da percepção de fronteiras prosódicas em um sistema binário categorizado entre terminal e não terminal, mas busca uma categorização mais refinada

de subtipos de fronteiras prosódicas além de terminal e não terminal, realizados por diferentes configurações de parâmetros fonético-acústicos.

1.3. Parâmetros fonético-acústicos frequentemente citados nos estudos de fronteiras prosódicas

Uma revisão de literatura da área indica que os estudos linguísticos já identificaram uma série de parâmetros fonético-acústicos básicos para descrever fronteiras prosódicas ao longo do fluxo da fala. Alguns destes parâmetros são recorrentes em diversos trabalhos, como, o *reset* da frequência fundamental (f_0), o alongamento pré-fronteiriço e fronteiriço, a mudança na taxa de elocução e a pausa silenciosa (Cruttenden, 1997; Crystal, 1969; Du Bois *et al.*, 1992; Du Bois, 2008; Kelly e Local, 1989; Amir, Silver-Varod e Izre'el, 2004; Mo, 2008; Blaauw; 1994). Entretanto, nestas análises, algumas vezes foram utilizados *corpora* de fala gravados em laboratório, a maior parte com fala lida. Os resultados obtidos por meio de fala gravada em laboratório são questionáveis, porque essas situações de comunicação não refletem necessariamente situações reais de fala. Os próximos parágrafos apresentam os parâmetros fonético-acústicos apontados em diversos trabalhos como parâmetros fonético-acústicos relevantes para o estabelecimento de fronteiras prosódicas. Serão usados dados de Português Brasileiro (PB) extraídos do *corpus* C-ORAL-BRASIL para exemplificar estes parâmetros.

A pausa silenciosa é um parâmetro fonético-acústico tradicionalmente apontado como um parâmetro que estabelece fronteiras prosódicas no fluxo da fala. O exemplo (13), ilustrado pela figura (3), mostra um exemplo de enunciado complexo em que há a ocorrência de pausa silenciosa após uma das fronteiras não terminais e após a fronteira terminal.

Exemplo (13) – bfammn24 – Áudio 9

*MIC: mas eu / não tenho vocação nenhuma / pra / administração //

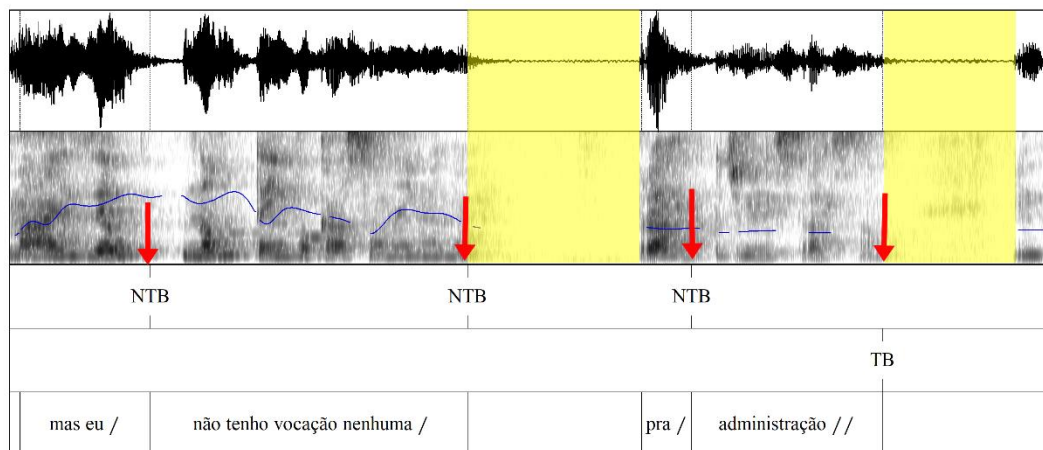


Figura 3: Exemplos de fronteiras (não terminal e terminal) seguidas por pausa

As áreas sombreadas do espectrograma mostram exemplos de duas pausas localizadas após rupturas perceptivelmente relevantes para a avaliação da organização da fala espontânea. A fronteira não terminal localizada após o item lexical “*nenhuma*” é seguida por uma pausa de 880 ms. A fronteira terminal localizada após o item lexical “*administração*” é seguida por uma pausa de 650 ms. Já as fronteiras não terminais localizadas após as palavras “*eu*” e “*pra*” não são seguidas por nenhuma pausa silenciosa.

O exemplo (14), representado pela figura (4), mostra outros exemplos de fronteiras prosódicas sem pausas.

Tabela 1: Número de fronteiras terminais e não terminais coincidentes com pausas

Duração da pausa	Coincidência terminais		Coincidência não terminais	
	Número de fronteiras terminais coincidentes	Porcentagem de fronteiras terminais coincidentes	Número de fronteiras não terminais coincidentes	Porcentagem de fronteiras não terminais coincidentes
>= 10 ms	842	67%	675	26%
>= 20 ms	842	67%	675	26%
>= 30 ms	841	67%	675	26%
>= 40 ms	839	67%	675	26%
>= 50 ms	837	67%	673	26%
>= 60 ms	833	67%	673	26%
>= 70 ms	827	66%	668	26%
>= 80 ms	823	66%	667	26%
>= 90 ms	815	65%	663	26%
>= 100 ms	810	65%	660	26%
>= 110 ms	807	65%	655	25%
>= 120 ms	799	64%	646	25%
>= 130 ms	792	63%	639	25%
>= 140 ms	785	63%	630	24%
>= 150 ms	776	62%	629	24%
>= 160 ms	767	61%	623	24%
>= 170 ms	763	61%	615	24%
>= 180 ms	759	61%	607	24%
>= 190 ms	750	60%	595	23%
>= 200 ms	746	60%	587	23%

Adaptado de: Raso, T.; Mittmann, M. M. ; Oliveira, A; 2015., p. 910.

Raso *et al.* (2015) demonstram que não é possível determinar o tipo de fronteira prosódica (terminal ou não terminal) orientando-se por uma duração de pausa específica. Assim, nenhuma duração de pausa silenciosa pode ser considerada como uma variável preditiva de distinção entre fronteiras terminais e não terminais.

Os resultados de Raso *et al.* (2015) e os exemplos dados sugerem que a pausa silenciosa pode ser vista como um parâmetro relevante para a segmentação da fala em unidades marcadas como fronteiras prosódicas, tendo em vista a correlação estabelecida entre certas fronteiras (terminais e não terminais) e a interrupção do sinal sonoro. Todavia, ela não pode ser considerada como o único parâmetro fonético-acústico que orienta a segmentação da fala em

unidades perceptualmente relevantes, pois, muitas fronteiras prosódicas não são seguidas por pausas. Por isso, são necessários outros parâmetros fonético-acústicos, além da presença de pausa silenciosa, para estabelecer a percepção de fronteiras prosódicas ao longo fluxo da fala espontânea.

A frequência fundamental (f_0) é um parâmetro fonético-acústico que expressa o número de ciclos completos de vibração das pregas vocais por segundo. O *pitch* é o correlato perceptual deste parâmetro. Normalmente, o início de uma unidade comunicativamente relevante para a fala é assinalado por um *pitch* alto e o seu final é assinalado por um tom descendente, atingindo muitas vezes o *pitch* mais baixo do falante. Em virtude dessa característica, uma fronteira pode ser estabelecida por meio de uma diferença significativa entre o valor de f_0 medido antes da fronteira e o valor de f_0 medido depois da fronteira. O resultado desta diferença é denominado *reset* de f_0 . De acordo com Zhang (2012, p. 20), “*pitch reset* refers to the readjustment of the pitch height to a higher value in the course of the f_0 declination at junctures.” O *reset* de f_0 pode ser visualizado na imagem abaixo.

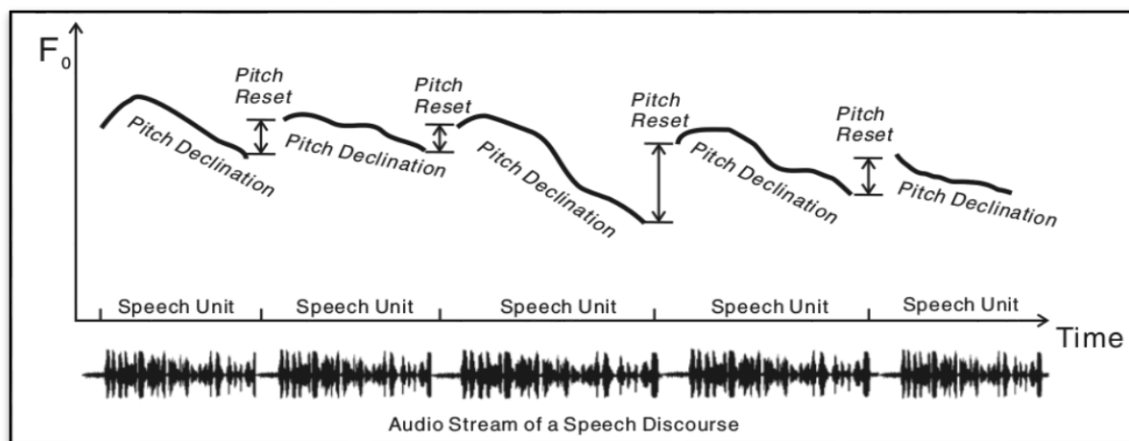


Figura 5: *Reset* de f_0 entre unidades marcadas por fronteiras prosódicas

Fonte: Xie, 2008

Tradicionalmente, a literatura da área estabelece que o *reset* de f_0 marca fronteiras de unidades relevantes para a comunicação humana (Cruttenden, 1997; Yang e Wang, 2002). Contudo, observa-se que diversas unidades delimitadas por fronteiras perceptivelmente

relevantes não são correlacionadas com o *reset* de f0 (Schuetze-Coburn *et al.*, 1991; Couper-Kuhlen, 2006).

Estes resultados sugerem que o *reset* de f0 pode ser relevante para estabelecer fronteiras prosódicas, porque, de fato, há fronteiras que são marcadas por este parâmetro. Contudo, este parâmetro não é suficiente para delimitar de forma precisa as fronteiras que delimitam o enunciado (fronteiras terminais) e as fronteiras internas a ele (fronteiras não terminais). Por isso, enfatiza-se novamente a necessidade de analisar uma série de parâmetros fonético-acústicos que podem contribuir de algum modo na segmentação da fala.

A taxa de elocução é um parâmetro que expressa o número de sílabas fonéticas produzidas pelo falante a cada segundo. A taxa de elocução é responsável pela percepção de uma velocidade de fala mais acelerada ou lenta. A literatura argumenta que a diferença do número de sílabas que o falante produz em um determinado período de tempo também pode ser considerado um parâmetro acústico relevante para o estabelecimento de uma fronteira prosódica, se houver um aumento ou uma diminuição da taxa de elocução em volta da fronteira da unidade (Chafe, 1994; Amir, Silver-Varod e Izre'el, 2004; Swerts *et al.*, 1994; Tyler, 2013).

O alongamento, pré-fronteiriço e fronteiriço, referem-se comumente ao fenômeno fonético em que a duração da sílaba é percebida como maior em relação a outra sílaba. O alongamento é comumente apontado como um fenômeno fonético-acústico marcador de fronteiras prosódicas em várias línguas, como inglês (Price *et al.*, 1991; Turk e Shattuck-Hufnagel, 2007; Wightman *et al.*, 1992), chinês (Duanmu, 1996; Shen, 1992), coreano (Cho e Keating, 2001), francês (Martin, 1982), alemão (Cambier-Langeveld, 1997), sueco (Lindblom e Rapp, 1973), hebraico (Amir, Silver-Varod e Izre'el, 2004).

A forma mais rigorosa de verificar alongamentos em posições pré-fronteiriças e fronteiriças é apresentada por Barbosa (2006). A proposta de Barbosa utiliza a segmentação em unidades VVs⁴, a normalização e a suavização da duração das unidades VVs para avaliar alongamentos ou encurtamentos em posições fronteiriças e pré-fronteiriças.

⁴ Unidade Vogal-Vogal (unidade VV). A unidade VV se diferencia da sílaba fonológica, porque agrupa a rima de uma sílaba com o ataque da sílaba subsequente. Para mais detalhes da importância da unidade VV, veja-se Barbosa (2006).

Exemplo (15) - bmidmasc01 – Áudio 11

*ROB: é uma descida que cê fica se segurando assim / que cê vai perdendo a capacidade de &m / de memória //

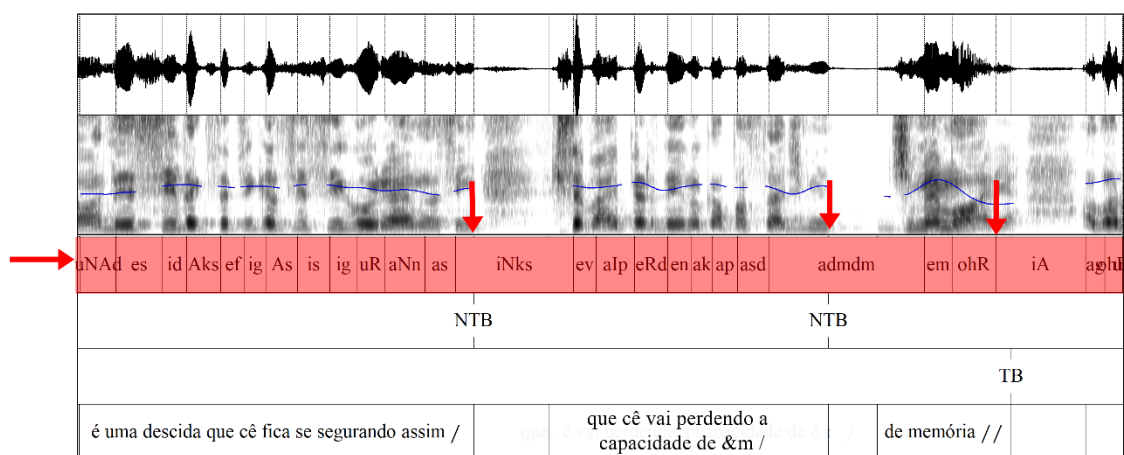


Figura 6: Exemplo de segmentação do sinal acústico em unidades VV utilizando transcrição fonética larga com caracteres ASCII

A normalização da duração das unidades VVs é um método mais adequado para o estudo da estrutura rítmica do enunciado, porque leva em consideração a média e a variância de cada fone que compõe a unidade VV em análise. Este tipo de análise é consistente tanto em relação à produção da fala como também à sua percepção, conforme vários estudos experimentais realizados em diversas línguas mostram (Barbosa, 1994; Barbosa, 1996a; Barbosa, 1996b; Dogil, 1987; Chistovich e Ogorodnikova, 1982; Geisler, 1988; Scott e Wise, 2003 entre outros). Para mais detalhes deste procedimento, veja-se Barbosa (2006).

1.4. Segmentação automática da fala

Há várias ferramentas computacionais já desenvolvidas para a segmentação do sinal acústico em unidades menores que a palavra, como a segmentação em fones ou em sílabas. Uma das ferramentas computacionais já desenvolvidas para a segmentação do sinal acústico em unidades menores que a palavra é o *script* para Praat *BeatExtractor* (Barbosa, 2006). O *BeatExtractor* é uma ferramenta que permite analisar a evolução da duração ao longo dos enunciados de fala, porque segmenta automaticamente o sinal acústico em *onsets* de vogais consecutivos, isto é, em unidades VVs. Outras ferramentas amplamente utilizadas para segmentar a fala em unidades definidas acusticamente são as ferramentas destinadas à segmentação fonética (Svendsen e Soong, 1987; Van Hemert, 1991; Selmini e Violaro, 2008).

Até então, há poucas ferramentas disponíveis para segmentação da fala em unidades maiores do que a palavra. Normalmente, este tipo de segmentação automática da fala é aplicada na tentativa de fornecer uma “pontuação” aos textos de fala, isto é, identificar limites de sentenças no texto falado (Cutler, Dahan e Donselaar, 1997). Todavia, é importante ressaltar que a adoção deste tipo de segmentação se manifesta como uma alternativa inadequada, porque as fronteiras das sentenças não são sempre correlacionadas com as fronteiras de natureza prosódica.

Outras ferramentas mais adequadas buscam segmentar o fluxo da fala em unidades maiores que a palavra por meio da avaliação do sinal acústico. Uma das ferramentas computacionais já desenvolvidas para identificar fronteiras prosódicas é o algoritmo denominado ANALOR (Avanzi, Lacheret-Dujour e Victorri, 2008). A partir do francês, a ferramenta busca detectar por meio de um único algoritmo pausas de pelo menos 300 ms, alongamentos silábicos, variações de f_0 , *resets* de f_0 e proeminências relevantes, para segmentar a fala em unidades marcadas por meio de fronteiras prosódicas. A ferramenta tem como grande vantagem o fato de ser destinada à segmentação da fala lida e da fala espontânea. Os resultados apontam uma convergência de 83,8% entre a segmentação manual de segmentadores e a segmentação automática. O algoritmo requer, no entanto, uma segmentação silábica prévia.

O modelo de detecção automática de fronteiras prosódicas proposto por Ni *et al.* (2012) detecta com alta acurácia fronteiras prosódicas terminais marcadas manualmente por

segmentadores. O modelo proposto por estes autores utiliza dois classificadores distintos, variáveis de previsão acústicas, lexicais e sintáticas para prever a realização de fronteiras em inglês e mandarim. Para os fins deste trabalho, serão expostos apenas os resultados obtidos pelo algoritmo utilizando exclusivamente as variáveis de previsão fonético-acústicas. Em mandarim, com 30 medidas, o modelo alcança 85% de precisão para detectar fronteiras. Em inglês, com 24 medidas, o modelo alcança 82% de precisão.

Outros autores também propõem modelos baseados em tipos mistos de variáveis de previsão. Kim (2004) desenvolveu um modelo para detectar fronteiras prosódicas que utiliza variáveis de previsão acústicas e lexicais. O teste final do modelo mostra que os parâmetros prosódicos identificam com 74% e 56% de precisão fronteiras terminais e não terminais respectivamente.

Com exceção do modelo proposto por Kim (2004), as ferramentas citadas não fazem distinção entre os dois principais tipos de fronteiras prosódicas apontados na literatura da área. Normalmente, detecta-se a fronteira, mas não é especificado se ela corresponde a fronteira terminal ou não terminal. Diferentemente destes autores, o trabalho aqui desenvolvido busca compreender o estabelecimento de fronteiras prosódicas nos planos fonético-acústico e perceptual para futuramente possibilitar a construção de uma ferramenta computacional de detecção automática de fronteiras prosódicas, que seja capaz de distinguir fronteiras terminais de não terminais.

2. Metodologia

Neste capítulo, são apresentados em sequência o *corpus* C-ORAL-BRASIL, os dados usados, o tratamento recebido pelos dados e as análises estatísticas desenvolvidas.

2.1. Apresentação do *corpus* C-ORAL-BRASIL

Os *corpora* de fala espontânea, como o *corpus* C-ORAL-BRASIL (Raso e Mello 2012; Raso e Mello em preparação), são de suma importância para várias áreas de investigação linguística. Eles são importantes para desenvolver pesquisas em linguística teórica descritiva e em linguística aplicada nas diversas linhas de pesquisa. Os *corpora* de fala espontânea também são muito importantes para desenvolver novas tecnologias destinadas à fala, como por exemplo, sistemas de reconhecimento, síntese e extração de informação oral. Deste modo, um *corpus* que inclui produções linguísticas de fala decorrentes de situações naturais de interação verbal entre falantes proporciona uma compreensão mais ampla e profunda sobre o funcionamento da língua em uso e os fenômenos linguísticos em contextos reais.

A grande quantidade de material oferecida pelos *corpora* C-ORAL-BRASIL I e II constitui um conjunto de dados extremamente rico para ser analisado com o objetivo de investigar a produção e a percepção de fronteiras prosódicas. A compilação dos *corpora* C-ORAL-BRASIL seguiu os mesmos critérios adotados pelo projeto C-ORAL-ROM para compilação de *corpora* de espanhol, francês, italiano e português europeu (Cresti e Moneglia 2005), sendo desta forma comparável em arquitetura e amostragem ao C-ORAL-ROM. O C-ORAL-BRASIL se configura como um *corpus* de terceira geração, com a transcrição do texto e o áudio alinhado. Os dados dos *corpora* C-ORAL-BRASIL I e II são representativos do PB na diatopia mineira, sobretudo da região metropolitana de Belo Horizonte. Estes *corpora* possibilitam o desenvolvimento de uma grande quantidade de estudos linguísticos, dentre eles, o estudo sobre os parâmetros fonético-acústicos associados às fronteiras prosódicas, já que o C-ORAL-BRASIL é segmentado prosodicamente em unidades entonacionais por segmentadores *experts*. O processo de transcrição e segmentação do *corpus* passou por diversas

etapas, desde a formação dos transcritores, diferentes fases de revisão, até a realização de testes estatísticos que validam o acordo entre os transcritores. Essa validação foi realizada por meio do teste Kappa, resultando em um acordo geral de 0,87 para quebras terminais e de 0,78 para não terminais (Raso e Mello, 2012). Para mais detalhes sobre os critérios de segmentação prosódica adotada no C-ORAL-BRASIL, veja-se Mello *et al.*, 2012.

O C-ORAL BRASIL prevê a compilação de textos de fala espontânea divididos em um *corpus* informal em contexto natural, um formal em contexto natural, um em diamesia telefônica e um em mídia televisiva e radiofônica. A fala espontânea compreende as instâncias de comunicação humana em contexto natural, sem intervenção do pesquisador, abrangendo interações dialógicas, monológicas e conversacionais. As tipologias interacionais incluídas na parte informal do *corpus* são diálogos, monólogos e conversações em contexto natural informal público e privado. As tipologias interacionais incluídas no C-ORAL-BRASIL II são diálogos, monólogos e conversações em contexto natural formal público, em mídia (televisiva ou radiofônica) e telefone.

2.2. Dados

Este trabalho usou 7 trechos de fala espontânea monológica, extraídos tanto da parte informal, quanto da parte formal e de mídia do *corpus* C-ORAL-BRASIL. Os textos são, portanto, relativos a três *corpora* de fala espontânea: fala formal em contexto natural, fala informal em contexto natural e fala televisiva. Os trechos são compostos por em média 191 palavras e são distribuídos da seguinte forma em uma amostra (I):

Tabela 2: Amostra I

Contexto	Tipologia	Sexo	Texto	Duração	Palavras
Natural informal	Monólogo	Masculino	bfammn11	01'11''	189
Natural informal	Monólogo	Masculino	bfammn24	00'58''	151
Mídia formal	Monólogo	Masculino	bmidmasc01	01'23''	212
Mídia formal	Monólogo	Masculino	bmidmasc02	01'21''	238
Mídia formal	Monólogo	Masculino	bmidmasc03	01'07''	183
Natural formal	Monólogo	Masculino	bnatmasc01	01'30''	205
Natural formal	Monólogo	Masculino	bnatmasc02	01'09''	161
		Total	7	08'39''	1339

2.3. Tratamento dos dados

Cada trecho foi segmentado autonomamente por quatorze segmentadores *experts*, membros da equipe do Laboratório de Estudos Empíricos e Experimentais da Linguagem (LEEL) da Universidade Federal de Minas Gerais. Os segmentadores eram apresentados às gravações e aos textos transcritos, sem nenhuma marcação de fronteiras. A tarefa dos segmentadores consistia em marcar os pontos em que as fronteiras prosódicas eram percebidas, inserindo uma barra simples (/) para fronteira não terminal, uma barra dupla (//) para fronteira terminal e não inserindo nenhum tipo de marcação para não fronteira.

Todos os segmentadores haviam sido treinados anteriormente e já apresentavam, mesmo que em diferentes graus, experiência em segmentação prosódica da fala. O treinamento, de duração média de três meses, teve como propósito levar os transcritores a tomar consciência das pistas prosódicas que, consistentemente, promovem a percepção de fronteiras prosódicas no fluxo da fala. O treinamento consistiu em um processo complexo que envolveu várias etapas. Inicialmente, introduziu-se o tema aos anotadores. Em seguida, os anotadores passaram por uma formação que envolveu a realização de exercícios de anotação de percepção de fronteiras, sessões semanais de treinamento e discussão sobre os exercícios. Para mais detalhes, veja-se Mello *et al.*, 2012. O exemplo (15) mostra a segmentação perceptual de um trecho de fala espontânea monológica realizada por um dos segmentadores.

Exemplo (15) – bnatmasc01 – Áudio 12

*JAC: basta lembrar que / pra Platão / o mundo / o mundo em que a gente vive / né / ele / ele é feito pelo demiurgo / que mimetiza / os paradigmas / ou as ideias // então a gente tem uma concepção de que isso faz parte do processo da natureza / né / essa / essa / perspectiva / &he / mimética // e Platão que / &he / critica / em várias esferas a / mimese / principalmente a mimese dos poetas / ele considera / que a / a / a mimese é um certo tipo de produção / né “mimesis poiesis tis estin” // né // tanto é um certo tipo de produção que ela produz / o mundo // na própria perspectiva dele / não é / inclusive não é um tipo de &pro / de produção / &he / desclassificado // &he / eu diria que / essa / primeira perspectiva / ela &he / a mimese / &he / nós poderíamos dizer que ela / é responsável pela forma como o homem / tá se apropriando do mundo // &en / enquanto ela implica um reconhecimento / de que o mundo nu é tocado / diretamente / mas a gente tem / as mediações aí / que é a nossa ciência / o nosso discurso / e / e e / e essa nossa ciência e o nosso discurso ele necessita dessas / mediações que o representa / na qualidade dum / de / do conhecimento / né / que / a qualidade do conhecimento / dependendo da qualidade dessas / representações //

O acordo entre segmentadores foi avaliado pelo coeficiente de Kappa (Fleiss, 1971). O Kappa, tanto em sua versão proposta por Cohen (1960), como na versão proposta por Fleiss (1971), é uma medida de concordância entre observadores. O coeficiente Kappa pode assumir qualquer valor entre 0 e 1. Se o valor for 0, não existe acordo entre segmentadores. Se o Kappa obtido for >0 , existe alguma concordância. Quanto maior for o coeficiente de Kappa, maior é o acordo entre segmentadores.

Os coeficientes de Kappa foram calculados através da função `kappam.fleiss()` do pacote `irr` (Gamer *et al.*, 2012) disponível no *software* estatístico R (R Development Core Team, 2017). Para os trechos de fala incluídos na amostra I, o acordo geral obtido foi de 0,80 na marcação de fronteiras terminais e de 0,75 para fronteiras não terminais, resultados de grande relevância estatística.

Após a segmentação perceptual realizada pelo grupo de segmentadores com base de oitiva, os trechos foram anotados em cinco camadas usando o *software* para análises fonéticas e acústicas Praat (Boersma e Weenink, 2015). Foram adotadas as seguintes camadas de anotação do TextGrid do Praat:

1. Segmentação em unidades Vogal-Vogal (unidades VVs) e etiquetagem utilizando transcrição fonética larga em caracteres ASCII;
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 do intervalo referente a pausas silenciosas;
5. Transcrição ortográfica do texto do enunciado.

As unidades VVs, 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. Em outras palavras, as unidades VVs compreendem o segmento acústico que vai de um *onset* vocálico até o *onset* seguinte, incluindo os segmentos consonantais e as pausas presentes entre eles (Barbosa, 2006). A unidade VV 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 VVs, porque estas unidades permitem detectar os pontos de transição consoante-vogal (CV). É importante marcar a transição CV, porque esta transição permite detectar o início de cada vogal, que funciona como ponto de ancoramento para a produção e a percepção da fala. A consequência disso é a implementação de um fluxo de fala vocálico contínuo perturbado, ou seja, interrompido pela realização de consoantes (Barbosa, 2006).

A transição CV funciona como ponto de ancoramento por três motivos. Primeiramente, porque a transição CV é o ponto específico (*beat*) de maior taxa de crescimento de amplitude da produção vocálica. Depois, porque se observa uma maior precisão perceptiva do ouvido humano para perceber a transição consoante e vogal subsequente (CV) do que entre vogal e consoante seguinte (VC) (Marcus, 1981). Por fim, porque a transição CV 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).

Sendo assim, a segmentação em unidades VVs é mais adequada para o estudo da estrutura rítmica da fala e é consistente tanto em relação à produção da fala como também à sua

percepção, conforme mostram vários estudos experimentais realizados em diversas línguas (Barbosa, 2006).

A figura (7) mostra um exemplo das cinco camadas de anotação do Praat para uma unidade entonacional marcada pelos segmentadores.

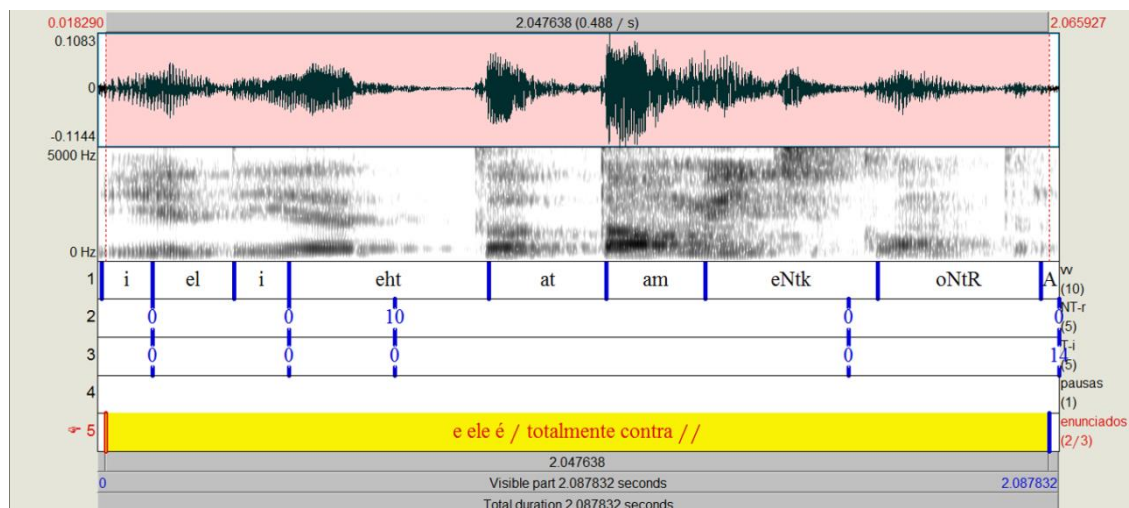


Figura 7: De cima para baixo: forma de onda, espectrograma de banda larga e camadas de anotação no Praat para trecho de fala espontânea. Destaca-se exemplo de camadas de anotação do Praat.

Trecho: e ele é / totalmente contra //

Na figura (7), é possível observar o seguinte. A fronteira não terminal correspondente à unidade VV “eht” foi percebida por dez segmentadores, a fronteira terminal correspondente à unidade VV “A” foi percebida por todos os 14 segmentadores. Não há ocorrências de pausa silenciosa diante das unidades VVs correspondentes às fronteiras. Nas camadas 2 e 3 foram marcadas com o número zero as posições em que nenhum segmentador percebeu fronteira prosódica. Estas posições também foram anotadas, porque cada fronteira entre palavras fonológicas pode ser considerada uma potencial posição de fronteira prosódica.

Desenvolveu-se uma versão estendida do *script ProsodyDescriptor* (Barbosa, 2013) para extrair uma série de parâmetros acústicos ao longo do sinal de fala. Essa versão estendida, denominada *BreakDescriptor* (Barbosa, 2016), extrai os parâmetros acústicos em

todas as unidades VVs em uma janela centrada em toda fronteira de palavra fonológica⁵, o que inclui as posições percebidas pelos segmentadores como fronteiras e também as posições percebidas como não fronteiras. O *BreakDescriptor* é capaz de extrair uma série de parâmetros fonético-acústicos que permitiram o aprendizado automático do que é relevante para assinalar os dois tipos de fronteira bem como a ausência de qualquer fronteira.

É importante lembrar que este trabalho tem como objetivo principal estabelecer uma correlação entre a percepção humana de fronteiras e os parâmetros fonético-acústicos. O *BreakDescriptor* extrai os parâmetros em todos os casos, inclusive nos casos de não fronteira, porque o conjunto de parâmetros associados às não fronteiras são fundamentais para a classificação estatística. Deste modo, o classificador requer pelo menos duas categorias para realizar a classificação. No nosso caso, as categorias usadas são a presença e a ausência de fronteira prosódica através de dois modelos, um para fronteiras terminais e outro para as fronteiras não terminais.

A presença de fronteira terminal se restringe às posições de fronteira conclusiva. Quando se trata da ausência de fronteira terminal, naturalmente são inclusas as posições de não fronteira, mas também são inclusas as posições de fronteira não terminal, porque as posições de fronteiras não terminais necessariamente são posições de ausência de fronteira terminal. Semelhantemente, a categoria presença de fronteira não terminal inclui as posições de fronteira não conclusiva. A categoria ausência de fronteira não terminal inclui as posições de não fronteira, como também posições de fronteira terminal.

A posição exata da fronteira é dada por meio de um processo fonológico que leva em conta a coincidência entre o ponto de fronteira prosódica e o ponto de fronteira de palavra fonológica. É fundamental ressaltar também que os parâmetros fonético-acústicos mais relevantes para o estabelecimento de certos fenômenos linguísticos marcados prosodicamente não são necessariamente localizados exatamente no ponto em que o fenômeno prosódico é percebido pelos falantes. Um trabalho que mostra este fenômeno é o trabalho de Frazier *et al.* (2004).

Frazier *et al.* (2004) mostram que o ponto de fronteira prosódica se relaciona com a localização da fronteira em si, mas também se relaciona com padrões prosódicos ao longo da fala. Por este motivo, o *script* focaliza a extração dos parâmetros nas posições em fronteira de palavra fonológica e nas posições em torno das fronteiras entre palavras fonológicas, a fim de verificar na análise estatística se algum parâmetro acústico relevante para a marcação de

⁵ Entende-se como palavra fonológica uma unidade prosódica identificada por recair um acento lexical primário.

fronteiras prosódicas ou para a sua não marcação localiza-se pouco antes ou pouco após a posição à qual a fronteira ou a não fronteira prosódica foi atribuída.

O critério de referência do *script* para estabelecer se uma unidade VV localizada em limite de palavra fonológica será usada para a análise de correlato de fronteira ou não fronteira é o acordo entre segmentadores humanos. Estabeleceu-se que uma posição é usada para análise de correlato de fronteira, desde que pelo menos 7 segmentadores treinados indicassem a unidade VV como uma posição de fronteira prosódica, ou seja, pelo menos 50% de acordo entre anotadores humanos. Assim, as demais unidades VVs, também localizadas em limites de palavras fonológicas, são consideradas pelo *script* como não fronteiras. Como consequência deste critério, são consideradas não fronteiras todos os casos em que menos de 7 segmentadores marcaram a presença de fronteira.

O critério de referência do *script* para extrair os parâmetros em torno das unidades VVs percebidas como fronteiras e não fronteiras indicadas pelos segmentadores é estabelecer duas janelas de trabalho antes e após a VV de análise. As janelas estabelecidas incluem 10 unidades VVs à esquerda e 10 unidades VVs à direita de cada unidade VV analisada. O *script* adota como padrão de anotação das unidades inclusas em cada janela o uso de letras seguidas por números, conforme a figura (8) mostra.

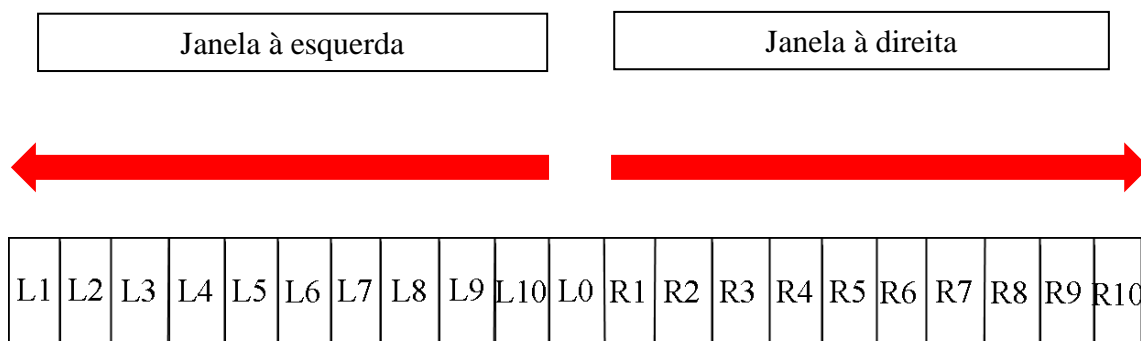


Figura 8: Janelas de trabalho do *script*

As letras L (*left*) e R (*right*) indicam a posição das duas janelas de trabalho em relação à posição de análise, indicada como L0. Os demais valores numéricos especificam as unidades VV à esquerda ou à direita dessa posição. A seguir, a figura (9) mostra no ponto

central uma fronteira terminal percebida pelos segmentadores e as janelas usadas pelo *BreakDescriptor* para extrair os parâmetros fonético-acústicos.

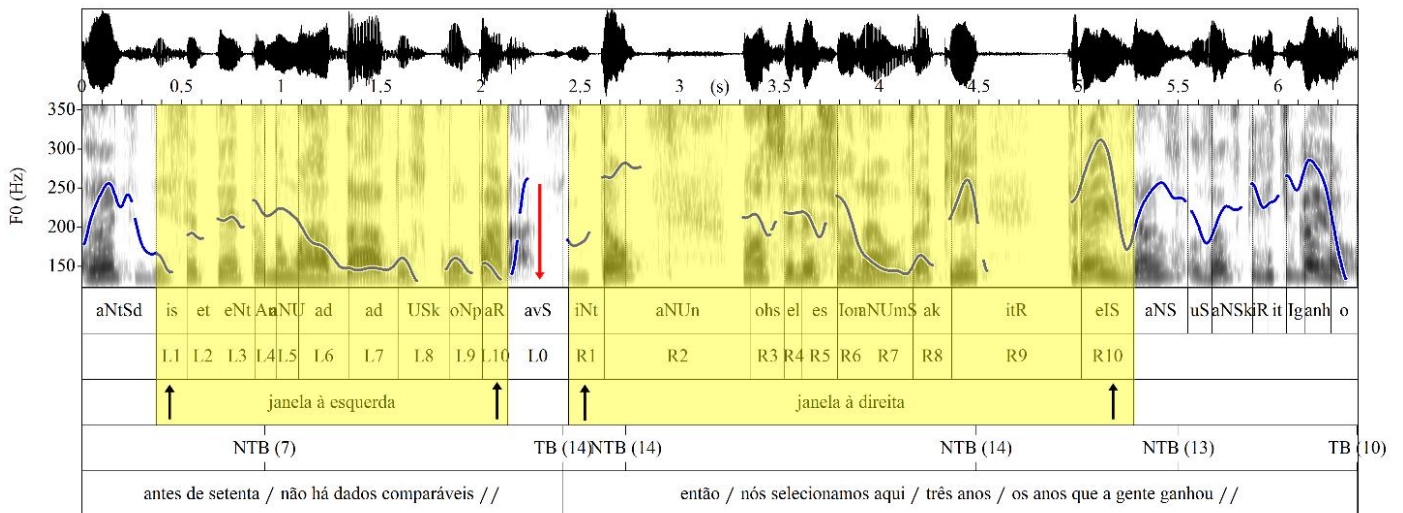


Figura 9: De cima para baixo: forma de onda, espectrograma de banda larga e camadas de anotação no Praat para trecho de fala espontânea. Destacam-se as janelas usadas para análise de fronteira.

Trecho: antes de setenta / não há dados comparáveis // então / nós selecionamos aqui / três anos / os anos que a gente ganhou //

Na figura (9), destaca-se uma unidade VV em que todos os 14 segmentadores marcaram fronteira terminal. Para esta posição destacada, são calculados parâmetros fonético-acústicos nas 10 unidades VVs anteriores, na unidade que imediatamente marca a posição da fronteira e nas 10 unidades VVs posteriores à posição de análise (área sombreada do espectrograma).

Na figura (9), também são destacadas por meio das setas pretas as unidades VVs mais próximas e distantes da fronteira terminal central percebida por todos os 14 segmentadores. Observa-se que a unidade VV L0 é a unidade que precede imediatamente a posição da fronteira terminal; as unidades VV L1 e R10 são as unidades mais distantes da fronteira entre aquelas consideradas pelo *script*; as unidades L10 e R1, por sua vez, são as mais próximas. Para detalhar melhor o funcionamento do *script*, a figura (10), apresentada a seguir,

mostra exemplos de unidades VVs consideradas pelo *script* como ponto de fronteiras e não fronteiras, em virtude do acordo obtido entre os segmentadores treinados.

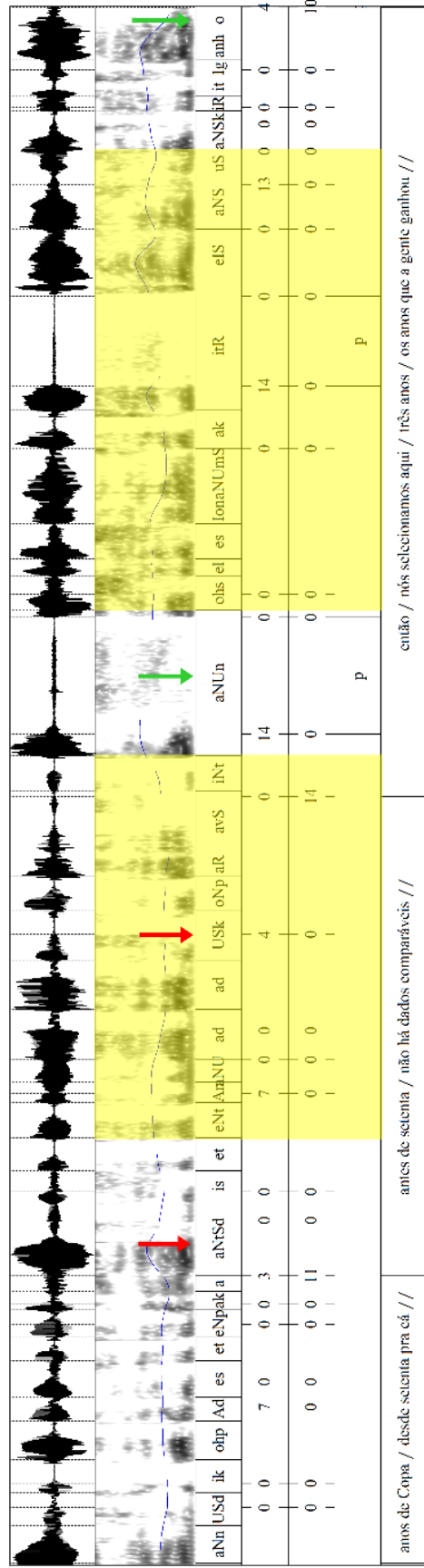


Figura 10: De cima para baixo: forma de onda, espectrograma de banda larga e camadas de anotação no Praat para trecho de fala espontânea. Destacam-se as janelas usadas para análise de correlato de fronteira do ponto central.

Trecho: anos de Copa / desde setenta pra cá // antes de setenta / não há dados comparáveis // então / nós selecionamos aqui / três anos / os anos que a gente ganhou //

Na figura (10), evidencia-se a unidade VV “aNUn” e as janelas usadas pelo *BreakDescriptor* para análise de correlato de fronteira prosódica não terminal, porque, na posição central destacada, 14 segmentadores marcaram como fronteira não terminal. A área sombreada indica as dez unidades VV anteriores e posteriores consideradas para análise desta fronteira. Por meio das setas vermelhas são evidenciadas duas posições que foram consideradas pelo *BreakDescriptor* como não fronteiras, porque nenhum segmentador (ou menos de sete) indicou a presença de fronteira. Assim, a unidade VV “aNtsd” e a unidade VV “USk”, destacadas na imagem 5, são exemplos de unidades VVs usadas para a análise de correlato de não fronteira.

Observa-se que a unidade VV “o”, indicada por uma das setas verdes, obteve desacordo entre os segmentadores quanto ao tipo de fronteira prosódica. Entre os segmentadores, quatro indicaram fronteira não terminal e dez indicaram fronteira terminal. Para este caso específico, a posição foi considerada pelo *script* como uma posição de fronteira terminal, pois, o *script* sempre considera o acordo de pelo menos 50% entre segmentadores.

Observou-se também a ocorrência de casos de empate entre segmentadores humanos no que diz respeito à percepção de fronteiras. Os casos de empate correspondem aos pontos em que todos os 14 segmentadores indicaram a presença de fronteira prosódica, contudo, sete segmentadores indicaram fronteira terminal e os outros sete indicaram fronteira não terminal. Para estes casos, como pelo menos sete pessoas obtiveram acordo quanto ao tipo de fronteira prosódica, o *script* considera a unidade VV correspondente como ponto de fronteira prosódica terminal e não terminal. Esta escolha metodológica é passível de discussão, porque uma única unidade VV é considerada na análise tanto como ponto de fronteira terminal, quanto de fronteira não terminal. Todavia, optou-se por manter o critério de acordo de pelo menos 50% entre segmentadores humanos, independentemente dos casos de empate entre segmentadores, porque este trabalho visa fornecer uma exploração inicial sobre um modelo automático de detecção de unidades entonacionais a partir de parâmetros fonético-acústicos relevantes para produção e percepção das fronteiras prosódicas.

No total, para cada unidade VV, são calculadas 111 medidas fonético-acústicas de natureza global e local, que permitem a análise de fenômenos prosódicos ao longo do sinal acústico. Para o treinamento inicial da ferramenta de segmentação automática, foi usada a amostra I correspondente à fala monológica masculina.

Optou-se por realizar o treinamento inicial usando somente os trechos de fala monológica masculina por duas razões. A primeira razão é a diferença de frequência

fundamental (f0) entre homens e mulheres; a segunda é a exclusão da variável sexo na constituição de um primeiro modelo estatístico para detectar automaticamente fronteiras prosódicas no fluxo da fala espontânea. A tabela (3) mostra todas as ocorrências de fronteiras prosódicas que pelos 50% dos segmentadores perceberam na amostra I.

Tabela 3: Fronteiras prosódicas

Tipo de fronteira	Quantidade	Porcentagem
Terminal	70	19,3
Não terminal	293	80,7
TOTAL	363	100

A amostra I é composta por 363 fronteiras prosódicas, das quais 70 são terminais e 293 não terminais. Contudo, entre as fronteiras não terminais, há 46 fronteiras não terminais localizadas após *retractings*, 14 após tomadas de tempo e 8 após o item lexical “né”, totalizando 68 ocorrências. Os *retractings* são unidades retratadas pelos falantes, as tomadas de tempos são unidades que refletem o esforço processual do falante para elaborar verbalmente um pensamento complexo, elas compreendem o que comumente é chamado de pausa preenchida. A interjeição “né” geralmente é uma unidade, que tem como função principal regular a interação verbal oral. Tradicionalmente, ela pertence à categoria de marcador discursivo.

Inicialmente, a análise estatística usou todas as ocorrências de fronteiras terminais e não terminais presentes na amostra I. Depois, em virtude de uma alta concentração de erros para detectar fronteiras localizadas diante de *retractings*, tomadas de tempo e “né”, a amostra I foi parcialmente reduzida, totalizando 295 fronteiras, das quais 225 são não terminais e 70 terminais.

2.3.1. Parâmetros fonético-acústicos extraídos pelo *BreakDescriptor*

As propriedades físicas relativas à produção dos sons extraídos automaticamente pelo *script BreakDescriptor* compreendem cinco grupos de parâmetros fonético-acústicos clássicos:

- I. Medidas de taxa de elocução (*speech rate*) e ritmo;
- II. Medidas de duração normalizada dos segmentos silábicos;
- III. Medidas de frequência fundamental (f_0);
- IV. Medidas de intensidade;
- V. Medidas de pausa.

As medidas do grupo de *speech rate* e ritmo, de duração normalizada do segmento silábico e de pausa são parâmetros rítmicos e temporais, as medidas do grupo de f_0 são estritamente melódicos, as medidas do grupo de intensidade se relacionam ao esforço vocal do falante. As próximas seções explicam cada um dos parâmetros fonético-acústicos e as medidas calculadas automaticamente nos cinco grupos. A fim de se evitar ambiguidades com nomenclaturas, optou-se por padronizar dois aspectos inter-relacionados entre si no que diz respeito às janelas de trabalho do *script*. Como já foi dito, para extrair os parâmetros em cada fronteira de palavra fonológica, o *script* utiliza duas janelas de trabalho, compostas cada uma delas por 10 unidades VVs. Os termos janela à esquerda e janela à direita referem-se às janelas de trabalho, considerando-as de forma global. Os termos unidade L1, L2, L3, L4, L5, L6, L7, L8, L9 ou L10 (à esquerda da unidade VV de análise) e unidade R1, R2, R3, R4, R5, R6, R7, R8, R9 ou R10 (à direita da unidade VV de análise) referem-se às unidades VVs, considerando-as de forma local, lembrando que L10 e R1 são as unidades mais próximas à posição de análise.

Serão utilizados três trechos de fala espontânea monológica masculina para explicar e exemplificar todas as medidas calculadas automaticamente pelo *BreakDescriptor*. Estes trechos são compostos pelas janelas à esquerda e à direita utilizadas para análise de três

fronteiras terminais percebidas por pelo menos 50% dos segmentadores humanos. Para facilitar a visualização das janelas globais consideradas, nas transcrições dos três trechos, serão marcados em negrito as janelas à esquerda e à direita das fronteiras.

Áudio 13

Trecho (1): antes de setenta / não há dados comparáveis // então / nós selecionamos aqui / três anos /

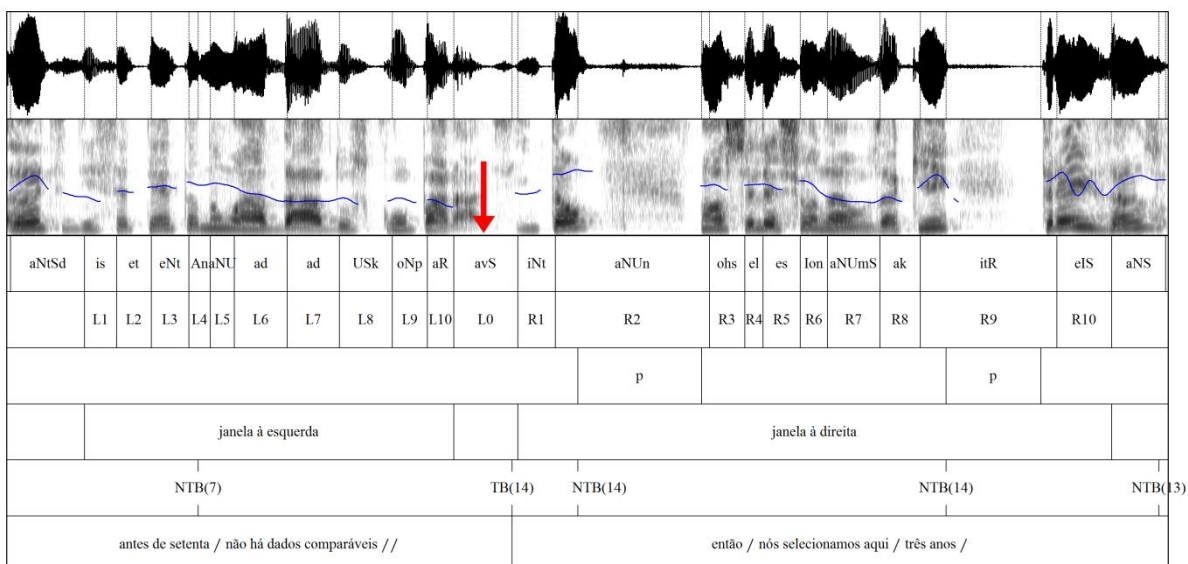


Figura 11: Trecho (1)

A fronteira terminal destacada no trecho (1) foi percebida por 14 segmentadores. No trecho (1), a janela à esquerda da fronteira terminal é iniciada pela a unidade VV “is” e finalizada pela unidade VV “aR”, a janela à direita da fronteira terminal é iniciada pela unidade VV “iNt” e finalizada pela unidade VV “eIS”.

Abaixo, o trecho (2) é transcrito e ilustrado.

Áudio 14

Trecho (2): **daquele / daquele cômodo / acesa // a / a madrugada ainda tava / ainda /**

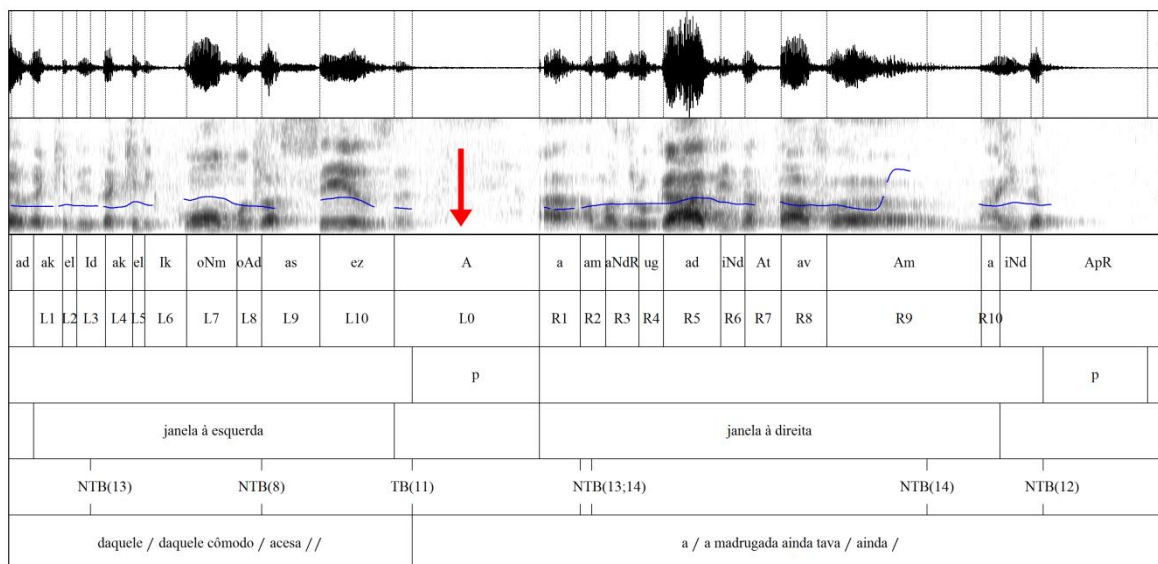


Figura 12: Trecho (2)

A fronteira terminal destacada no trecho (2) foi percebida por 11 segmentadores. No trecho (2), a janela à esquerda da fronteira terminal é iniciada pela a unidade VV “ak” e finalizada pela unidade VV “ez”, a janela à direita da fronteira terminal é iniciada e finalizada pela unidade VV “a”.

Áudio 15

Trecho (3): cada um deles / faz mais de dez mil sinapses // quer dizer / o que a / a forma como essas

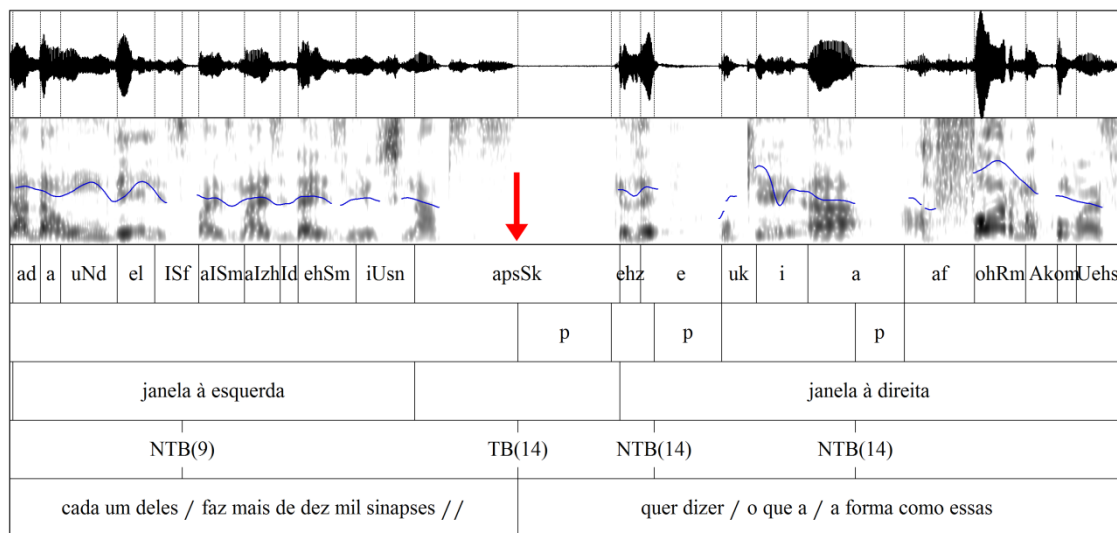


Figura 13: Trecho (3)

A fronteira terminal destacada no trecho (3) foi percebida por 14 segmentadores. No trecho (3), a janela à esquerda da fronteira terminal é iniciada pela a unidade VV “ad” e finalizada pela unidade VV “iUsn”, a janela à direita da fronteira terminal é iniciada pela unidade VV “ehz” e finalizada pela unidade VV “Uehs”.

I. Medidas de taxa de elocução (*speech rate*) e ritmo

A taxa de elocução é um parâmetro fonético-acústico que expressa o número de unidades VVs que o falante produz em um determinado período de tempo. Normalmente, esta taxa é responsável pela percepção de uma velocidade de fala mais acelerada ou lenta. A taxa de

elocução calculada automaticamente pelo *BreakDescriptor* expressa o número de unidades VVs produzidas a cada segundo pelo falante em cada janela global de trabalho estabelecida pelo *script*. Em outras palavras, ela consiste da razão entre o número unidades VVs produzidas em uma janela de 10 unidades VV à direita e à esquerda da unidade VV analisada e a duração total em segundos da janela.

Na figura (14), destaca-se a unidade VV utilizada para análise de correlato de fronteira terminal. A medida taxa de elocução à esquerda da unidade VV de análise (*srl*) é obtida por meio da divisão do número de unidades VVs produzidas na janela à esquerda da fronteira pela duração total da janela à esquerda. Para a fronteira terminal destacada no trecho (1), a medida *srl* é obtida por meio da divisão de 10 por 1,76 s (= 5,7 unidades VV/segundo). A medida taxa de elocução à direita da unidade VV de análise (*srr*) é calculada de forma análoga. Para a fronteira terminal destacada no trecho (1), a medida *srr* é obtida por meio da divisão de 10 por 2,84 s (= 3,5 unidades VV/segundo).

O *BreakDescriptor* também calcula um terceiro parâmetro, qual seja, a diferença de taxa de elocução entre janela à direita e à esquerda da unidade VV de análise (*srd*). O parâmetro *srd* é obtido por meio da diferença entre *srr* e *srl*. Ele é importante para assinalar o aumento ou declínio da taxa de elocução, dada uma posição de análise. Deste modo, com a medida *srd* verifica-se o impacto da mudança da taxa de elocução na percepção de fronteira ou de não fronteira.

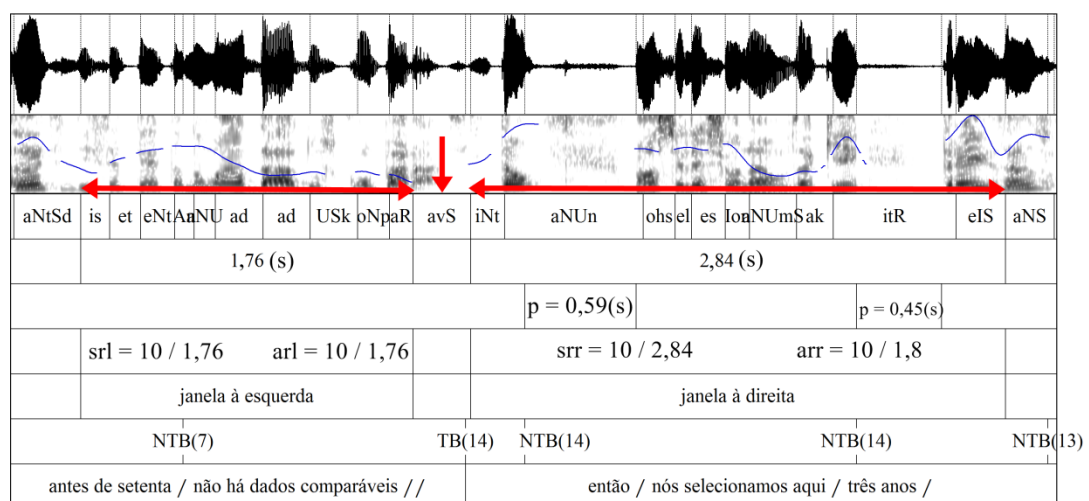


Figura 14: Medidas de taxa de elocução e articulação para a fronteira terminal destacada no trecho (1) – *srl*, *srr*, *arr* e *arl*. Trecho (1): antes de setenta / não há dados comparáveis // então / nós selecionamos aqui / três anos /

Outro parâmetro fonético-acústico calculado automaticamente no grupo de medidas de taxa de elocução e ritmo é a taxa de unidades VV não salientes produzidas por segundo. As unidades VVs não salientes são aquelas que não correspondem a picos de duração normalizada. Este parâmetro é semelhante à taxa de articulação, porque, em seu cálculo, são incluídas apenas as durações dos segmentos sonoros, excluindo assim, as durações de pausas silenciosas. Esta taxa consiste da razão entre o número de unidades VVs não salientes produzidas em uma janela global de 10 unidades VVs à direita e também à esquerda da unidade VV analisada e a duração total em segundos da janela, excluindo as durações de pausas silenciosas.

Na figura (14), destaca-se a unidade VV utilizada para análise de correlato de fronteira terminal. A taxa de articulação à esquerda da unidade VV de análise (arl) é obtida por meio da divisão do número de unidades VVs produzidas na janela à esquerda da unidade VV de análise pela duração da janela à esquerda, excluindo as durações de pausa silenciosa. Como a janela à esquerda da unidade VV de análise destacada na figura (14) não apresentou nenhuma pausa silenciosa, a taxa de unidades não salientes à esquerda é igual à taxa de elocução à esquerda. A medida taxa de articulação à direita da unidade VV de análise (arr) é calculada de forma semelhante, com a diferença que se considera a janela à direita da unidade VV de análise, também excluindo as durações de pausa silenciosa presentes na janela. No exemplo acima, são desconsiderados 0,59 e 0,45 s de pausa silenciosa para o cálculo de arr, o que resulta numa duração sem pausa de 1,8 s (assim a taxa de articulação à direita é de 5,6 unidades VV/s).

O último parâmetro calculado pelo *BreakDescriptor* relacionado à taxa de articulação é a diferença de taxa de articulação entre janela à direita e à esquerda da unidade VV de análise (ard). Deste modo, a medida ard é obtida por meio da subtração simples entre arr e arl. Ele é importante para assinalar o aumento ou declínio da taxa de unidades não salientes por segundo, dada uma posição de análise.

II. Medidas de duração normalizada dos segmentos silábicos

A normalização das durações através do cálculo de z-score é um procedimento padrão utilizado para atenuar o efeito da duração intrínseca do número de segmentos da unidade silábica e da implementação do acento lexical. Em geral, o z-score minimiza os efeitos

microprosódicos que não têm função linguística, ressaltando a informação prosódica linguisticamente relevante e indicando o quanto a duração das unidades VVs afasta-se em relação à média da soma da duração dos fones que compõem a unidades. Neste trabalho, a duração das unidades VVs, segmentadas na primeira camada do Praat, é normalizada automaticamente por meio do *BreakDescriptor*.

O tipo de afastamento indicado pelo cálculo do z-score relativo às unidades VVs ao longo dos trechos de fala é expresso em unidades de desvio-padrão. Deste modo, o z-score especifica o afastamento do valor medido na duração de cada unidade VV em relação à média da soma da duração dos fones de português brasileiro (PB) que compõem a unidade. Em outras palavras, o z-score assinala o alongamento ou a compressão da unidade VV, independentemente da duração intrínseca dos segmentos que a compõem.

Optou-se por calcular o z-score suavizado, porque a suavização é o procedimento padrão usado para atenuar os efeitos da realização local do acento lexical, assinalando assim apenas unidades potencialmente salientes no que diz respeito à prosódia do trecho de fala (Barbosa, 2006). O z-score suavizado para a duração das unidades VVs é calculado em duas etapas.

Na primeira, calcula-se a duração normalizada das unidades VVs, que é o z-score estatístico:

$$z - score = \frac{dur - \sum_i \mu_i}{\sqrt{\sum_i var_i}}$$

Na fórmula, a variável *dur* é a duração bruta da unidade VV em milissegundos, as variáveis μ_i e var_i são respectivamente a média e a variância de cada fone que compõe a unidade. A média e a variância dos fones transcritos na 1ª camada de anotação do Praat são retiradas de uma tabela construída a partir de um *corpus* da língua em análise. Neste trabalho, foram usados os valores de média e variância calculados em Barbosa (2006).

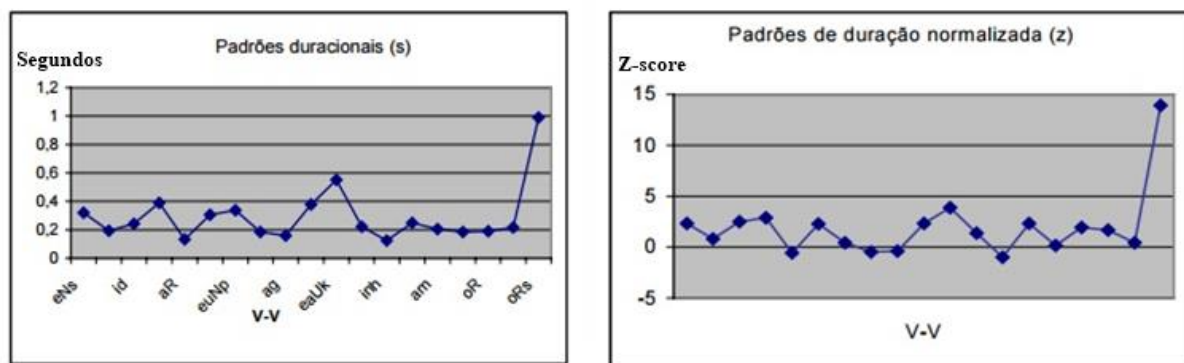


Figura 15: Duração bruta das unidades VVs e duração normalizada das unidades VVs. Fonte: Barbosa, 2004, p. 10-11.

Trecho: Em seguida apareceu um papagaio real que tinha fama de orador.

A segunda etapa para obtenção do z-score suavizado de duração da unidade VV é feita por meio de um procedimento padrão de média ponderada. Nesta etapa, as curvas das durações normalizadas são suavizadas por meio da fórmula abaixo:

$$z_{suav}^l = \frac{5 \cdot z^i + 3 \cdot z^{i-1} + 3 \cdot z^{i+1} + 1 \cdot z^{l-2} + 1 \cdot z^{l+2}}{13}$$

Na fórmula, z^i é o valor de z-score da duração da unidade VV, z^{i-1} é o valor de z-score da duração da unidade VV imediatamente anterior, z^{i+1} é o valor de z-score da duração da unidade VV imediatamente seguinte e assim sucessivamente.

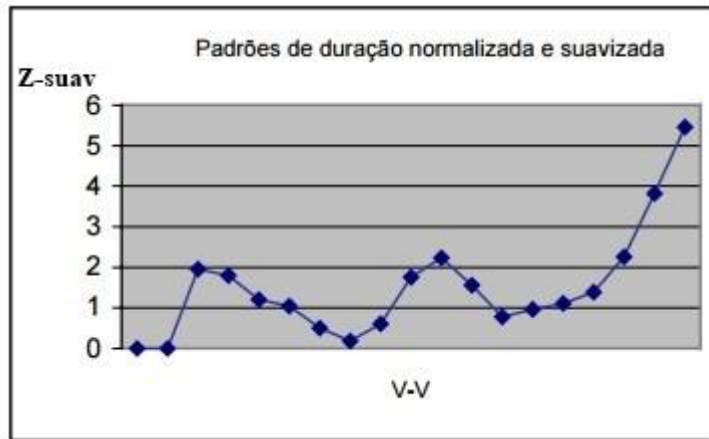


Figura 16: Duração normalizada e suavizada das unidades VVs.

Fonte: Barbosa, 2004, p. 11.

Trecho: Em seguida apareceu um papagaio real que tinha fama de orador.

Para as medidas de z-score suavizado do segmento, foram calculadas medidas em cada unidade VV em torno da posição de análise. Os parâmetros denominados z11, z12, z13, z14, z15, z16, z17, z18, z19, z110, z10, zr1, zr2, zr3, zr4, zr5, zr6, zr7, zr8, zr9, zr10 indicam a duração normalizada média em cada uma das 21 unidades locais estabelecidas pelo *script*, conforme esquema da figura (17) mostra.

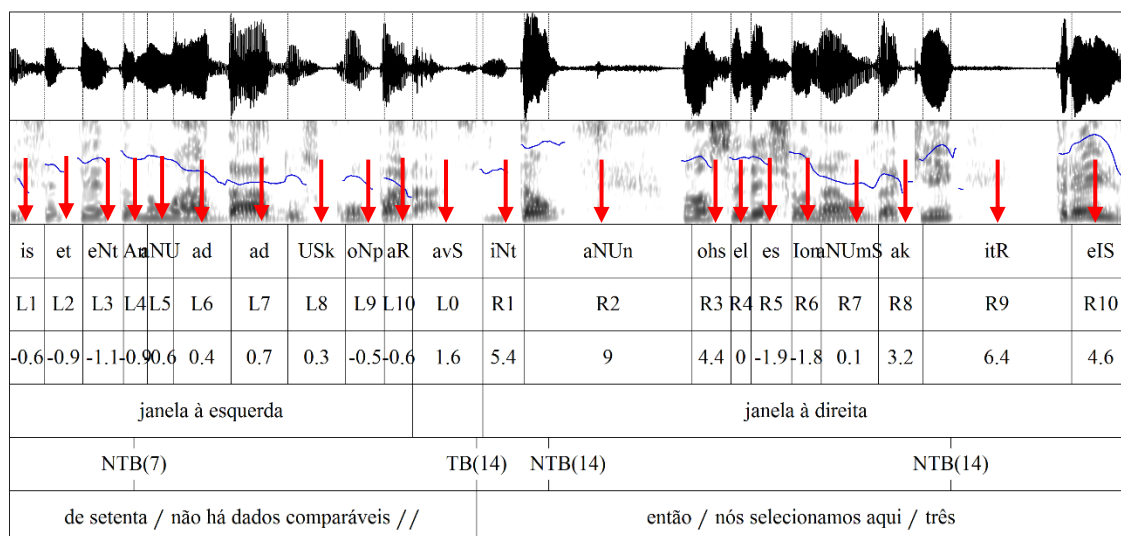


Figura 17: Medidas locais de duração normalizada para a fronteira terminal destacada no trecho (1).

Trecho (1): antes de setenta / não há dados comparáveis // então / nós selecionamos aqui / três

Na figura (17), destacam-se os valores de z-score suavizado de duração das unidades VVs em cada unidade, dada uma fronteira terminal percebida por 14 segmentadores.

Outra medida local de duração normalizada do segmento extraída automaticamente é a diferença entre a média de z-score normalizado e suavizado das unidades R1 e L0 em relação à posição de análise (zdloc). Em outras palavras, a medida zdloc foi obtida por meio da subtração simples entre o z-score de duração da unidade VV imediatamente após a posição de análise e o z-score de duração da unidade VV analisada, ou seja, entre zr1 e zl0. Para a fronteira terminal destacada no centro da figura (17), zdloc tem como resultado 3,8. A medida zdloc é importante para assinalar discrepância de z-score entre a unidade VV imediatamente anterior à posição de análise e a unidade imediatamente após.

Os descritores estatísticos calculados pelo *BreakDescriptor* para o z-score suavizado ao longo das janelas esquerda e direita compreendem medidas de média, assimetria, desvio-padrão e valores máximos de z-score suavizado de duração das unidades VVs (picos). Foram calculados estes descritores estatísticos, porque eles resumem informações disponíveis sobre o comportamento de uma variável qualquer. Na literatura da área, já é reconhecida a

importância destas medidas para a fala. De acordo com Barbosa e Silva (2012), os valores de média, desvio-padrão e assimetria dos picos de z-score suavizado de duração das unidades VVs revelam como se estruturam as proeminências e as fronteiras das unidades entonacionais.

A média, o desvio-padrão e a assimetria do z-score suavizado de duração das unidades VVs são calculados de três formas distintas pelo *BreakDescriptor*, independentemente da unidade VV ser usada para análise de correlato de fronteira ou não fronteira:

1. À esquerda da unidade VV de análise;
2. À direita da unidade VV de análise;
3. Diferença entre valor à direita e à esquerda da unidade VV de análise.

Em relação às medidas de diferença, elas foram obtidas a partir de uma subtração simples entre a média, o desvio-padrão e a assimetria de z-scores à direita e à esquerda da unidade VV de análise.

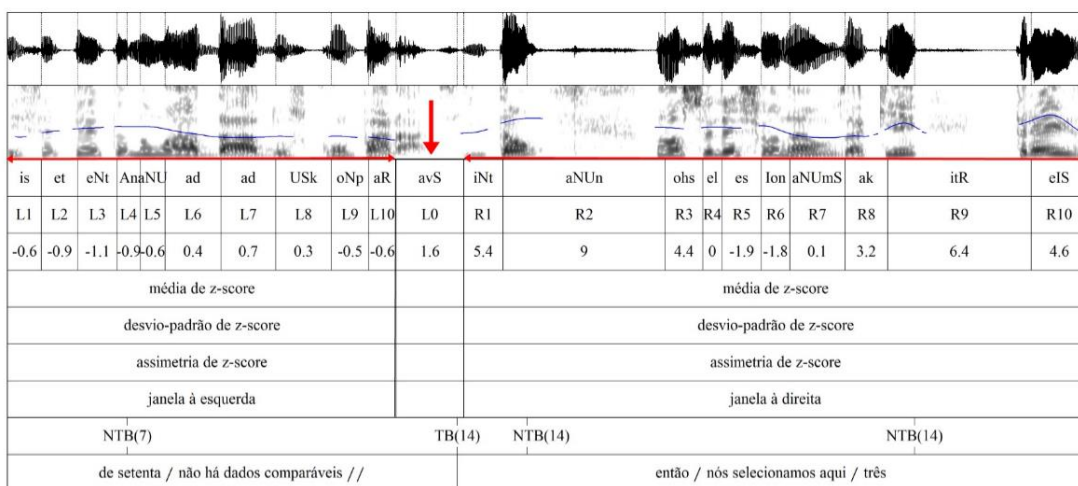


Figura 18: Medidas globais de duração normalizada para a fronteira terminal destacada no trecho (1) – mzl, m zr, SDzl, SDzr, skzl, skzr.

Trecho (1): antes de setenta / não há dados comparáveis // então / nós selecionamos aqui / três

O *BreakDescriptor* calcula a média do z-score suavizado de duração das unidades VVs localizadas na janela à esquerda e na janela à direita da unidade VV em análise. Essas medidas de média são parâmetros globais, que foram obtidos por meio do cálculo da média aritmética, utilizando os valores de z-score calculados nas unidades à esquerda e à direita (L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L0, R1, R2, R3, R4, R5, R6, R7, R8, R9 e R10) da posição em análise. A média global à esquerda da unidade VV de análise é chamada de *mzl*, já a média global à direita da unidade VV de análise é chamada de *mzr*. As medidas de média de z-score suavizado de duração das unidades VVs têm como objetivo geral verificar se nas janelas à esquerda e à direita da posição de análise existe um alongamento ou encurtamento geral das unidades VVs.

Para a fronteira terminal destacada no trecho (1) correspondente à fronteira também destacada na figura (18), as figuras (19) e (20) ilustram as medidas *mzl* e *mzr* calculadas. Estas figuras mostram a distribuição dos valores de z-score suavizado de duração das unidades VVs da unidade VV “avS” e das janelas à esquerda e à direita da unidade VV de análise. No eixo das abscissas, são apresentados os valores de z-score suavizado. No eixo das ordenadas, é apresentada a distribuição de frequências da variável z-score suavizado, ou seja, este eixo é proporcional a quantas vezes um determinado valor de z-score ocorreu. Por isso, conforme vemos abaixo, é possível representar também a probabilidade de ocorrência dos valores de z-score suavizado nas janelas à esquerda e à direita da unidade VV de análise.

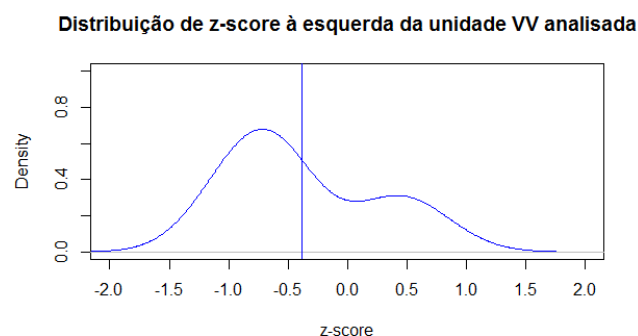


Figura 19: Medidas de duração normalizada para a fronteira terminal destacada no trecho (1) - *mzl*

Trecho (1): antes de setenta / não há dados comparáveis // então / nós selecionamos aqui / três

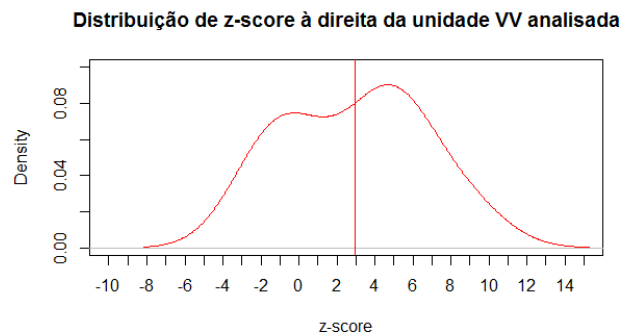


Figura 20: Medidas de duração normalizada para a fronteira terminal destacada no trecho (1) - m_{zr}.

Trecho (1): antes de setenta / não há dados comparáveis // então / nós selecionamos aqui / três

Nas figuras (19) e (20), as linhas verticais indicam respectivamente a média de z-score de duração das unidades VVs que compõem as janelas à esquerda e à direita da unidade VV de análise (“avS”). O valor da medida m_{zl} é -0,38, o valor de m_{zr} é 2,94.

Outra medida global calculada pelo *BreakDescriptor* para a duração normalizada e suavizada é a diferença entre a média de z-score suavizado de duração das unidades VVs à direita da VV de análise e a média de z-score suavizado de duração das unidades VV à esquerda da unidade VV de análise. Portanto, a diferença é obtida por meio da subtração dos valores médios m_{zr} e m_{zl}. Essa diferença é denominada m_{zd}. Para a fronteira terminal destacada no centro da figura (18), m_{zd} é 3,32.

O desvio-padrão do z-score suavizado de duração das unidades VVs mostra o grau de dispersão do z-score em relação à média. As medidas de desvio-padrão são parâmetros globais, que foram obtidos por meio do cálculo do desvio-padrão, utilizando os valores de z-score calculados em cada unidade VV. Estas medidas assinalam o quanto o ritmo das unidades VVs nas janelas globais varia, dado um ponto de fronteira ou não fronteira.

Os desvios-padrão do z-score suavizado de duração das unidades VVs à esquerda e à direita da unidade VV de análise são chamados respectivamente de SD_{zl} e SD_{zr}. Também é calculado outro parâmetro global de desvio-padrão para a duração normalizada, denominado SD_{zd}. O parâmetro SD_{zd} é obtido por meio da diferença entre SD_{zr} e SD_{zl}. A diferença foi obtida por meio da subtração entre os desvios-padrão de z-scores à direita e à esquerda da unidade de análise.

A assimetria é um tipo de descritor estatístico que distingue uma distribuição simétrica (aquela cuja média é igual à mediana e à moda) de uma assimétrica. Se assimétrica, isto é, assimetria diferente de zero, a distribuição da variável pode ter assimetria negativa (quando a cauda do lado esquerdo da distribuição for mais longa do que a do lado direito) ou assimetria positiva (quando a cauda do lado direito for mais longa do que a do lado esquerdo). Um valor nulo indica que os valores são distribuídos de maneira relativamente igual em ambos os lados da cauda, entretanto, não implica necessariamente em uma distribuição simétrica. Desse modo, a assimetria indica se, no caso de uma distribuição assimétrica, o desvio em relação à média é positivo ou negativo.

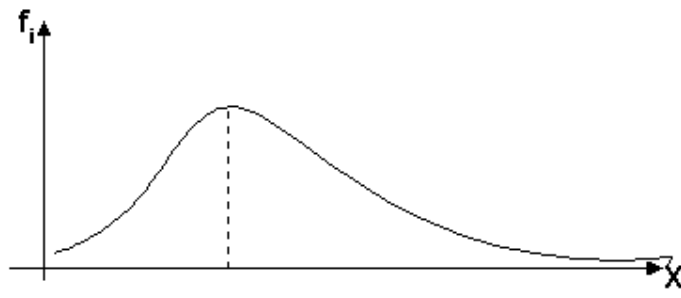


Figura 21: Distribuição assimétrica positiva

Fonte: Portal Action, 2018.

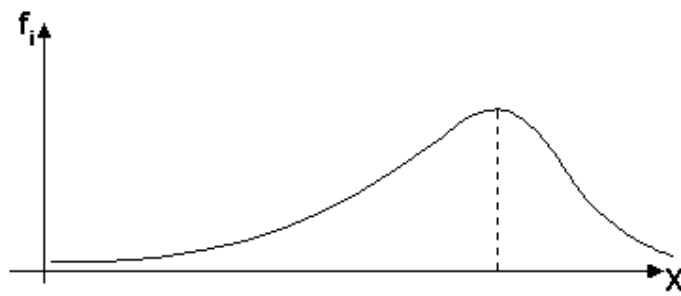


Figura 22: Distribuição assimétrica negativa

Fonte: Portal Action, 2018.

A assimetria do z-score suavizado de duração das unidades VVs mostra o quão longe a distribuição do z-score nas janelas à esquerda e à direita está de ser simétrica. As medidas de assimetria são parâmetros globais, que foram obtidos por meio do cálculo da assimetria, utilizando os valores de z-score calculados para cada janela individual de trabalho do *script*. A assimetria à direita tem um valor superior a 0, a assimetria à esquerda tem um valor inferior a 0.

A distribuição de frequência de uma variável que apresenta assimetria positiva tem uma cauda mais longa à direita da ordenada máxima, já a distribuição de frequência de uma variável que apresenta assimetria negativa tem uma cauda mais longa à esquerda da ordenada máxima. Normalmente, quando predominam os valores mais altos da variável, a assimetria é positiva, quando predominam os valores mais baixos da variável, a assimetria é negativa.

As medidas de assimetria de z-score suavizado de duração das unidades VVs têm como foco indicar se nas duas janelas globais à esquerda e à direita da unidade VV de análise predominam alongamentos ou encurtamentos em relação à duração intrínseca dos fones que compõem a janela em questão. As assimetrias do z-score suavizado de duração das unidades VVs à esquerda e à direita da unidade VV de análise são chamadas respectivamente de $skzl$ e $skzr$. A seguir, as figuras (23) e (24) mostram as distribuições de z-score suavizado extraídas para a fronteira terminal destacada no ponto central da figura (18).

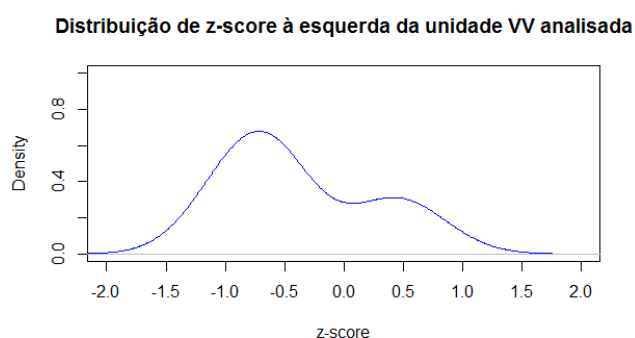


Figura 23: Medidas globais de duração normalizada do segmento para a fronteira terminal destacada no trecho (1) – $skzl$

Trecho (1): antes de setenta / não há dados comparáveis // então / nós selecionamos aqui / três

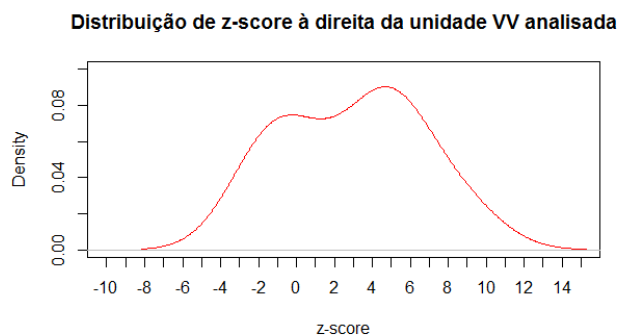


Figura 24: Medidas globais de duração normalizada do segmento para a fronteira terminal destacada no trecho (1) – skzr

Trecho (1): antes de setenta / não há dados comparáveis // então / nós selecionamos aqui / três

Para a fronteira terminal destacada no trecho (1), as medidas skzl e skzr são respectivamente 0,67 e 0,04. Observa-se que a janela global à esquerda da unidade VV de análise apresenta uma cauda mais longa à direita da ordenada máxima, porque predominam valores de z-score mais altos. Observa-se que a distribuição de z-score da janela à direita da unidade VV de análise é quase simétrica. Por isso, o valor de assimetria calculado é tão próximo de 0.

Também é calculado um parâmetro global de diferença de assimetria de z-score de duração das 10 unidades VVs localizadas antes e após a unidade VV analisada. Este parâmetro, denominado skzd, é obtido por meio da diferença de entre a assimetria de z-score de duração das unidades VVs à direita e a assimetria de z-score de duração das unidades VVs à esquerda da unidade VV de análise. Como nos outros descritores estatísticos, a medida de diferença (skzd) é calculada por meio de uma subtração simples. No caso de assimetria, a diferença é obtida por meio da subtração entre skzr e skzl. Para a fronteira terminal destacada no trecho (1), skzd é -0,63.

A taxa de saliência duracional é um parâmetro fonético-acústico global que indica a taxa de produção de picos de z-score suavizado de duração das unidades VVs por segundo. Foram calculadas três medidas globais de taxa de saliência. A primeira, denominada picos de z-score suavizado de duração das unidades VVs à esquerda da unidade VV de análise (prl), indica a taxa de produção de picos de z-score por segundo na janela à esquerda da unidade VV de análise; a segunda, denominada picos de z-score suavizado de duração das unidades VVs à direita da unidade VV analisada (prr), indica a taxa de produção de picos de z-score por segundo

na janela à direita da unidade VV de análise. As medidas prr e prl foram obtidas por meio da divisão do número de picos z-score produzidos nas janelas à direita e à esquerda da unidade VV de análise pela duração total de cada uma das janelas.

A terceira medida é denominada diferença de taxa de produção de picos de z-score por segundo (prd). Ela foi obtida por meio da diferença entre prr e prl, ou seja, prd foi obtida por meio da diferença entre as taxas de saliência à direita e à esquerda da unidade VV de análise. A medida prl é importante para assinalar o aumento ou declínio da taxa de produção de picos de z-score, dada uma unidade VV de análise. Deste modo, com a medida prd verifica-se o impacto da mudança da taxa de saliência duracional na percepção de fronteira ou de não fronteira.

III. Medidas de frequência fundamental (f0)

Como já foi dito anteriormente, a frequência fundamental (f0) é um parâmetro fonético-acústico que expressa o número de ciclos completos de vibração das cordas vocais por segundo e sua unidade de medida padrão é o Hertz (Hz). O *pitch* (tom) é o correlato perceptual deste parâmetro. Outra escala de medida utilizada para a frequência fundamental é o semitom, que é expresso em escala logarítmica e se aproxima mais da percepção humana de f0.

Quanto maior for a frequência fundamental, maior será a altura do som e vice-versa. A altura (*pitch*) do som é o parâmetro fonético-acústico que diferencia um som grave de um som agudo. Os sons de menor altura ou menores valores de f0 são mais graves, como por exemplo, as vozes masculinas. Já os sons com maiores alturas, isto é, maiores valores de f0 são mais agudos como por exemplo, as vozes de mulheres e crianças.

As medidas de f0 calculadas pelo *BreakDescriptor* foram calculadas na escala semitons re 1Hz⁶. Os descritores estatísticos calculados para a frequência fundamental compreendem medidas de mediana, média, assimetria e desvio-padrão. Também foram

⁶ Usa-se a expressão “semitons re 1 Hz” para indicar que os valores em semitons são calculados relativamente à frequência 1 Hz.

calculadas medidas locais em cada unidade VV, a taxa de picos de f0 produzidos por segundo e medidas de primeira derivada da mediana de f0.

As medidas locais de f0 calculadas em cada unidade VV pelo *BreakDescriptor* compreendem exclusivamente medidas de mediana. Foram extraídos os valores de mediana da frequência fundamental, porque os valores de mediana são mais resistentes a erros de medição do que a média. A figura (25) mostra as medianas de f0 calculadas em cada uma das dez unidades VVs à esquerda e à direita, dada uma fronteira terminal percebida por 11 segmentadores.

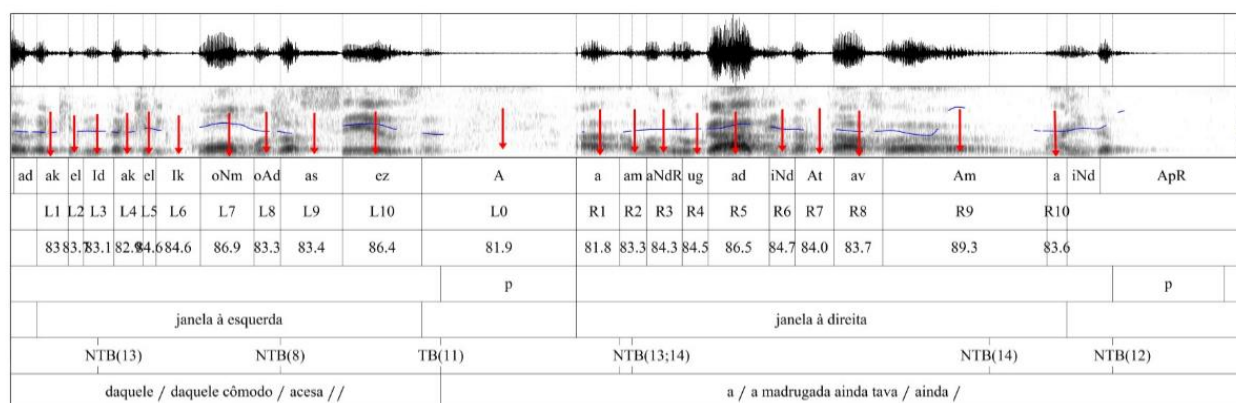


Figura 25: Medidas locais de f0 para a fronteira terminal destacada no trecho (2)

Trecho (2): daquele / daquele cômodo / acesa // a / a madrugada ainda tava / ainda /

De acordo com o critério de nomenclatura estabelecido, as medianas locais de f0 são denominadas f0medl1, f0medl2, f0medl3, f0medl4, f0medl5, f0medl6, f0medl7, f0medl8, f0medl9, f0medl10, f0medr1, f0medr2, f0medr3, f0medr4, f0medr5, f0medr6, f0medr7, f0medr8, f0medr9, f0medr10. No total, foram calculadas 22 medidas locais de mediana de f0, porque além das 21 medidas locais de mediana de f0 extraídas em cada unidade local de trabalho do *script*, também foi calculada uma medida local de mediana de f0 denominada f0meddloc. A medida f0meddloc é obtida por meio da subtração entre a mediana de f0 da unidade R1 e a mediana de f0 da unidade L0, isto é, entre f0medr1 e f0medl0. F0meddloc assinala o valor do *reset* de f0.

Os parâmetros globais calculados para a frequência fundamental compreendem medidas de média, assimetria e desvio-padrão. As medidas de média, assimetria e desvio-padrão de f_0 foram calculadas utilizando as medianas de f_0 calculadas em cada unidade VV (L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L0, R1, R2, R3, R4, R5, R6, R7, R8, R9 e R10) para extrair estes descritores estatísticos.

A média, o desvio-padrão e a assimetria das medianas de f_0 são calculadas de três formas distintas pelo *BreakDescriptor*, independentemente da unidade VV ser usada para análise de correlato de fronteira ou não fronteira:

1. À esquerda da unidade VV de análise;
2. À direita da unidade VV de análise;
3. Diferença entre valor à direita e à esquerda da unidade VV de análise.

Abaixo, a figura (26) ilustra a janela usada para calcular seis medidas de globais f_0 , que são: a média das medianas de f_0 à esquerda e à direita da unidade VV de análise, a assimetria das medianas de f_0 à esquerda e à direita da unidade VV de análise e o desvio-padrão das medianas de f_0 à esquerda e à direita da unidade VV de análise.

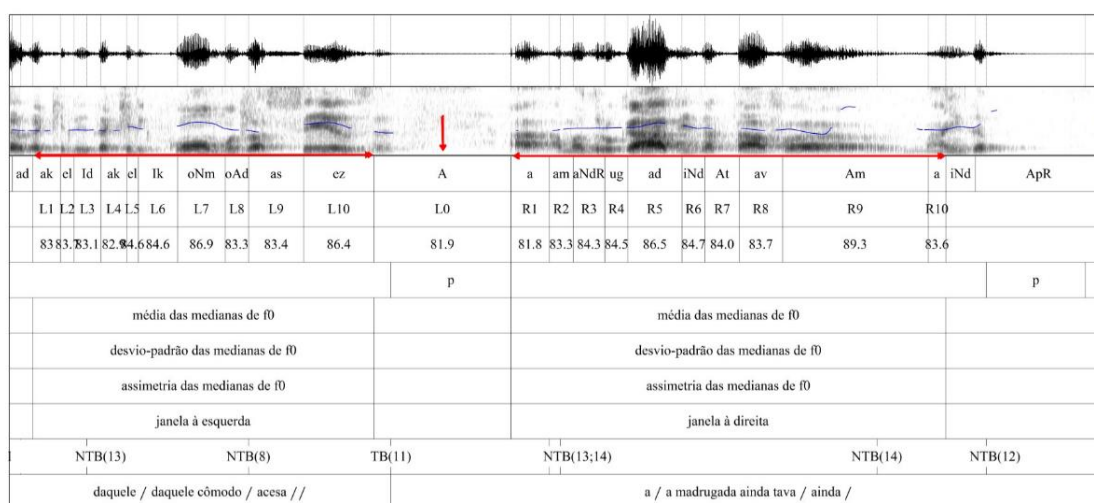


Figura 26: Medidas globais de f0 para a fronteira terminal destacada no trecho (2) - f0medl, sdf0l, skf0l, f0medr, sdf0r e skf0r.

Trecho (2): daquele / daquele cômodo / acesa // a / a madrugada ainda tava / ainda /

A média, o desvio-padrão e a assimetria das medianas de f0 à esquerda da unidade VV são chamados respectivamente de f0medl, sdf0l e skf0l. A média, o desvio-padrão e a assimetria das medianas de f0 à direita da unidade VV são chamados respectivamente de f0medr, sdf0r e skf0r.

Para a unidade VV destacada na figura (25), as figuras (27) e (28) ilustram por meio de linhas verticais as medidas de média das medianas de f0 calculadas.

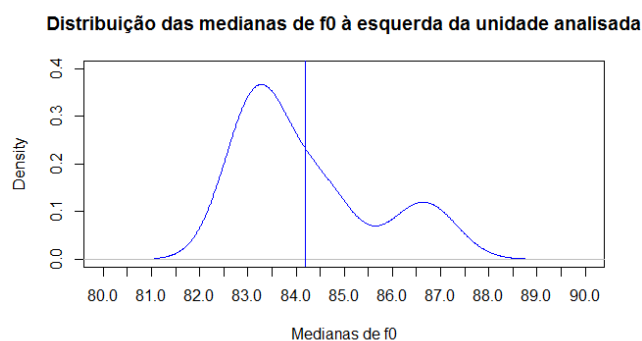


Figura 27: Medidas globais de f0 para a fronteira terminal destacada no trecho (2) - f0medl.

Trecho (2): daquele / daquele cômodo / acesa // a / a madrugada ainda tava / ainda /

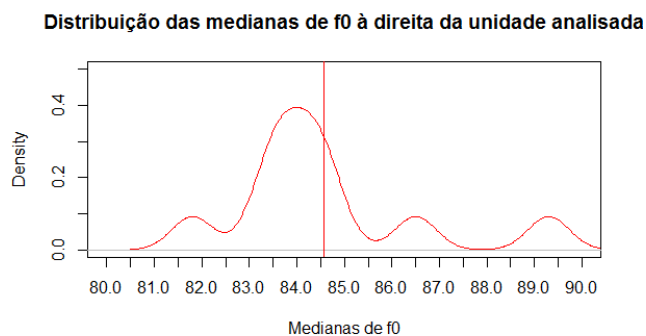


Figura 28: Medidas globais de f0 para a fronteira terminal destacada no trecho (2) – f0medr.

Trecho (2): daquele / daquele cômodo / acesa // a / a madrugada ainda tava / ainda /

As duas medidas de média das medianas de f0 à esquerda e à direita da unidade VV de análise têm como objetivo geral verificar o f0 central nas duas janelas globais. Observa-se que f0medl é menor do que f0medr, pois, f0medl é em torno de 84,2 st⁷ e f0medr é em torno de 84,6 st. A terceira medida calculada para a média das medianas de f0 é a diferença entre a média à direita e à esquerda da unidade VV de análise, ou seja, entre f0medr e f0medl. No exemplo dado, a subtração tem como resultado 0,4 st.

O desvio-padrão das medianas de f0 indica a variação de f0 nas duas janelas globais estabelecidas para análise em torno da unidade VV de análise. Estas medidas são importantes para assinalar o quanto o valor de f0 nas duas janelas globais estabelecidas pelo *BreakDescriptor* varia, dado um ponto de fronteira ou não fronteira. Para o trecho de fala ilustrado na figura (25), os desvios-padrão à esquerda (sdf0l) e à direita (sdf0r) da unidade VV analisada são respectivamente 1,43 e 2,04 st. A terceira medida calculada para o desvio-padrão das medianas de f0 é a diferença entre a mediana à direita e à esquerda da unidade VV de análise, ou seja, entre sdf0r e sdf0l. No exemplo dado, a diferença tem como resultado 0,61 st.

Com as medidas de assimetria das medianas de f0 é indicado se nas duas janelas globais à esquerda e à direita da unidade VV de análise predominam valores de f0 mais graves ou mais agudos. Valores mais altos de assimetria das medianas de f0 indicam que o f0 predominante é mais agudo, valores mais baixos de assimetria das medianas indicam que o f0 predominante é mais grave. A seguir, as figuras (29) e (30) mostram em sequência a assimetria

⁷ Usa-se st para indicar a escala de medida para a f0 em semitons.

das medianas de f_0 à esquerda e à direita (skf_{0l} e skf_{0r}) da fronteira terminal destacada no ponto central da figura (25).

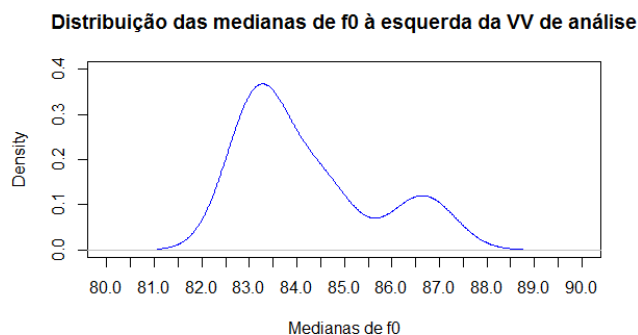


Figura 29: Medidas globais de f_0 para a fronteira terminal destacada no trecho (2) – skf_{0l}

Trecho (2): daquele / daquele cômodo / acesa // a / a madrugada ainda tava / ainda /

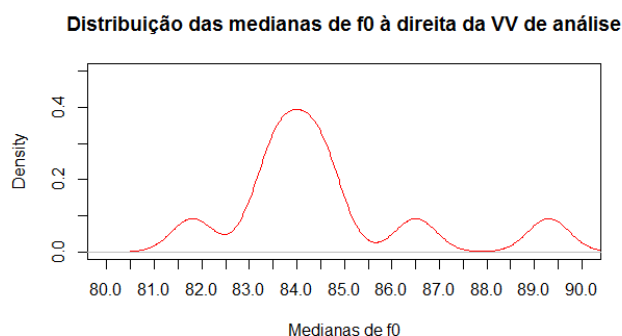


Figura 30: Medidas globais de f_0 para a fronteira terminal destacada no trecho (2) – skf_{0r} .

Trecho (2): daquele / daquele cômodo / acesa // a / a madrugada ainda tava / ainda /

Para a fronteira terminal destacada no trecho (2), as medidas skf_{0l} e skf_{0r} são respectivamente 0,97 e 1,21. Observa-se que as janelas globais à esquerda e à direita da unidade VV de análise apresentam uma cauda mais longa à direita da ordenada máxima, porque predominam valores de mediana de f_0 mais altos. Em outras palavras, para a fronteira terminal

destacada no trecho (2), nas janelas à esquerda e à direita da fronteira, predominam valores de f_0 mais agudos, pois os valores de assimetria skf_{0l} e skf_{0r} são superiores a 0. Entretanto, o valor de assimetria skf_{0r} é um pouco maior, se comparado ao valor de assimetria skf_{0l} . Deste modo, nas duas janelas predominam valores de f_0 mais agudos. Na janela à direita, entretanto, os valores de f_0 são mais agudos do que os valores de f_0 na janela à esquerda.

O *BreakDescriptor* também calcula outra medida para a assimetria das medianas de f_0 . Esta medida é denominada diferença de assimetria das medianas de f_0 entre janela à direita e à esquerda da unidade VV de análise (skf_{0d}). A medida skf_{0d} é calculada por meio da subtração entre skf_{0r} e skf_{0l} . No caso da fronteira terminal em discussão, skf_{0d} é 0,24

Em Cálculo Diferencial e Integral, a derivada de uma função representa a taxa de variação instantânea da função em um ponto qualquer. Em outras palavras, a derivada indica o crescimento ou decrescimento da função no ponto de análise. Além das medidas de mediana de f_0 , o *BreakDescriptor* calculou automaticamente uma série de medidas de derivadas discretas para a mediana da frequência fundamental. A definição de derivada discreta adotada expressa a taxa de variação da mediana de f_0 em pontos específicos. Os pontos em questão são as unidades VVs presentes nas duas janelas de trabalho estabelecidas pelo *script*.

As derivadas foram obtidas por meio da subtração entre o valor atual e o valor anterior de mediana de f_0 . Deste modo, para calcular a primeira derivada da mediana de f_0 na unidade L0, por exemplo, subtrai-se da mediana de f_0 da unidade L0 a mediana de f_0 da unidade L10. De forma semelhante, para calcular a primeira derivada discreta da mediana de f_0 na unidade R1, subtrai-se da mediana de f_0 da unidade R1 a mediana de f_0 da unidade L0 e assim sucessivamente. Valores positivos de derivadas indicam o crescimento de f_0 , valores negativos indicam o decrescimento.

Abaixo, a figura (31) ilustra o cálculo realizado para obtenção das derivadas discretas da mediana de f_0 .

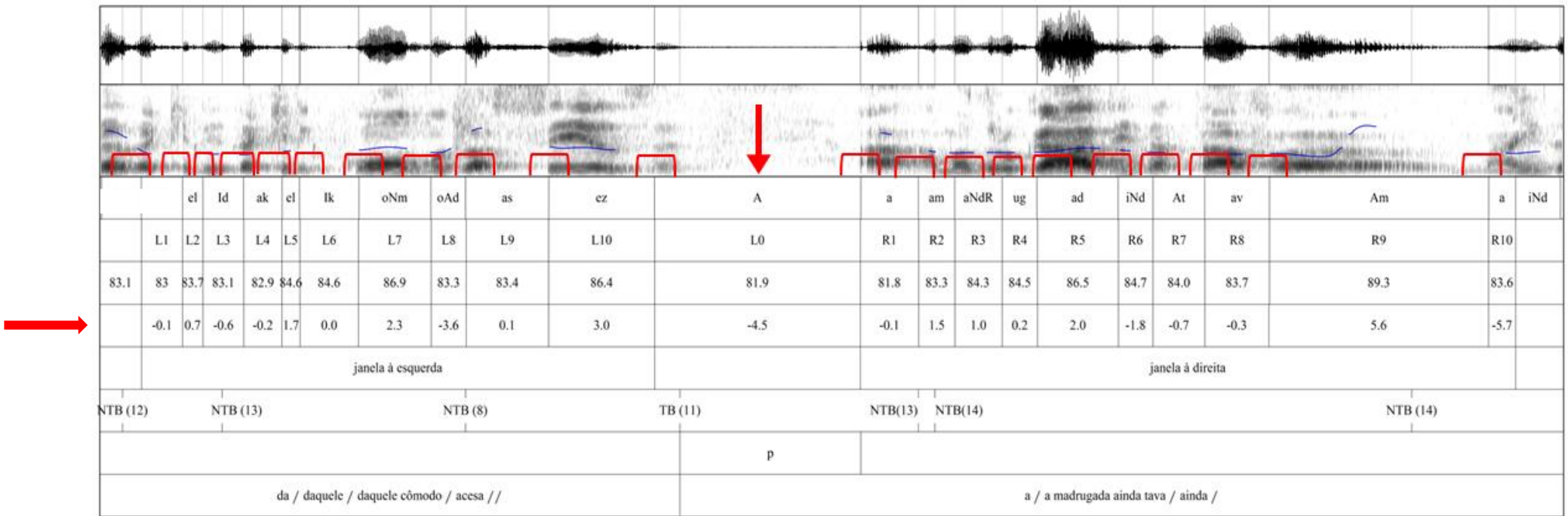


Figura 31: Medidas locais de f0 para a fronteira terminal destacada no trecho (2) – Derivadas discretas

Trecho (2): daquele / daquele cômodo / acesa // a / a madrugada ainda tava / ainda /

Na figura (31), destacam-se, na terceira camada de anotação do Praat, as medianas de f_0 calculadas em cada unidade VV individual de trabalho do *script*, na quarta camada, são destacados os valores das primeiras derivadas discretas da mediana de f_0 calculadas em cada unidade local considerada pelo *script*, dado um ponto de fronteira terminal percebida por 11 segmentadores humanos.

Para análise de uma unidade VV qualquer, adotou-se como padrão de anotação das primeiras derivadas as siglas $df0medl1$, $df0medl2$, $df0medl3$, $df0medl4$, $df0medl5$, $df0medl6$, $df0medl7$, $df0medl8$, $df0medl9$, $df0medl10$, $df0med0$, $df0medr1$, $df0medr2$, $df0medr3$, $df0medr4$, $df0medr5$, $df0medr6$, $df0medr7$, $df0medr8$, $df0medr9$ e $df0medr10$ para as respectivas unidades VVs.

Além das 21 medidas de primeira derivada da mediana de f_0 extraída em cada unidade local de trabalho do *script*, também foi calculada outra medida local para a derivada da mediana de f_0 . Esta medida é denominada $df0meddloc$ e foi obtida por meio da subtração simples entre $df0medr1$ e $df0med0$. Todas as medidas locais de derivada da mediana de f_0 mostram de forma aproximada mudanças abruptas no contorno entonacional de f_0 , indicando simultaneamente a velocidade da mudança e a direção da mudança. Para a unidade VV destacada na figura (25), $df0meddloc$ tem como resultado 4,4.

Os descritores estatísticos calculados para as derivadas das medianas de f_0 são a média, o desvio-padrão e a assimetria. A média das derivadas das medianas de f_0 assinala a tendência de mudanças no contorno entonacional de f_0 nas duas janelas globais estabelecidas, dada uma unidade VV de análise localizada em fronteira de palavra fonológica. O desvio-padrão das derivadas das medianas de f_0 é importante para indicar a variação das mudanças de f_0 nas duas janelas globais estabelecidas para análise de uma unidade VV, isto é, assinalar a regularidade ou irregularidade das mudanças de f_0 , dado um ponto de fronteira ou não fronteira. A assimetria das derivadas das medianas de f_0 assinala o tipo de mudança de f_0 que é predominante nas duas janelas globais de trabalho do *script*, ou seja, assinala se o crescimento ou decrescimento no contorno entonacional de f_0 é predominante nas duas janelas globais.

A média, o desvio-padrão e a assimetria das derivadas discretas das medianas de f_0 também foram calculados de três formas distintas, usando os valores de derivadas extraídos:

1. À esquerda da unidade VV de análise;
2. À direita da unidade VV de análise;

3. Diferença entre valor à direita e à esquerda da unidade VV de análise.

Abaixo, a figura (32) ilustra o cálculo da média, do desvio-padrão e da assimetria da primeira derivada discreta da mediana de f_0 . Na figura (32), destaca-se uma fronteira terminal percebida por 11 segmentadores. Para calcular a média, o desvio-padrão e a assimetria das primeiras derivadas das medianas de f_0 , foram usados os valores de derivadas das medianas de f_0 anotados na quarta camada de anotação do Praat. A média, o desvio-padrão e a assimetria das primeiras derivadas das medianas de f_0 à esquerda da unidade VV de análise são chamados respectivamente de df_0medl , $sddf0l$ e $skdf0l$. A média, o desvio-padrão e a assimetria das primeiras derivadas das medianas de f_0 à direita da unidade VV de análise são chamados respectivamente de df_0medr , $sddf0r$ e $skdf0r$.

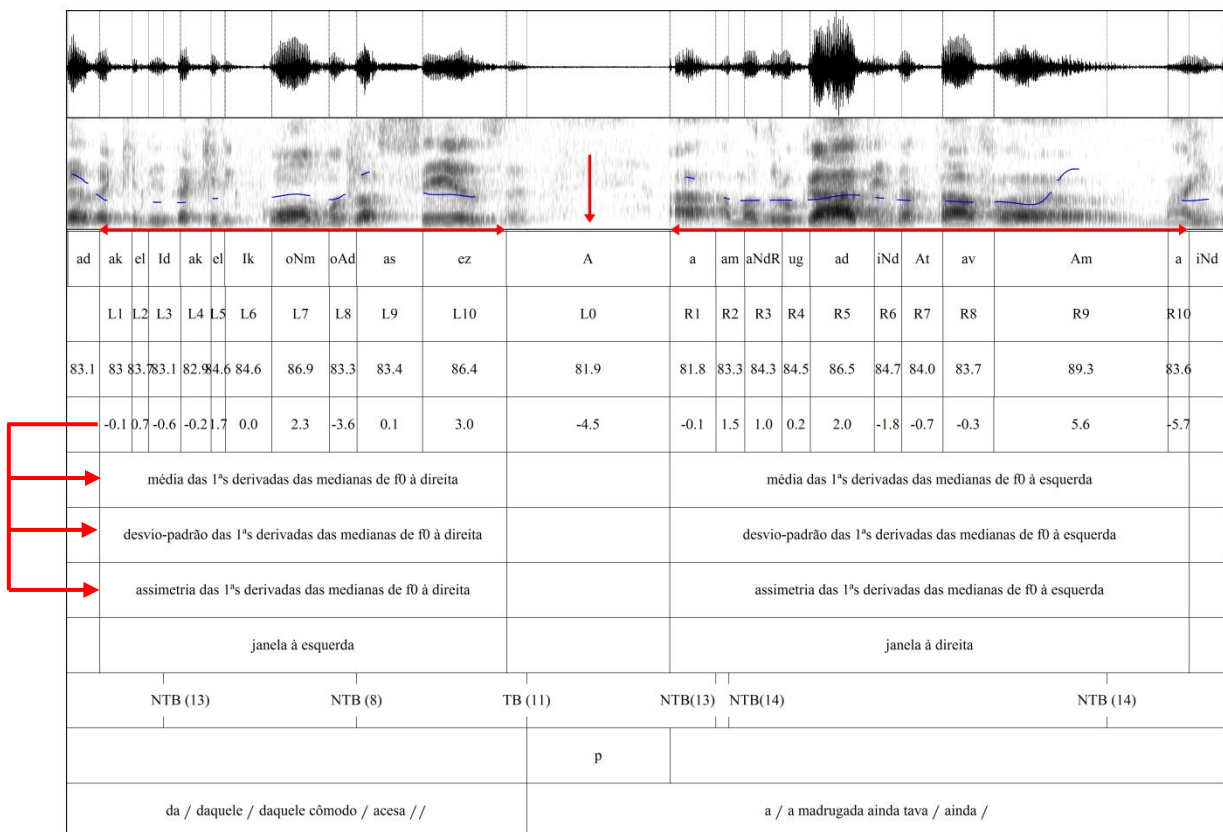


Figura 32: Medidas globais das primeiras derivadas das medianas de f_0 para a fronteira terminal destacada no trecho (2) – df_0medl , $sddf0l$, $skdf0l$, df_0medr , $sddf0r$ e $skdf0r$.

Trecho (2): da / aquele / aquele cômodo / acesa // a / a madrugada ainda tava / ainda /

Para visualizar melhor a média das derivadas das medianas de f_0 à esquerda e à direita da unidade VV de análise, abaixo é apresentado o gráfico de distribuição da variável derivada da mediana de f_0 . As figuras (33) e (34) ilustram por meio das linhas verticais as medidas df_0medl e df_0medr calculadas para a fronteira terminal destacada no trecho (2).

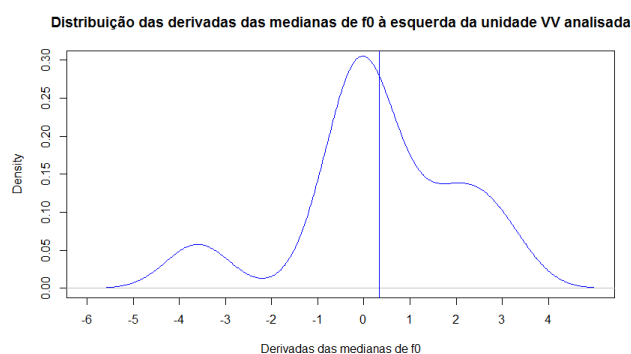


Figura 33: Medidas globais das primeiras derivadas das medianas de f_0 para a fronteira terminal destacada no trecho (2) - df_0medl

Trecho (2): daquele / daquele cômodo / acesa // a / a madrugada ainda tava / ainda /

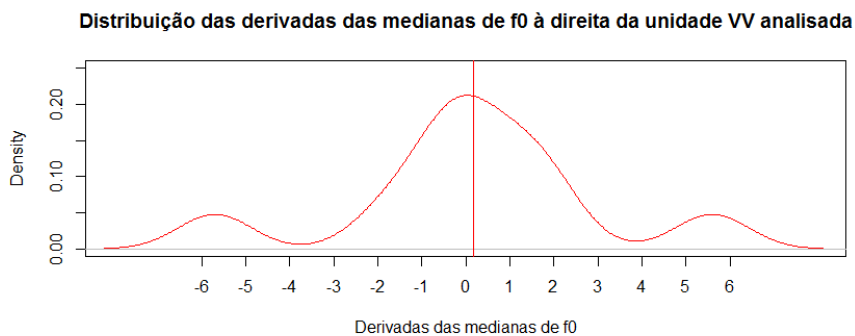


Figura 34: Medidas globais das primeiras derivadas das medianas de f_0 para a fronteira terminal destacada no trecho (2) - df_0medr

Trecho (2): daquele / daquele cômodo / acesa // a / a madrugada ainda tava / ainda /

Observa-se que $df0medl$ é um pouco maior do que $df0medr$, pois, $df0medl$ é em torno de 0,33 st e $df0medr$ é em torno de 0,17 st. Deste modo, para a fronteira terminal destacada no trecho (2), a tendência de mudanças no contorno entonacional de f_0 em ambas janelas globais estabelecidas é o aumento de f_0 . Contudo, na janela à esquerda da fronteira em análise a tendência de aumento de f_0 é maior do que a tendência de aumento de f_0 da janela à direita da fronteira terminal.

A terceira medida calculada para a média das derivadas das medianas de f_0 é a diferença entre $df0medr$ e $df0medl$. Esta diferença é denominada $df0medd$. No exemplo dado, $df0medd$ tem como resultado -0,16 st.

Para a fronteira terminal destacada no trecho de fala (2), os desvios-padrão das derivadas das medianas de f_0 à esquerda ($sddf0l$) e à direita ($sddf0r$) da unidade VV analisada são respectivamente 1,82 e 2,88 st. A terceira medida calculada para o desvio-padrão das derivadas de f_0 é a diferença de desvio-padrão das primeiras derivadas das medianas de f_0 entre a janela à direita e à esquerda da unidade VV de análise, isto é, entre $sddf0r$ e $sddf0l$. Esta diferença é chamada de $sddf0d$. No exemplo dado, $sddf0d$ tem como resultado 1,06 st.

As figuras (35) e (36) ilustram as medidas de assimetria das derivadas das medianas de f_0 para a fronteira terminal destacada no trecho (2).

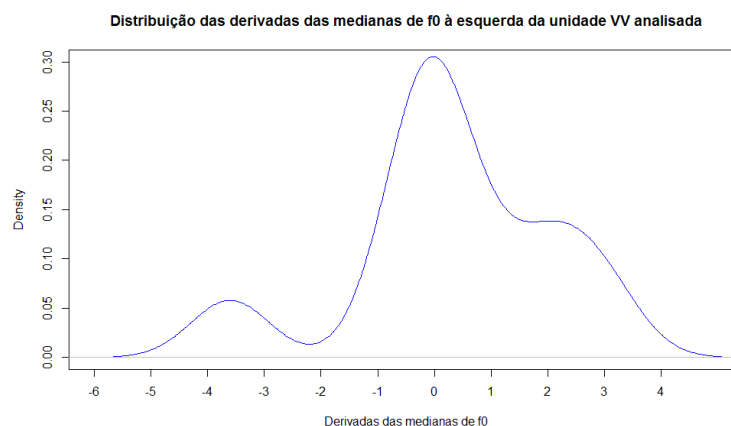


Figura 35: Medidas globais das primeiras derivadas das medianas de f_0 para a fronteira terminal destacada no trecho (2) – $skdf0l$.

Trecho (2): *daquele / daquele cômodo / acesa // a / a madrugada ainda tava / ainda /*

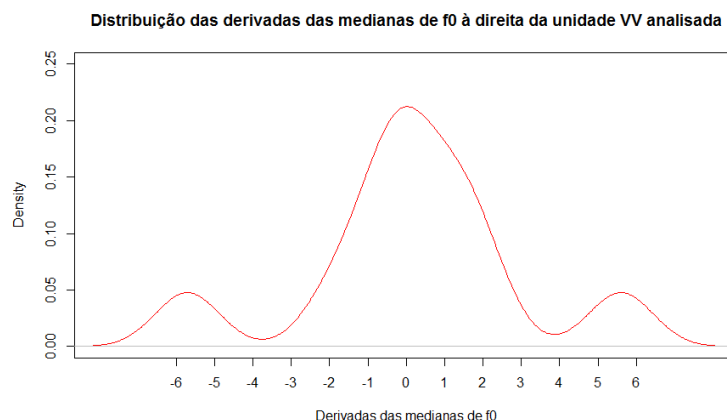


Figura 36: Medidas globais das primeiras derivadas das medianas de f_0 para a fronteira terminal destacada no trecho (2) – skdf0r.

Trecho (2): *daquele / daquele cômodo / acesa // a / a madrugada ainda tava / ainda /*

Para a fronteira terminal destacada no trecho (2), as medidas skdf0l e skdf0r são respectivamente -0,63 e -0,2. Observa-se que as janelas globais à esquerda e à direita da unidade VV de análise apresentam uma cauda mais longa à esquerda da ordenada máxima, porque predominam em ambas as janelas valores de derivadas das medianas de f_0 mais baixos. Por isso, os valores de assimetria skdf0l e skdf0r são menores que 0. Outra medida global calculada para as derivadas das medianas de f_0 é a diferença de assimetria das primeiras derivadas das medianas de f_0 entre a janela à direita e à esquerda da unidade VV de análise. Esta diferença é chamada de skdf0d e tem como resultado no exemplo dado 0,43.

Foram calculadas automaticamente também três medidas de taxa de saliência de f_0 . A taxa de saliência de f_0 é um parâmetro fonético-acústico global que indica a taxa de produção de picos de f_0 suavizados por segundo. Esta medida assinala a taxa de proeminências melódicas associadas ao *pitch accent* em um trecho de fala espontânea, sendo importante para assinalar a vivacidade (*liveliness*) da fala. Foram calculadas três medidas globais de taxa de proeminência de f_0 . A primeira, denominada de f0rl, calculou a taxa de produção de picos de f_0 na janela à esquerda da unidade VV de análise, a segunda, denominada f0rr, calculou a taxa de produção de picos de f_0 na janela à direita da unidade VV de análise. A terceira medida, denominada f0rd, foi obtida por meio da diferença entre f0rr e f0rl, ou seja, foi obtida por meio da subtração entre as taxas de proeminência de f_0 à direita e à esquerda da unidade VV de análise.

IV. Medidas de intensidade

A intensidade é um parâmetro fonético-acústico que expressa a quantidade de energia dispersa no sinal sonoro da fala em todo espectro. Geralmente, a intensidade é medida em dB e tem como correlato perceptual a sensação de volume (*loudness*). Observa-se entre os trabalhos da área que a intensidade não é muito citada na literatura que investiga parâmetros fonético-acústicos relevantes para o estabelecimento de fronteiras prosódicas. Uma das razões que justificam isso é o fato de que a intensidade é um tipo de medida muito afetada pela distância do falante em relação ao microfone de gravação da fala. Naturalmente, se a distância do microfone em relação ao aparelho fonador do falante for muito próxima, a intensidade medida será um valor alto, o que pode não ser necessariamente verdadeiro.

Uma forma de contornar esse problema é usar uma medida de intensidade relativa de bandas, denominada ênfase espectral. Esta medida é capaz de atenuar os efeitos da posição do microfone sob as medidas de intensidade e fornecer uma medida indireta do esforço vocal. Quanto maior o valor de ênfase espectral, maior o esforço vocal produzido pelo falante.

Traunmüller e Eriksson (2000) observaram que as vogais prosodicamente proeminentes são produzidas com um maior esforço vocal, o que faz com que as faixas de alta frequência do espectro apresentem maior energia. A ênfase espectral é uma medida de intensidade relativa de bandas que permite analisar a produção de vogais proeminentes por meio da presença de mais energia em faixas de frequências mais altas, como atestado anteriormente por Traunmüller e Eriksson (2000). A ênfase espectral é uma medida que indica a contribuição relativa das altas frequências para a intensidade total do espectrograma.

Traunmüller e Eriksson (2000) definem ênfase espectral como diferença de energia, em decibéis, entre a intensidade acústica total do sinal original e a intensidade do sinal após ser submetido a um filtro passa-baixas pouco acima da frequência fundamental. A fórmula usada para o cálculo da ênfase espectral é:

$$\hat{E}nfase\ espectral = L - L0$$

Em que, L é a intensidade total do espectro como um todo, L_0 é a intensidade de frequências baixas até o valor $1,43 \times f_0$ média. Uma proposta de simplificação apresentada pelos autores é adotar o filtro passa-baixas com frequência de corte (L_0) de aproximadamente 400 Hz. Abaixo, a figura (37) ilustra como calcular a ênfase espectral.

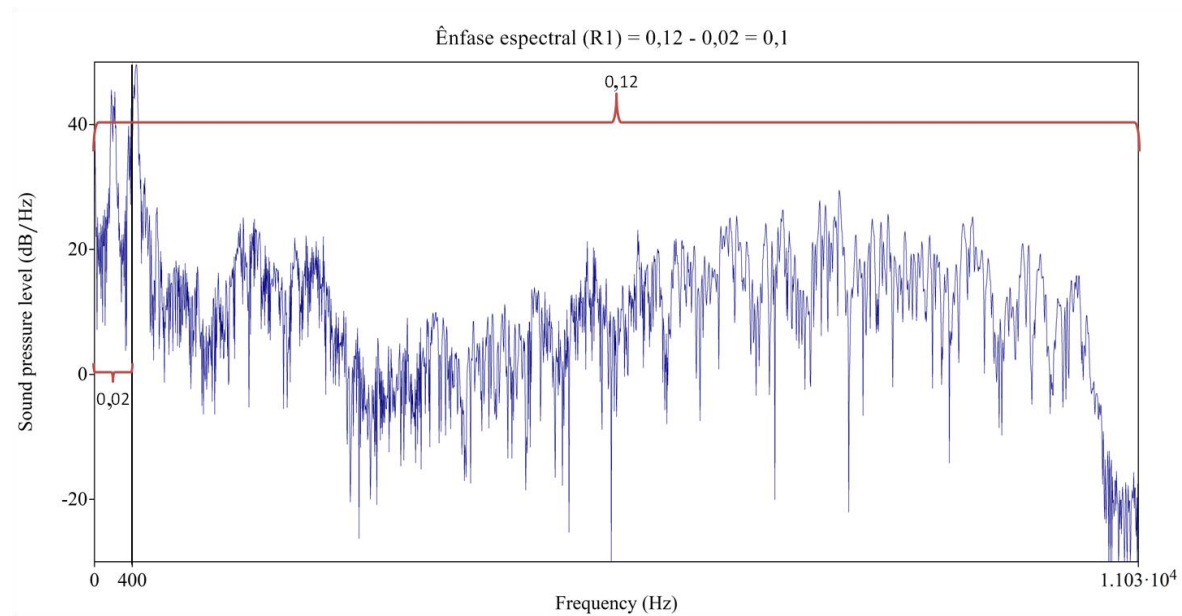


Figura 37: Medidas de intensidade - Ênfase espectral para a unidade R1 localizada após a fronteira terminal destacada no trecho (3)

Trecho (3): cada um deles / faz mais de dez mil sinapses // quer dizer / o que a / a forma como essas

A figura (37) mostra o nível de pressão sonora (dB) por frequência fundamental (Hz) ao longo da unidade VV R1 localizada imediatamente após a fronteira terminal destacada no trecho (3). Podemos observar na imagem que as frequências altas apresentam maiores valores de energia. Contudo, nem sempre as altas frequências apresentam de fato valores de intensidade mais altos. Por isso, é muito importante adotar medidas de ênfase espectral para que as medidas de intensidade não sejam afetadas pela intensidade das altas frequências. No exemplo dado na figura (37), as intensidades L e L_0 são respectivamente 0,12 e 0,02. Assim, a ênfase espectral desta unidade VV tem como resultado 0,1.

Neste trabalho, optou-se pelo cálculo da ênfase espectral considerando a energia total (L) até 11025 Hz. Fixou-se 11025 Hz como intensidade total, porque esta é a taxa de Nyquist dos textos do *corpus* C-ORAL BRASIL. O filtro passa-baixas adotado ($L0$) tem como frequência de corte 400Hz. Os descritores estatísticos calculados para a ênfase espectral compreendem apenas medidas de média.

Foram calculadas pelo *BreakDescriptor* quatro medidas de ênfase espectral, das quais três são globais e uma é local. Como nos outros casos de parâmetros fonético-acústicos globais, as medidas foram calculadas de três formas distintas:

1. À esquerda da unidade VV de análise;
2. À direita da unidade VV de análise;
3. Diferença entre valor à direita e à esquerda da unidade VV de análise.

As médias de ênfase espectral foram calculadas de forma global em cada uma das duas janelas estabelecidas. As siglas usadas para se referir às medidas globais de média de ênfase espectral à esquerda e à direita da unidade VV de análise são respectivamente *emphl* e *emphr*. Outra medida global calculada para a ênfase espectral é a diferença de média de ênfase espectral à direita e à esquerda da unidade VV de análise. A diferença, chamada de *emphd*, foi obtida por meio da subtração entre *emphr* e *emphl*.

A única medida local calculada para a média de ênfase espectral é denominada *emphdloc*. Ela é obtida por meio da subtração entre a média de ênfase espectral após a unidade VV de análise e a média de ênfase espectral medida na unidade VV de análise. Assim, a medida *emphdloc* é obtida por meio da subtração entre a média de ênfase espectral na unidade R1 e na unidade L0, dado um ponto de fronteira ou não fronteira analisado. O parâmetro *emphdloc* é importante para assinalar o aumento ou declínio da intensidade, dada uma unidade VV de análise. Deste modo, com a medida *emphdloc* verifica-se o impacto da mudança da intensidade na percepção de fronteira prosódica ou de não fronteira.

V. Medidas de pausa

Frequentemente, entre os trabalhos da área, distingue-se “pausa silenciosa” de “pausa preenchida” (Zellner, 1994; Soriano, 2006; Merlo e Barbosa, 2012). A pausa silenciosa é composta por uma interrupção do sinal acústico, sendo assim, um silêncio no fluxo da fala. As pausas preenchidas são interrupções, que são realizadas por meio de algum material vocalizado ou vocalizações também chamadas de tomadas de tempo, como por exemplo, as vogais alongadas em inglês (Fletcher, 2010).

As medidas de pausa calculadas pelo *script* compreendem apenas medidas de pausa silenciosa, conforme mostra a figura (38). As medidas de pausas silenciosa calculadas automaticamente são a presença ou não a presença de pausa após a unidade VV de análise e a duração dessa pausa silenciosa. Como toda pausa silenciosa necessariamente é um parâmetro marcador de fronteira, adotou-se como padrão de anotação psp para referir à interrupção do sinal sonoro após a fronteira e psdur para referir à duração da interrupção do sinal acústico. Contudo, é importante ressaltar que não há uma correspondência absoluta entre a presença de fronteira prosódica e a presença de pausa silenciosa. São frequentes as fronteiras sem pausa e a pausa, quando ocorre, não diz muito sobre a distinção entre fronteira terminal e não terminal (Raso *et al.*, 2015).

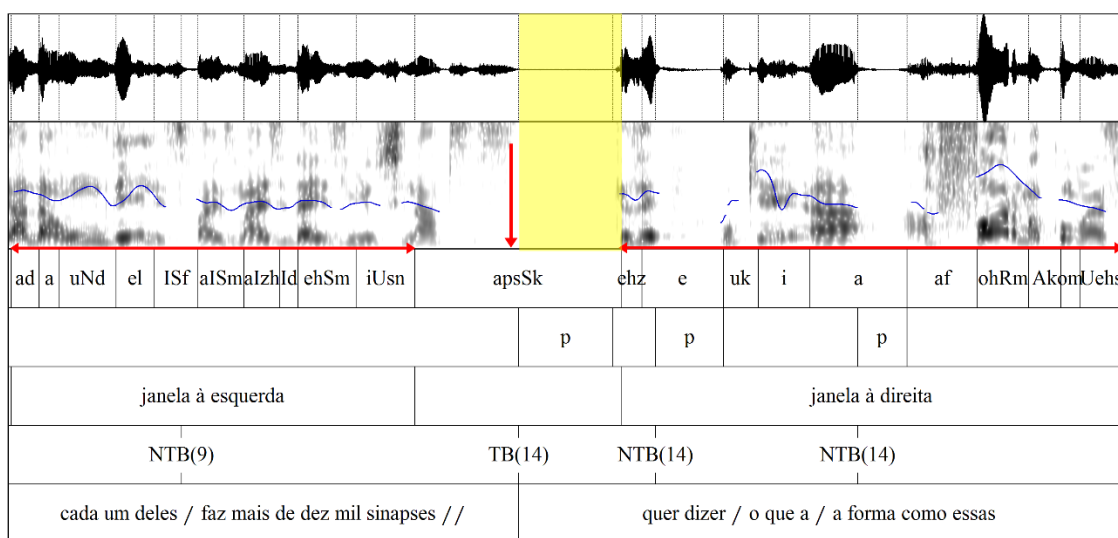


Figura 38: Medidas de pausa silenciosa destacadas no trecho (3) – psp e psdur.

Trecho (3): cada um deles / faz mais de dez mil sinapses // quer dizer / o que a / a forma como essas

Na figura (38), destaca-se uma fronteira terminal percebida por 14 segmentadores. Esta fronteira é seguida imediatamente por uma pausa silenciosa de 520 ms (indicada pela área sombreada da imagem). Na janela à direita da fronteira terminal destacada no centro da figura (38), também há duas outras ocorrências de pausas silenciosas, entretanto, estas pausas silenciosas são usadas para análise de correlato de fronteiras não terminais.

2.3.2 Visão geral dos parâmetros fonético-acústicos calculados pelo *BreakDescriptor*

As tabelas 4, 5, 6, 7 e 8 apresentam uma visão geral das 111 medidas fonético-acústicas calculadas automaticamente para a análise de correlato de fronteiras e não fronteiras. As colunas Medidas, Tipo, Sigla e Fenômeno Fonético indicam respectivamente a medida extraída pelo *script*, o tipo de medida extraída (Global ou Local), a nomenclatura adotada pelo

script e o fenômeno fonético indicado pela medida. Abaixo, a tabela (4) apresenta primeiramente as medidas de taxa de elocução e ritmo.

Tabela 4: Parâmetros fonético-acústicos - Medidas de taxa de elocução (speech rate) e ritmo

Medidas	Tipo	Sigla	Fenômeno fonético
Taxa de unidades VV por segundo	G	srl	Taxa de elocução antes da unidade VV de análise
	G	srr	Taxa de elocução após a unidade VV de análise
	G	srd	Aumento ou declínio da taxa de elocução da janela global localizada após a unidade VV de análise em relação à taxa de elocução da janela global localizada antes da unidade VV de análise
Taxa de unidades VV não-salientes por segundo	G	arl	Taxa de articulação antes da unidade VV de análise
	G	arr	Taxa de articulação após a unidade VV de análise
	G	ard	Aumento ou declínio taxa de articulação da janela global localizada após a unidade VV de análise em relação à taxa de articulação da janela global localizada antes da unidade VV de análise

A tabela (5) apresenta as medidas de duração normalizada do segmento

Tabela 5: Parâmetros fonético-acústicos - Medidas de duração normalizada do segmento

Medidas	Tipo	Sigla	Fenômeno fonético
Média de z-score suavizado de duração da unidade VV	L	zl1	Alongamento ou compressão da unidade VV L1
	L	zl2	Alongamento ou compressão da unidade VV L2
	L	zl3	Alongamento ou compressão da unidade VV L3
	L	zl4	Alongamento ou compressão da unidade VV L4
	L	zl5	Alongamento ou compressão da unidade VV L5
	L	zl6	Alongamento ou compressão da unidade VV L6
	L	zl7	Alongamento ou compressão da unidade VV L7
	L	zl8	Alongamento ou compressão da unidade VV L8
	L	zl9	Alongamento ou compressão da unidade VV L9
	L	zl10	Alongamento ou compressão da unidade VV L10
Média de z-score suavizado de duração da unidade VV	L	zl0	Alongamento ou compressão da unidade VV de análise
Média de z-score suavizado de duração da unidade VV	L	zr1	Alongamento ou compressão da unidade VV R1
	L	zr2	Alongamento ou compressão da unidade VV R2
	L	zr3	Alongamento ou compressão da unidade VV R3
	L	zr4	Alongamento ou compressão da unidade VV R4
	L	zr5	Alongamento ou compressão da unidade VV R5

	L	zr6	Alongamento ou compressão da unidade VV R6
	L	zr7	Alongamento ou compressão da unidade VV R7
	L	zr8	Alongamento ou compressão da unidade VV R8
	L	zr9	Alongamento ou compressão da unidade VV R9
	L	zr10	Alongamento ou compressão da unidade VV R10
Média de z-score suavizado de duração da unidade VV	L	zdloc	Aumento ou declínio da duração normalizada e suavizada da unidade VV de análise em relação à unidade VV imediatamente seguinte
Média de z-score suavizado de duração da unidade VV	G	mzl	Alongamento ou encurtamento geral antes da unidade VV de análise
	G	mzr	Alongamento ou encurtamento geral após a unidade VV de análise
	G	mzd	Aumento ou declínio da duração normalizada e suavizada da janela global localizada após a unidade VV de análise em relação à janela global localizada antes da unidade VV de análise
Desvio-padrão de z-score suavizado de duração das unidades VVs	G	SDzl	Variação no ritmo antes da unidade VV de análise
	G	SDzr	Variação no ritmo após a unidade VV de análise
	G	SDzd	Aumento ou declínio da variação do ritmo da janela global localizada após a unidade VV de análise em relação à janela global localizada antes da unidade VV analisada
Assimetria de z-score suavizado de duração das unidades VVs	G	skzl	Predominância de alongamentos ou encurtamentos em relação à duração intrínseca dos fones que compõem a janela antes da unidade VV de análise
	G	skzr	Predominância de alongamentos ou encurtamentos em relação à duração intrínseca dos fones que compõem a janela após a unidade VV de análise
	G	skzd	Mudança de predominância de alongamentos ou encurtamentos entre as janelas à esquerda e à direita da unidade VV de análise
Taxa de picos de z-score suavizado de duração das unidades VVs por segundo	G	prl	Taxa de saliência duracional antes da unidade VV de análise
	G	prr	Taxa de saliência duracional após a unidade VV de análise
	G	prd	Aumento ou declínio da taxa de saliência duracional da janela global localizada após a unidade VV de análise em relação à taxa de saliência duracional da janela global localizada antes da unidade VV de análise

A tabela (6) apresenta as medidas de frequência fundamental calculadas.

Tabela 6: Parâmetros fonético-acústicos - Medidas de frequência fundamental (f0) em semitons (re 1 Hz)

Medidas	Tipo	Sigla	Fenômeno fonético
Mediana de f0	L	f0medl1	F0 da unidade VV L1
	L	f0medl2	F0 da unidade VV L2
	L	f0medl3	F0 da unidade VV L3
	L	f0medl4	F0 da unidade VV L4
	L	f0medl5	F0 da unidade VV L5
	L	f0medl6	F0 da unidade VV L6
	L	f0medl7	F0 da unidade VV L7
	L	f0medl8	F0 da unidade VV L8
	L	f0medl9	F0 da unidade VV L9
	L	f0medl10	F0 da unidade VV L10
Mediana de f0	L	f0med0	F0 da unidade VV de análise
Mediana de f0	L	f0medr1	F0 da unidade VV R1
	L	f0medr2	F0 da unidade VV R2
	L	f0medr3	F0 da unidade VV R3
	L	f0medr4	F0 da unidade VV R4
	L	f0medr5	F0 da unidade VV R5
	L	f0medr6	F0 da unidade VV R6
	L	f0medr7	F0 da unidade VV R7
	L	f0medr8	F0 da unidade VV R8
	L	f0medr9	F0 da unidade VV R9
	L	f0medr10	F0 da unidade VV R10
Mediana de f0	L	f0meddloc	<i>Reset</i> de f0
Média das medianas de f0	G	f0medl	F0 geral antes da unidade VV de análise
	G	f0medr	F0 geral após a unidade VV de análise
	G	f0medd	Aumento ou declínio da F0 da janela global localizada após a unidade VV de análise em relação à janela global localizada antes da unidade VV de análise
Desvio-padrão das medianas de f0	G	sdf0l	Varição de f0 antes da unidade VV de análise
	G	sdf0r	Varição de f0 após a unidade VV de análise
	G	sdf0d	Aumento ou declínio da variação de f0 da janela global localizada após a unidade VV de análise em relação à janela global localizada antes da unidade VV de análise
Assimetria das medianas de f0	G	skf0l	Predominância de valores de f0 mais graves ou mais agudos antes da unidade VV de análise
	G	skf0r	Predominância de valores de f0 mais graves ou mais agudos após a unidade VV de análise
	G	skf0d	Mudança de predominância de valores de f0 (de graves para agudos ou vice versa) entre as

			janelas à esquerda e à direita da unidade VV de análise
Primeira derivada da mediana de f0	L	df0medl1	Aumento ou declínio de f0 da unidade VV L1 em relação à f0 da unidade VV anterior
	L	df0medl2	Aumento ou declínio de f0 da unidade VV L2 em relação à f0 da unidade VV L1
	L	df0medl3	Aumento ou declínio de f0 da unidade VV L3 em relação à f0 da unidade VV L2
	L	df0medl4	Aumento ou declínio de f0 da unidade VV L4 em relação à f0 da unidade VV L3
	L	df0medl5	Aumento ou declínio de f0 da unidade VV L5 em relação à f0 da unidade VV L4
	L	df0medl6	Aumento ou declínido de f0 da unidade VV L6 em relação à f0 da unidade VV L5
	L	df0medl7	Aumento ou declínio de f0 da unidade VV L7 em relação à f0 da unidade VV L6
	L	df0medl8	Aumento ou declínio de f0 da unidade VV L8 em relação à f0 da unidade VV L7
	L	df0medl9	Aumento ou declínio de f0 da unidade VV L9 em relação à f0 da unidade VV L8
	L	df0medl10	Aumento ou declínio de f0 da unidade VV L10 em relação à f0 da unidade VV L9
Primeira derivada da mediana de f0	L	df0med0	Aumento ou declínio de f0 da unidade VV VV de análise (L0) em relação à f0 da unidade VV L10
Primeira derivada da mediana de f0	L	df0medr1	Aumento ou declínio de f0 da unidade VV R1 em relação à f0 da unidade VV L0
	L	df0medr2	Aumento ou declínio de f0 da unidade VV R2 em relação à f0 da unidade VV R1
	L	df0medr3	Aumento ou declínio de f0 da unidade VV R3 em relação à f0 da unidade VV R2
	L	df0medr4	Aumento ou declínio de f0 da unidade VV R4 em relação à f0 da unidade VV R3
	L	df0medr5	Aumento ou declínio de f0 da unidade VV R5 em relação à f0 da unidade VV R4
	L	df0medr6	Aumento ou declínio de f0 da unidade VV R6 em relação à f0 da unidade VV R5
	L	df0medr7	Aumento ou declínio de f0 da unidade VV R7 em relação à f0 da unidade VV R6
	L	df0medr8	Aumento ou declínio de f0 da unidade VV R8 em relação à f0 da unidade VV R7
	L	f0medr9	Aumento ou declínio de f0 da unidade VV R9 em relação à f0 da unidade VV R8
	L	df0medr10	Aumento ou declínio de f0 da unidade VV R10 em relação à f0 da unidade VV R9
Primeira derivada da mediana de f0	L	df0meddloc	Indica se a forma de subida ou descida de f0 entre as unidades R1 e L0 é igual ou diferente
Média das primeiras derivadas das medianas de f0	G	df0medl	Tendência geral de mudanças no contorno entonacional de f0 antes da unidade VV de análise

	G	df0medr	Tendência geral de mudanças no contorno entonacional de f0 após a unidade VV de análise
	G	df0medd	Diferença entre médias de declínio ou aumento de f0 antes e depois da posição de análise
Desvio-padrão das primeiras derivadas das medianas de f0	G	sddf0l	Variação das mudanças no contorno entonacional de f0 (regularidade das mudanças de f0) antes da unidade VV de análise
	G	sddf0r	Variação das mudanças no contorno entonacional de f0 (regularidade das mudanças de f0) após a unidade VV de análise
	G	sddf0d	Aumento ou declínio da variação das mudanças no contorno entonacional de f0 (regularidade das mudanças de f0) da janela global localizada após a unidade VV de análise em relação à variação das mudanças no contorno entonacional de f0 (regularidade das mudanças de f0) da janela global localizada antes da unidade VV de análise
Assimetria das primeiras derivadas das medianas de f0	G	skdf0l	Predominância de contorno entonacional de f0 crescente ou decrescente antes da unidade VV de análise
	G	skdf0r	Predominância de contorno entonacional de f0 crescente ou decrescente após a unidade VV de análise
	G	skdf0d	Mudança de Predominância de contorno entonacional de f0 após a unidade VV de análise
Taxa de picos de f0 suavizado por segundo	G	f0rl	Taxa de proeminência de f0 antes da unidade VV de análise
	G	f0rr	Taxa de proeminência de f0 após a unidade VV de análise
	G	f0rd	Aumento ou declínio da taxa de proeminência de f0 da janela global localizada após a unidade VV de análise em relação à taxa de proeminência de f0 da janela global localizada antes da unidade VV de análise

A tabela (7) apresenta as medidas de intensidade calculadas.

Tabela 7: Parâmetros fonético-acústicos - Medidas de intensidade

Medidas	Tipo	Sigla	Fenômeno fonético
Média de ênfase espectral	G	emphl	Intensidade antes da unidade VV de análise
	G	emphr	Intensidade após a unidade VV de análise
	G	emphd	Aumento ou declínio de intensidade da janela global localizada após a unidade VV de análise em relação à intensidade da janela global localizada antes da unidade VV de análise
Média de ênfase espectral	L	emphdloc	Aumento ou declínio de intensidade da unidade VV de análise em relação à unidade VV imediatamente seguinte

A tabela (8) apresenta as medidas de pausa silenciosa extraídas.

Tabela 8: Parâmetros fonético-acústicos - Medidas de pausa silenciosa

Medidas	Tipo	Sigla	Fenômeno fonético
Presença de pausa	L	psp	Interrupção do sinal acústico após a unidade VV de análise
Duração de pausa	L	psdur	Duração da interrupção do sinal acústico após a unidade VV de análise

2.4. Análise estatística

Os classificadores estatísticos são técnicas estatísticas capazes de realizar decisões em contextos complexos. Os classificadores possuem uma alta capacidade estatística para categorizar dados em grupos distintos, pois eles têm como ponto de referência um conjunto de dados em que já são conhecidos os grupos analisados, as variáveis e os valores das variáveis apresentados em cada grupo. Por isso, é recorrente dizer que os classificadores têm um alto poder de predição.

Os classificadores estatísticos também podem ser chamados de modelos de regressão logística. Uma regressão logística é um tipo de modelo preditivo que necessita de variáveis dependentes discretas e variáveis independentes (também chamadas de variáveis de previsão) contínuas, discretas ou ambas para realizar suas análises.

As variáveis dependentes discretas deste trabalho são as categorias associadas às fronteiras prosódicas percebidas pelos segmentadores, as variáveis independentes são os parâmetros fonético-acústicos extraídos automaticamente para cada posição de análise localizada em limite de palavra fonológica. Os classificadores operam inferindo a contribuição de cada variável independente para prever e discriminar estatisticamente as variáveis dependentes. Para que a predição seja a melhor possível, é necessário um conjunto de dados coletados em duas partes disjuntas. A importância de se coletar os dados em duas partes disjuntas é justificada pelo fato de que as técnicas de regressão logística, em sua maioria, são realizadas por meio de duas etapas diferentes entre si: uma etapa de treinamento e outra de teste.

O treinamento é a etapa responsável pela análise multivariada dos dados previamente classificados em grupos distintos. Durante esta etapa, busca-se uma explicação estatística para os diferentes grupos usando as variáveis de predição disponíveis nos grupos. Esta explicação pode ser chamada de equação discriminante, função discriminante, modelo ou classificador. O teste tem como objetivo avaliar a eficácia do classificador em distinguir os grupos analisados com base na função discriminante encontrada para os grupos. É fundamental que o teste utilize dados diferentes daqueles empregados durante a etapa de treino, porque desta forma é possível avaliar a eficácia das decisões tomadas pelo classificador em dados não treinados. Este procedimento final de teste é denominado classificação, porque em geral, ele classifica dados não treinados baseando-se na função discriminante obtida pelo classificador.

Um dos possíveis contextos de aplicação dos classificadores estatísticos é a classificação das fronteiras e não fronteiras prosódicas. Os classificadores parecem ser uma boa alternativa para o problema das fronteiras, porque a produção e a percepção de fronteiras e não fronteiras dependem de uma série de variáveis complexas, que são os parâmetros fonético-acústicos do fluxo da fala.

Os parâmetros fonético-acústicos extraídos automaticamente pelo *BreakDescriptor* foram submetidos a dois modelos estatísticos de classificação amplamente conhecidos, a saber, o *Random Forest* (RF) e o *Linear Discriminant Analysis* (LDA). Em geral, o RF e o LDA são métodos estatísticos usados para o reconhecimento de padrões e

características presentes em duas ou mais categorias. Como são modelos de classificação estatística, para reconhecer e distinguir padrões presentes em duas ou mais categorias, tanto o RF, quanto o LDA, usam variáveis de previsão em seus cálculos. Para os fins deste trabalho, o RF e o LDA foram usados para identificar a combinação de medidas que melhor explicam a segmentação realizada pelos segmentadores perceptualmente, buscando uma maior discriminação possível entre os dois tipos de fronteiras. Evidentemente, as variáveis de previsão usadas neste trabalho são o conjunto de parâmetros fonético-acústicos extraídos nas posições de fronteiras de palavras fonológicas e em torno delas, quer coincidente com fronteira prosódica ou não.

Primeiramente, estas técnicas observam o comportamento das variáveis disponíveis em dados já classificados. Em seguida, as técnicas buscam uma combinação de características capaz de separar as categorias analisadas. Para isso, ambas estimam uma hierarquia de variáveis e o peso atribuído a cada variável dentro das categorias que se busca diferenciar e caracterizar.

O LDA e RF atribuem a cada uma das variáveis de previsão um peso que reflete a contribuição individual de cada variável para classificar adequadamente a categoria analisada. Tendo em vista este papel dos pesos, a variável que apresentar o menor valor absoluto de peso é a variável menos relevante hierarquicamente para realização da classificação estatística da categoria analisada. Assim, quanto maior for o peso atribuído a uma variável, maior é a importância desta para a classificação adequada da categoria.

LDA e Random Forest são duas técnicas estatísticas distintas que apresentarão modelos diferentes para aprender a associar dados numéricos e categóricos às categorias já classificadas previamente. O LDA distingue as categorias por meio de uma regressão linear, já o Random Forest distingue as categorias por meio de árvores, também chamadas de nós de decisão. Deste modo, serão produzidos resultados diferentes para correlacionar a presença ou a ausência de fronteiras prosódicas aos parâmetros fonético-acústicos.

Num primeiro momento, a principal justificativa de se trabalhar com dois classificadores distintos é desenvolver uma análise parcial dos dados com os modelos de classificação estatística RF e LDA. O desenvolvimento de uma análise parcial dos dados é uma etapa crucial da pesquisa, porque ela pode ser vista como um bom indicador de qual classificador é mais adequado para classificar e distinguir as categorias analisadas.

Tanto para as fronteiras terminais, quanto para as fronteiras não terminais, foram considerados, em ambos os modelos de classificação, duas categorias distintas: a presença e a

ausência de fronteira prosódica. Para a categoria presença de fronteira, foi considerado também a concordância de pelo menos 50% entre segmentadores *experts*, isto é, pelo menos sete pessoas. Considerou-se também o poder de predição do LDA e do RF. A predição compara a classificação perceptual prévia dos grupos analisados com a classificação automática feita pelo modelo obtido, mostrando acertos e falsos alarmes para um conjunto de dados.

Assim, a predição indica ao pesquisador o quanto o modelo estatístico obtido se aproxima às decisões tomadas pela maioria dos segmentadores, ou seja, em que medida o modelo reproduz os resultados obtidos com base na percepção humana. Em função disso, os classificadores permitem investigar os parâmetros que supostamente guiam a produção e a percepção de fronteiras prosódicas no fluxo da fala espontânea.

Visando esclarecer o que é uma predição de um determinado classificador estatístico, nos próximos parágrafos, são apresentados em sequência exemplos de predições e a forma de analisá-las.

Por exemplo, se os segmentadores indicarem em uma amostra de fala 10 presenças de fronteiras terminais e 100 ausências de fronteiras terminais, a predição mostrará o número de fronteiras terminais que foram detectadas pela função discriminante do modelo. A predição mostrará também quantas ausências de fronteira terminal, de acordo com a percepção dos segmentadores, foram consideradas incorretamente pelo modelo como presença de fronteira terminal. Caso o modelo indique 10 fronteiras terminais e 100 ausências de fronteiras terminais, o modelo alcançou o melhor desempenho possível, porque ele obteve 100% de acerto para a presença de fronteira terminal e 0% de falso alarme para a ausência de fronteira terminal.

Outra possibilidade seria o modelo identificar, dentre as 10 fronteiras terminais marcadas pelos segmentadores na amostra, somente sete fronteiras terminais. Isto é, o modelo deixou de identificar três fronteiras marcadas pelos segmentadores. Neste caso, o modelo obteve 70% de acerto para a presença de fronteira terminal. Caso o modelo identifique incorretamente 10 presenças de fronteira terminal classificadas perceptualmente pelos segmentadores como ausência de fronteira terminal, o modelo obteve 10% de falsos alarmes para a ausência de fronteira terminal, pois, ele identificou 10 fronteiras terminais em casos em que os segmentadores marcaram como ausência de fronteira.

A análise estatística foi realizada em três etapas principais explicadas nas próximas seções. Toda a análise estatística foi realizada com o ambiente para computação estatística R (R Core Team, 2017). O classificador LDA é incluído no pacote MASS (Venables e Ripley,

2002) - função `lda` (). O classificador RF é encontrado no pacote `randomForest` (Liaw e Wiener, 2002) - função `randomForest` ()).

2.4.1 Avaliação Inicial

A etapa de avaliação inicial dos classificadores *Random Forest* e *Linear Discriminant Analysis* utilizou 70% das unidades VVs localizadas em limites de palavras fonológicas na amostra I, independentemente de as VVs de análise serem usadas para a análise de correlato de fronteiras ou não fronteiras. As unidades VVs usadas foram escolhidas aleatoriamente. A opção de utilizar apenas 70% das unidades VVs durante a etapa de treinamento é justificada pela necessidade de realizar posteriormente uma etapa de teste para ambos os modelos. A avaliação inicial foi composta por apenas duas etapas de treinamento, porque o foco desta fase era exclusivamente avaliar qual método alcançaria o melhor desempenho para identificar a combinação de medidas capazes de prever a realização de fronteiras prosódicas. Assim, na etapa de avaliação inicial, não foi realizado o teste no subconjunto de dados com 30% restante das unidades VVs não treinadas. Em um primeiro momento, o treinamento usou todas as 111 medidas fonético-acústicas extraídas, em seguida, foi desenvolvido outro treinamento usando apenas 28 medidas.

2.4.2. Refinamento do classificador

Após a avaliação inicial dos modelos RF e LDA, apenas o LDA foi refinado de várias formas distintas. Evidentemente, o modelo que foi refinado foi aquele que apresentou, durante a avaliação inicial, melhores resultados para identificar a presença dos dois tipos de fronteiras prosódicas. O processo de refinamento tem como foco melhorar o desempenho do classificador, buscando uma melhor reprodução do comportamento dos segmentadores obtido com base na percepção humana.

O refinamento do classificador que alcançou melhor desempenho na avaliação inicial também foi composto por duas etapas, uma etapa de treinamento e outra de teste. Contudo, ao invés de usar apenas uma parte dos 111 parâmetros extraídos automaticamente pelo *BreakDescriptor*, o treinamento utilizou todos eles. Semelhantemente à etapa de avaliação inicial do RF e LDA, as unidades VVs usadas foram escolhidas aleatoriamente. Foram usadas 70% das unidades VVs localizadas em limites de palavras fonológicas na amostra I, independentemente das unidades VVs escolhidas serem correspondentes a pontos de fronteiras ou não fronteiras.

O teste foi realizado com os 30% restantes de unidades VVs excluídas do treinamento. Este teste foi realizado somente nos modelos finais, ou seja, aqueles cujo treinamento foi considerado satisfatório. O objetivo do teste é comparar a percepção humana de fronteiras prosódicas ao longo do fluxo da fala com a marcação automática de fronteiras adotando o modelo classificador estatístico refinado. Este procedimento final é denominado classificação, porque, em geral, ele classifica novos dados com base no modelo obtido pelas técnicas do classificador estatístico obtidas durante a etapa de treinamento.

Como na avaliação inicial, as categorias usadas foram a presença e a ausência de fronteira terminal e não terminal. Foram considerados também todos os 111 parâmetros fonético-acústicos extraídos automaticamente pelo *BreakDescriptor*, o acordo de pelo menos 50% entre segmentadores *experts* e o poder de predição alcançado pelo modelo.

Os 111 parâmetros foram eliminados de duas formas diferentes. Primeiramente, foram eliminados progressivamente os parâmetros fonético-acústicos que não se mostraram relevantes hierarquicamente para a classificação, de acordo com o peso atribuído pelo modelo às variáveis de previsão. Em seguida, as medidas não apontadas na literatura da área como relevantes para o estabelecimento de fronteiras foram eliminadas.

Este duplo processo de treinamento tem como objetivo reduzir o "ruído" no modelo gerado por parâmetros não relevantes para o estabelecimento de fronteiras, aumentar a porcentagem de acertos para a presença de fronteira e reduzir aquela de falsos alarmes para a ausência de fronteira. Deste modo, o refinamento do modelo é importante para identificar o momento em que o modelo atinge o pico de acertos (maior porcentagem de acertos) para a presença de fronteira e, de forma complementar, a menor porcentagem de falsos alarmes para a ausência de fronteira.

Embora demorada, a eliminação progressiva de medidas em conjunto com a eliminação de medidas com base nos fenômenos fonéticos que elas representam traz vantagens.

A primeira delas é a exclusão dos parâmetros que podem prejudicar o desempenho de classificação do modelo. A segunda vantagem é a possibilidade de obter um melhor conhecimento sobre os parâmetros mais relevantes que orientam o estabelecimento de fronteiras no fluxo da fala espontânea. Em termos práticos, a terceira vantagem está diretamente ligada à exclusão de medidas não relevantes para a marcação de fronteiras, pois, ao eliminá-las, o modelo refinado estatisticamente permite a modelagem de uma ferramenta computacional econômica no que diz respeito à quantidade de variáveis necessárias para segmentar automaticamente, ou pelo menos, semiautomaticamente, o fluxo da fala em unidades entonacionais marcadas por fronteiras.

2.4.3. Refinamento do modelo de fronteiras não terminais

Como já foi dito, os classificadores estatísticos são técnicas estatísticas multivariadas utilizadas para discriminar e classificar objetos de uma população em dois ou mais grupos a partir de uma série de variáveis de previsão. Os resultados obtidos tanto na avaliação inicial de dois classificadores, quanto no refinamento de apenas um deles indicaram uma maior dificuldade para classificar fronteiras não terminais. Por isso, a detecção automática de fronteiras não conclusivas no fluxo da fala tende a ser uma tarefa muito mais complexa.

A fim de se obter uma melhor correlação entre a percepção de fronteira não terminal e parâmetros fonético-acústicos, desenvolveu-se, paralelamente ao refinamento de um dos modelos RF ou LDA, um refinamento exclusivo para as fronteiras não terminais. O refinamento partiu inicialmente da percepção de fronteiras prosódicas em um sistema binário categorizado entre terminal e não terminal, contudo, buscou uma categorização mais ampla de subtipos de fronteiras não terminais. Este refinamento tem como foco investigar subgrupos de fronteiras não conclusivas e melhorar a classificação automática do modelo.

O refinamento do classificador que alcançou melhor desempenho para classificar fronteiras não terminais foi composto apenas pela etapa de treinamento. Assim, ao invés de selecionar aleatoriamente 70% das unidades VVs da amostra I localizadas em limites de palavras fonológicas, 100% das unidades VVs correspondentes às fronteiras não conclusivas, de acordo com a percepção dos segmentadores humanos, foram submetidas ao treinamento. Optou-se por este tipo de treinamento para minimizar o risco de que fronteiras casualmente não

usadas no treinamento do modelo apareçam em outros, dificultando a classificação estatística. Deste modo, a etapa de teste utilizando 30% de dados não treinados não foi realizada, pois, todas as ocorrências de fronteiras não terminais foram submetidas ao treinamento do classificador estatístico.

Nesta fase de treinamento, as categorias usadas foram a presença e a ausência de fronteira não terminal. Foram considerados também todos os 111 parâmetros fonético-acústicos extraídos automaticamente pelo *BreakDescriptor*, o acordo de pelo menos 50% entre segmentadores *experts* e o poder de predição do modelo.

3. Análises estatísticas e discussão dos resultados

O trabalho foi desenvolvido em diversas fases interligadas entre si. Neste capítulo, são apresentados os resultados e a discussão dos resultados obtidos nas análises estatísticas desenvolvidas. A análise estatística compreendeu várias etapas de treinamento dos modelos e uma etapa final.

Três dos quatro treinamentos desenvolvidos utilizaram em suas análises 70% das unidades VVs localizadas em limites de palavras fonológicas, independentemente das unidades VVs serem correspondentes à presença ou ausência de fronteira prosódica. Um único treinamento realizado utilizou 100% das unidades VVs correspondentes às fronteiras não terminais. Nos treinamentos em que foi utilizada uma parte das unidades VVs, elas foram retiradas aleatoriamente da amostra I correspondente à fala monológica masculina. Nos quatro treinamentos realizados, considerou-se como posição de fronteira prosódica as unidades VVs, localizadas em limite de palavras fonológicas, que foram indicadas como posições de fronteira prosódica por pelo menos sete dos 14 segmentadores humanos. As demais unidades VVs, também localizadas em limite de palavras fonológicas, são consideradas posições de ausência de fronteira prosódica. O teste no subconjunto de dados correspondente a 30% dos dados não treinados foi realizado para os dois melhores modelos de classificação automática de fronteiras prosódica.

A primeira etapa desenvolvida avaliou o desempenho dos classificadores *Random Forest* e *Linear Discriminant Analysis*. O principal objetivo desta primeira fase é verificar qual dos classificadores é mais adequado para classificar automaticamente as categorias presença e ausência de fronteira prosódica terminal e não terminal percebidas pelos segmentadores. A partir dos resultados obtidos na avaliação inicial dos dois classificadores, verificou-se que o classificador LDA obteve um melhor desempenho para identificar as categorias analisadas. Por isso, o LDA foi refinado por meio de várias etapas de treinamento diferentes.

A primeira fase de refinamento e treinamento do classificador LDA eliminou progressivamente algumas das 111 medidas extraídas automaticamente pelo *BreakDescriptor*, de acordo com o peso atribuído pelo classificador a cada uma das variáveis de previsão. As medidas foram eliminadas, retirando-se do modelo aquelas que foram menos relevantes hierarquicamente. Elas foram eliminadas progressivamente até os modelos de classificação de

fronteiras terminais e não terminais serem reduzidos a 21 variáveis. A eliminação de medidas foi realizada retirando-se simultaneamente do modelo 10 ou cinco medidas.

Assim, o classificador começou o treinamento com todas as 111 medidas; sucessivamente, as 10 medidas menos relevantes hierarquicamente foram eliminadas, permanecendo 101 medidas; em seguida, mais 10 medidas menos relevantes hierarquicamente foram eliminadas, permanecendo, então, 91 medidas e assim sucessivamente. A eliminação foi feita de 10 em 10 medidas até o modelo ser reduzido a 71 medidas. A partir do ponto em que o modelo foi reduzido a 71 medidas, as medidas foram eliminadas de cinco em cinco até o modelo ser reduzido a 21 medidas.

Os resultados obtidos através da eliminação progressiva de medidas com base no peso atribuído pelo LDA mostraram que o desempenho do modelo para classificar fronteiras terminais e não terminais diminuiu sempre com a retirada de medidas. Por isso, desenvolveu-se um novo método de treinamento do classificador LDA.

A segunda etapa de refinamento e treinamento do LDA eliminou as medidas extraídas pelo *BreakDescriptor* com base nos fenômenos fonéticos representados por cada uma das variáveis de previsão. Nesta fase, tendo como base a revisão de literatura sobre os parâmetros que estabelecem fronteiras prosódicas nas diversas línguas, foram eliminadas as medidas consideradas não relevantes para o estabelecimento de fronteiras independentemente do peso indicado pelo modelo. Deste modo, avaliou-se o poder de predição do LDA, mantendo no modelo as medidas amplamente citadas na literatura da área como relevantes para o estabelecimento de fronteiras.

Inicialmente, a análise estatística usou todas as ocorrências de fronteiras terminais e não terminais discriminadas na amostra I. A fim de minimizar os erros para detectar fronteiras não terminais, optou-se por excluir da análise alguns casos de fronteiras não terminais: *retractings*, “né” e tomadas de tempo. Por isso, a amostra I foi parcialmente reduzida para realizar uma terceira etapa de refinamento e treinamento destinada exclusivamente às fronteiras não terminais. Esta etapa de treinamento também eliminou as medidas consideradas não relevantes para o estabelecimento de fronteiras prosódicas.

Os dois melhores modelos encontrados para classificar fronteiras terminais e não terminais foram submetidos ao teste final, usando o subconjunto de dados correspondente a 30%. No teste, o modelo alcançou 80% de acerto para classificar fronteiras terminais e 50% de acerto para classificar não terminais.

As etapas de treinamento e de teste desenvolvidas até então enfatizaram por meio de seus resultados a necessidade de desenvolver um melhor método de classificação automática para as fronteiras não terminais. Uma forma de buscar um aumento da porcentagem de acerto para detectar fronteiras não terminais é submeter 100% dos dados ao treinamento e não realizar o teste final em dados não treinados, porque aumenta-se o conjunto de dados submetido à análise estatística. A quarta e última fase de treinamento realizada adotou este procedimento e buscou investigar subtipos de fronteiras não terminais, eliminando as fronteiras acertadas pelo modelo anterior e criando modelos novos para as fronteiras remanescentes. Os resultados apontam que há três subgrupos de fronteiras não conclusivas, caracterizados cada um deles por configurações de parâmetros fonético-acústicos distintos. Na amostra I, com estes três subgrupos, são identificadas 220 de 225 de fronteiras não terminais.

3.1. Resultados da avaliação inicial dos classificadores *Random Forest* e *Linear Discriminant Analysis*

Apresentamos abaixo os resultados obtidos durante a avaliação inicial dos classificadores RF e LDA. A avaliação inicial dos classificadores RF e LDA foi composta por duas fases de treinamento (I e II). A fase I de treinamento de avaliação inicial dos dois classificadores utilizou todas as 111 medidas extraídas automaticamente pelo *BreakDescriptor* para cada unidade VV localizada em fronteira de palavra fonológica, a fase II utilizou 28 medidas.

Abaixo, na tabela (9), são apresentadas as medidas usadas na segunda fase de treinamento da avaliação inicial.

Tabela 9: Medidas usadas na segunda fase de treinamento da avaliação inicial dos classificadores RF e LDA

Grupo de medidas	Sigla	Fenômeno fonético
Medidas de pausa	psp	Interrupção do sinal acústico após a unidade VV de análise
	psdur	Duração da interrupção do sinal acústico após a unidade VV de análise
Medidas de frequência fundamental	df0med0	Aumento ou declínio de f0 na unidade VV de análise
	df0medl	Tendência geral de mudanças no contorno entonacional de f0 antes da unidade VV de análise
	df0medl10	Aumento ou declínio de f0 na unidade VV L10
	df0medr1	Aumento ou declínio de f0 na unidade VV R1
	f0med0	F0 da unidade VV de análise
	f0medd	Aumento ou declínio da f0 da janela global localizada após a unidade VV de análise em relação à f0 da janela global localizada antes da unidade VV de análise
	f0meddloc	Reset de f0
	f0medl	F0 média antes da unidade VV de análise
	f0medl10	F0 da unidade VV L10
	f0medr1	F0 da unidade VV R1
	f0rl	Taxa de proeminência de f0 antes da unidade VV de análise
	f0rr	Taxa de proeminência de f0 após a unidade VV de análise
	sddf0l	Variação das mudanças de f0 (regularidade das mudanças de f0) antes da unidade VV de análise
	sdf0l	Variação de f0 antes da unidade VV de análise
	Medidas de duração normalizada e suavizada dos segmentos silábicos	mzl
mzr		Alongamento ou encurtamento geral após a unidade VV de análise
prl		Taxa de saliência duracional antes da unidade VV de análise
prr		Taxa de saliência duracional após a unidade VV de análise
SDzl		Predominância de alongamentos ou encurtamentos em relação à duração intrínseca dos fones que compõem a janela antes da unidade VV de análise
SDzr		Predominância de alongamentos ou encurtamentos em relação à duração intrínseca dos fones que compõem a janela após a unidade VV de análise
zl0		Alongamento ou compressão da unidade VV de análise
zl10		Alongamento ou compressão da unidade VV L10 em relação à unidade VV de análise
zr1		Alongamento ou compressão da unidade VV R1 em relação à unidade VV de análise
Medidas de taxa de elocução (speech rate) e ritmo		srl
	arl	Taxa de articulação antes da unidade VV de análise

Medidas de intensidade	emphdloc	Aumento ou declínio de intensidade da unidade VV de análise em relação à unidade VV imediatamente seguinte
------------------------	----------	--

Na tabela (10), são apresentadas as ocorrências de presença e ausência de fronteiras prosódicas escolhidas aleatoriamente para a avaliação inicial dos dois classificadores estatísticos adotados.

Tabela 10: Treinamento inicial – Frequência bruta de unidades VVs escolhidas aleatoriamente

Fronteira	Presença	Ausência
Terminal	49	744
Não terminal	179	614

No total, 49 unidades VVs correspondem a fronteiras terminais, 744 à ausência de fronteira terminal, 179 a fronteiras não terminais e 614 à ausência de fronteira não terminal.

3.1.1. Poder de predição da fase I de avaliação inicial dos classificadores estatísticos

A seguir, a tabela (11) mostra os resultados alcançados pelo classificador *Random Forest* na primeira fase de avaliação inicial dos classificadores.

Tabela 11: Poder de predição alcançado pelo RF usando 111 medidas

Fronteira	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
Terminal	10%	100%	90%	0%
Não terminal	16%	94%	84%	6%

Com todas as 111 medidas extraídas pelo *BreakDescriptor*, o classificador RF acertou 10% das ocorrências de fronteiras terminais percebidas pelos segmentadores, que

foram escolhidas aleatoriamente para a análise estatística inicial. Em outras palavras, o modelo RF apresentou uma convergência de 10% em relação às fronteiras terminais marcadas perceptualmente pelos segmentadores humanos. Por isso, 90% das fronteiras terminais não foram identificadas automaticamente pelo modelo.

O modelo RF apresentou também uma convergência de 100% com os pontos em que os segmentadores indicaram a ausência de fronteira conclusiva. Com isso, o modelo RF não obteve falsos alarmes em relação à ausência de fronteiras terminais. Os falsos alarmes são erros relativos às posições em que os segmentadores indicaram ausência de fronteira conclusiva e o modelo, entretanto, indicou inadequadamente a presença de fronteira prosódica terminal.

O modelo RF acertou 16% das fronteiras não terminais percebidas pelos segmentadores, que foram escolhidas aleatoriamente para análise estatística inicial. Em outras palavras, o modelo RF apresentou uma convergência de 16% em relação às fronteiras não terminais. Conseqüentemente, 84% das fronteiras não terminais deixaram de ser identificadas automaticamente pelo modelo. O modelo RF apresentou uma convergência de 94% com os pontos em que os segmentadores marcaram a ausência de fronteiras não terminais. Assim, o modelo RF obteve 6% de falsos alarmes em relação à ausência de fronteiras não terminais. Os falsos alarmes neste caso são erros relativos às posições em que os segmentadores indicaram a ausência de fronteiras não terminais e o modelo, contudo, indicou inadequadamente a presença de fronteira prosódica não conclusiva.

Abaixo, a tabela (12) mostra os resultados alcançados pelo classificador *Linear Discriminant Analysis* na primeira fase de avaliação inicial dos classificadores.

Tabela 12: : Poder de predição alcançado pelo LDA usando 111 medidas

Fronteira	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
Terminal	73%	97%	27%	3%
Não terminal	37%	95%	63%	5%

Com todas as 111 medidas extraídas pelo *BreakDescriptor*, o classificador LDA acertou 73% das ocorrências de fronteiras terminais percebidas pelos segmentadores. Em outras palavras, o modelo LDA apresentou uma convergência de 73% em relação às fronteiras terminais marcadas com base na percepção de segmentadores humanos. Deste modo, 27% das

fronteiras terminais não foram identificadas automaticamente pelo modelo. O modelo LDA apresentou uma convergência de 97% com os pontos em que os segmentadores indicaram a ausência de fronteira terminal. Com isso, o modelo LDA obteve 3% falsos alarmes em relação à ausência de fronteiras terminais.

O modelo LDA acertou 37% das fronteiras não terminais percebidas pelos segmentadores. Em outras palavras, o modelo LDA apresentou uma convergência de 37% em relação às fronteiras não terminais. Conseqüentemente, 63% das fronteiras não terminais deixaram de ser identificadas automaticamente pelo modelo. O modelo LDA apresentou uma convergência de 95% com os pontos em que os segmentadores marcaram a ausência de fronteiras não terminais. Assim, o modelo LDA obteve 5% de falsos alarmes em relação à ausência de fronteiras não terminais.

3.1.2. Poder de predição da fase II de avaliação inicial dos classificadores estatísticos

A seguir, a tabela (13) mostra os resultados alcançados pelo classificador *Random Forest* na segunda fase de avaliação inicial dos classificadores.

Tabela 13: Poder de predição alcançado pelo RF usando 28 medidas

Fronteira	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
Terminal	27%	100%	73%	0%
Não terminal	24%	93%	76%	7%

Com 28 medidas, o classificador RF acertou 27% das ocorrências de fronteiras terminais percebidas pelos segmentadores, isto é, o modelo RF apresentou uma convergência de 27% em relação às fronteiras terminais. Por isso, 73% das fronteiras terminais não foram identificadas automaticamente pelo modelo. O modelo RF apresentou também uma convergência de 100% com os pontos em que os segmentadores indicaram a ausência de fronteira conclusiva. Com isso, o modelo RF não obteve falsos alarmes em relação à ausência

de fronteiras terminais.

O modelo RF acertou 24% das fronteiras não terminais percebidas pelos segmentadores. Em outras palavras, o modelo RF apresentou uma convergência de 24% em relação às fronteiras não terminais. Conseqüentemente, 76% das fronteiras não terminais deixaram de ser identificadas automaticamente pelo modelo. Em relação à ausência de fronteiras não terminais, o modelo RF apresentou uma convergência de 93% com os pontos em que os segmentadores marcaram. Assim, o modelo RF obteve 7% de falsos alarmes em relação à ausência de fronteiras não terminais.

A tabela (14) mostra os resultados alcançados pelo classificador LDA na segunda fase de avaliação inicial dos classificadores.

Tabela 14: Poder de predição alcançado pelo LDA usando 28 medidas

Fronteira	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
Terminal	65%	95%	35%	5%
Não terminal	27%	95%	73%	5%

Com 28 medidas, o modelo LDA acertou 65% das fronteiras terminais, ou seja, o modelo LDA apresentou uma convergência de 65% com a categoria presença de fronteira terminal marcada perceptualmente pelos segmentadores. Por outro lado, 35% das fronteiras terminais não foram identificadas pelo LDA. O modelo LDA apresentou uma convergência de 95% com os pontos em que os segmentadores não perceberam fronteiras terminais. Por isso, o modelo LDA obteve 5% de falsos alarmes em relação à ausência de fronteiras conclusivas.

O modelo LDA acertou 27% das fronteiras não terminais, isto é, o LDA apresentou uma convergência de 27% com a categoria presença de fronteira não terminal marcada com base na percepção de segmentadores *experts*. Deste modo, 73% das fronteiras não terminais não foram identificadas. O modelo LDA apresentou uma convergência de 95% com os pontos em que os segmentadores indicaram ausência de fronteira não terminal. Assim sendo, o modelo LDA obteve 5% de falsos alarmes em relação à ausência de fronteiras não terminais.

3.1.3. Discussão da avaliação inicial de classificadores estatísticos

Dois modelos de classificação estatística, *Random Forest* e *Linear Discriminant Analysis*, foram utilizados para gerar modelos de combinações de parâmetros fonético-acústicos capazes de prever a realização das fronteiras prosódicas percebidas pelos segmentadores.

O LDA, nas fases I e II de avaliação inicial dos classificadores, apresentou uma maior porcentagem de acerto na previsão de fronteiras terminais e não terminais. Porém, nas duas fases de avaliação inicial dos classificadores, observa-se que o LDA apresentou uma maior porcentagem de falsos alarmes para identificar a ausência de fronteiras terminais. Adotando 111 medidas, o classificador LDA obteve 3% de falsos alarmes para a ausência de fronteira terminal, enquanto o RF obteve 0%. Usando 28 medidas, o LDA obteve 5% de falsos alarmes para a ausência de fronteira terminal, enquanto o RF obteve 0%.

Em relação às fronteiras não terminais, o LDA, nas duas fases de avaliação inicial, apresentou uma menor porcentagem de falsos alarmes. Com 111 medidas, o LDA obteve 5% de falsos alarmes, enquanto RF obteve 6%. Usando 28 medidas, o LDA obteve 5% de falsos alarmes, enquanto o RF obteve 7%.

Em geral, o classificador LDA parece ser uma melhor alternativa para identificar tanto a presença de fronteiras prosódicas (terminais e não terminais), quanto a ausência de fronteiras não terminais. O classificador RF classificou melhor unicamente a categoria ausência de fronteiras terminais.

A comparação entre os resultados obtidos por meio do LDA com 111 e 28 medidas revela que a eliminação de algumas medidas piorou o desempenho do modelo para classificar automaticamente a presença de fronteiras terminais e não terminais, pois, o poder de predição do LDA reduziu 8% para classificar fronteira terminal e 10% para classificar fronteira não terminal.

A comparação entre os resultados obtidos por meio do RF com 111 e 28 medidas revela que a redução das medidas melhorou o desempenho do modelo para classificar automaticamente a presença de fronteiras terminais e não terminais, porque, após a eliminação de 83 medidas, o poder de predição do RF aumentou 17% para fronteira terminal e 8% para fronteira não terminal.

Mesmo assim, optou-se por refinar o classificador LDA de várias formas distintas, porque o LDA, se comparado ao desempenho do RF, apresentou um desempenho indubitavelmente superior na classificação de fronteiras prosódicas. Para as fronteiras não terminais, o poder de predição alcançado pelo LDA é 3% superior. Para as fronteiras terminais, o poder de predição alcançado pelo LDA é 38% maior.

Foi observado também um pior desempenho do LDA para classificar a categoria ausência de fronteira terminal. Contudo, a porcentagem de falsos alarmes para a ausência de fronteira terminal é relativamente baixa, não ultrapassando 5%. Para melhorar o desempenho do LDA, se concluiu que seria necessário reincluir no modelo de classificação estatística medidas diferentes das 28 medidas avaliadas até então e eventualmente retirar do modelo medidas que possam gerar ruído no desempenho do classificador, prejudicando-o.

Os resultados obtidos na avaliação inicial sugerem uma maior dificuldade para detectar e classificar automaticamente a presença de fronteiras não terminais. Esta maior dificuldade pode ser parcialmente justificada pela percepção humana de fronteiras não terminais ao longo do fluxo da fala, pois, os valores de Kappa obtidos mostram que o acordo para indicar a percepção de fronteiras não conclusivas, inclusive entre segmentadores humanos treinados, é um pouco menor.

3.2. Refinamento do classificador LDA

Nesta sessão, é apresentada a etapa de eliminação progressiva de medidas e a etapa de eliminação de medidas não apontadas na literatura como relevantes para o estabelecimento de fronteiras prosódicas.

3.2.1. Resultados obtidos por meio da eliminação progressiva de medidas com baixo peso no modelo

I. Fronteiras terminais

A seguir, a tabela (15) mostra a frequência bruta de unidades VVs escolhidas aleatoriamente para o refinamento do LDA e o poder de predição alcançado por ele ao longo da eliminação progressiva de medidas para as categorias presença e ausência de fronteira terminal.

Tabela 15: Poder de predição alcançado pelo LDA ao longo da eliminação progressiva de medidas – Fronteiras terminais

Medidas usadas	Frequência Presença	Frequência Ausência	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
111	38	759	76%	97,4%	24%	2,6%
101	47	758	72%	97,5%	28%	2,5%
91	47	744	74,4%	98%	25,6%	2%
81	42	758	76,1%	98,2%	23,9%	1,8%
71	50	769	62%	98,2%	38%	1,8%
66	40	750	62,5%	98%	37,5%	2%
61	40	782	65%	98,3%	35%	1,7%
56	45	753	64,4%	97,7%	35,6%	2,3%
51	55	750	67,2%	97%	32,8%	3%
46	47	782	68%	97,7%	32%	2,3%
41	46	792	58,6%	97,1%	41,4%	2,9%
36	49	798	65,3%	97,2%	34,7%	2,8%
31	47	763	59,5%	96,6%	40,5%	3,4%
26	44	783	54,5%	96,1%	45,5%	3,9%
21	48	800	60,4%	95,5%	39,6%	4,5%

Os resultados obtidos ao longo da eliminação progressiva de medidas revelam que o modelo que melhor explica e prevê a categoria presença de fronteira terminal contém 81 das 111 medidas extraídas. Ainda de acordo com estes resultados expostos acima, o modelo que melhor prediz a categoria ausência de fronteira terminal contém 61 medidas.

II. Fronteiras não terminais

Abaixo, a tabela (16) mostra a frequência bruta de unidades VVs escolhidas aleatoriamente para o refinamento do LDA e o poder de predição alcançado por ele ao longo da eliminação progressiva de medidas para as categorias presença e ausência de fronteira não terminal.

Tabela 16: Poder de predição alcançado pelo LDA ao longo da eliminação progressiva de medidas – Fronteiras não terminais

Medidas usadas	Frequência Presença	Frequência Ausência	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
111	179	718	39%	94,9%	61%	5,1%
101	191	614	39,2%	94,5%	60,8%	5,5%
91	179	612	41,3%	95%	58,7%	5%
81	174	616	39%	94,9%	61%	5,1%
71	178	641	37,6%	94,9%	62,4%	5,1%
66	184	606	35,3%	94,4%	64,7%	5,6%
61	186	636	34,9%	95,3%	65,1%	4,7%
56	185	613	31,3%	95,5%	68,7%	4,5%
51	171	597	27,4%	95,7%	72,6%	4,3%
46	181	610	30,3%	95,1%	69,7%	4,9%
41	171	637	29,2%	95,2%	70,8%	4,8%
36	189	616	28,5%	94,7%	71,5%	5,3%
31	171	592	28,6%	95,8%	71,4%	4,2%
26	180	609	31,1%	95,1%	68,9%	4,9%
21	174	636	25,2%	96,1%	74,8%	3,9%

Os resultados obtidos ao longo da eliminação progressiva de medidas revelam que o modelo que melhor explica e prevê a categoria presença de fronteira não terminal contém 91 das 111 medidas extraídas automaticamente pelo *BreakDescriptor*. Já o modelo que melhor prediz a categoria ausência de fronteira não terminal contém 21 medidas.

3.2.2. Discussão dos resultados obtidos por meio da eliminação progressiva de medidas com baixo peso

Com a finalidade de alcançar o melhor desempenho possível para detectar a presença e a ausência de fronteiras terminais e não terminais, a eliminação progressiva de medidas com base no peso atribuído pelo classificador buscava retirar as medidas menos relevantes hierarquicamente para que o LDA aprendesse a associar as diversas variáveis de previsão (numéricas e categóricas) às categorias já classificadas previamente pelos segmentadores, usando apenas as variáveis mais relevantes hierarquicamente.

Conforme observa-se nas tabelas (15) e (16), após eliminar as medidas menos relevantes hierarquicamente, o desempenho do classificador para classificar a presença de fronteiras prosódicas teve a sua eficácia reduzida. Com as 21 medidas mais relevantes, o desempenho do LDA reduziu 15,6%, pois o classificador LDA classificou 60,4% das fronteiras terminais analisadas. No caso das fronteiras não terminais, com as 21 medidas mais relevantes, o LDA reduziu 13,8% de acerto, pois, o LDA classificou corretamente 25,2% das fronteiras não terminais. Por isso, é necessário desenvolver um melhor método de refinamento do classificador LDA, distinto da simples eliminação de medidas menos relevantes hierarquicamente, para aumentar a porcentagem de acerto para a presença de fronteiras prosódicas e diminuir a porcentagem de falsos alarmes para a ausência de fronteiras.

3.2.3. Resultados obtidos por meio da eliminação de medidas não apontadas na literatura como relevantes para o estabelecimento de fronteiras prosódicas

Tendo como base a revisão de literatura da área, as medidas não indicadas como relevantes para o estabelecimento de fronteiras prosódicas foram eliminadas, retirando-se sempre uma medida por vez e buscando diversas combinações, até os modelos atingirem a maior porcentagem de acerto.

I. Fronteiras terminais

A seguir, a tabela (17) mostra a frequência bruta de unidades VVs escolhidas aleatoriamente para o refinamento do LDA-TB.

Tabela 17: Refinamento do treinamento LDA-TB – Frequência bruta de unidades VVs escolhidas aleatoriamente

Fronteira	Presença	Ausência
Terminal	45	837

Os resultados obtidos após a eliminação de 91 medidas revelam que o melhor modelo que explica e prevê as categorias presença e ausência de fronteira terminal contém 20 das 111 medidas extraídas automaticamente pelo *BreakDescriptor*. Abaixo, a tabela (18) mostra o poder de predição alcançado pelo classificador LDA-TB, usando 20 medidas.

Tabela 18: Poder de predição alcançado pelo LDA-TB para predizer as categorias presença e ausência de fronteira terminal usando 20 medidas

Fronteira	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
Terminal	80%	95,2%	20%	4,8%

Para este modelo de classificação estatística, a tabela (19) mostra todas as 20 medidas relevantes, sua hierarquia e pesos para predizer a realização de fronteiras terminais.

Tabela 19: Parâmetros relevantes para predizer as categorias presença e ausência de fronteira terminal usando o LDA e 20 medidas

Hierarquia	Sigla	Medida	Peso
1°	psdur	Duração da interrupção do sinal acústico após a unidade VV de análise	2,641517
2°	psp	Interrupção do sinal acústico após a unidade VV de análise	1,948023
3°	f0meddloc	Reset de f0	0,328589
4°	df0medr1	Aumento ou declínio de f0 na unidade VV R1	0,264072
5°	df0medl	Tendência geral de mudanças no contorno entonacional de f0 antes da unidade VV de análise	0,257113
6°	sddf0d	Aumento ou declínio da variação das mudanças no contorno entoacional de f0 (regularidade das mudanças de f0) da janela global localizada após a unidade VV de análise em relação à variação das mudanças no contorno entoacional de f0 (regularidade das mudanças de f0) da janela global localizada antes da unidade VV de análise	0,157061
7°	prd	Aumento ou declínio da taxa de saliência duracional da janela global localizada após a unidade VV de análise em relação à taxa de saliência duracional da janela global localizada antes da unidade VV de análise	0,10095
8°	sdf0l	Variação de f0 antes da unidade VV de análise	0,091317
9°	df0medl10	Aumento ou declínio de f0 na unidade VV L10	0,065812
10°	f0rl	Taxa de proeminência de f0 antes da unidade VV de análise	0,032949
11°	df0meddloc	Indica se a forma de subida ou descida de f0 entre as unidades R1 e L0 é igual ou diferente	0,032238
12°	f0medd	Aumento ou declínio da f0 da janela global localizada após a unidade VV de análise em relação à f0 da janela global localizada antes da unidade VV de análise	0,029097
13°	zl10	Alongamento ou compressão da unidade VV L10	0,027885
14°	skf0d	Mudança de predominância de valores de f0 (de graves para agudos ou vice versa) após a unidade VV de análise	0,025454
15°	mzd	Aumento ou declínio da duração normalizada e suavizada da janela global localizada após a unidade VV de análise em relação à duração normalizada e suavizada da janela global localizada antes da unidade VV de análise	0,014874
16°	skdf0d	Mudança de predominância de contorno entonacional de f0 (crescente ou decrescente) após a unidade VV de análise	0,010724
17°	SDzl	Variação no ritmo antes da unidade VV de análise	0,009925
18°	ard	Aumento ou declínio taxa de articulação da janela global localizada após a unidade VV de análise em	0,003481

		relação à janela global localizada antes da unidade VV de análise	
19°	zdlloc	Aumento ou declínio da duração normalizada e suavizada da unidade VV de análise em relação à unidade VV imediatamente seguinte	0,001426
20°	emphl	Intensidade antes da unidade VV de análise	0,000584

II. Fronteiras não terminais

A seguir, a tabela (20) mostra a frequência bruta de unidades VVs escolhidas aleatoriamente para o refinamento do LDA-NTB.

Tabela 20: Refinamento do treinamento LDA-NTB – Frequência bruta de unidades VVs escolhidas aleatoriamente

Fronteira	Presença	Ausência
Não terminal	180	595

Os resultados obtidos após a eliminação de 59 medidas revelam que o melhor modelo que explica e prevê a categoria presença de fronteira não terminal contém 52 das 111 medidas extraídas automaticamente pelo *BreakDescriptor*. Abaixo, a tabela (21) mostra o poder de predição alcançado pelo classificador LDA-TB, usando 52 medidas.

Tabela 21: Poder de predição alcançado pelo LDA-NTB para predizer as categorias presença e ausência de fronteira não terminal usando 52 medidas

Fronteira	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
Não terminal	41,6%	95%	58,4%	5%

Para este modelo de classificação estatística de fronteiras não conclusivas, a tabela (22) mostra todas as 52 medidas relevantes hierarquicamente para prever a realização de fronteiras não terminais.

Tabela 22: Parâmetros relevantes para prever a categoria presença de fronteira não terminal usando o LDA e 52 medidas

Hierarquia	Sigla	Medidas	Pesos
1°	srd	Aumento ou declínio da taxa de elocução da janela global localizada após a unidade VV de análise em relação à taxa de elocução da janela global localizada antes da unidade VV de análise	35,4497
2°	srl	Taxa de elocução antes da unidade VV de análise	35,32849
3°	srr	Taxa de elocução após a unidade VV de análise	35,30309
4°	skdf0l	Predominância de contorno entonacional de f0 crescente ou decrescente antes da unidade VV de análise	23,75122
5°	skf0l	Predominância de valores de f0 mais graves ou mais agudos antes da unidade VV de análise	23,59795
6°	skf0d	Mudança de predominância de valores de f0 (graves ou agudos) após a unidade VV de análise	23,57102
7°	skdf0d	Mudança de predominância de contorno entonacional de f0 (crescente ou decrescente) após a unidade VV de análise	23,56964
8°	psdur	Duração da interrupção do sinal acústico após a unidade VV de análise	4,499951
9°	psp	Interrupção do sinal acústico após a unidade VV de análise	4,380292
10°	skzl	Predominância de alongamentos ou encurtamentos em relação à duração intrínseca dos fones que compõem a janela antes da unidade VV de análise	3,435028
11°	skzd	Mudança de predominância de alongamentos ou encurtamentos entre as janelas à esquerda e à direita da unidade VV de análise	3,309259
12°	skzr	Predominância de alongamentos ou encurtamentos em relação à duração intrínseca dos fones que compõem a janela após a unidade VV de análise	3,25352
13°	df0medr1	Aumento ou declínio de f0 da unidade VV R1 em relação à f0 da unidade VV L0	1,291951
14°	df0medl10	Aumento ou declínio de f0 da unidade VV L10 em relação à f0 da unidade VV L9	0,698202
15°	df0med0	Aumento ou declínio de f0 da unidade VV VV de análise (L0) em relação à f0 da unidade VV L10	0,687864
16°	df0medr3	Aumento ou declínio de f0 da unidade VV R3 em relação à f0 da unidade VV R2	0,555977
17°	df0medl9	Aumento ou declínio de f0 da unidade VV L9 em relação à f0 da unidade VV L8	0,501705

18°	f0medr1	F0 da unidade VV R1	0,501575
19°	df0medl7	Aumento ou declínio de f0 da unidade VV L7 em relação à f0 da unidade VV L6	0,495478
20°	f0medr3	F0 da unidade VV R3	0,491169
21°	df0medr	Tendência geral de mudanças no contorno entonacional de f0 após a unidade VV de análise	0,47987
22°	f0medl6	F0 da unidade VV L6	0,429624
23°	f0meddloc	<i>Reset</i> de f0	0,382925
24°	f0medl8	F0 da unidade VV L8	0,31627
25°	f0medl7	F0 da unidade VV L7	0,302669
26°	df0medr2	Aumento ou declínio de f0 da unidade VV R2 em relação à f0 da unidade VV R1	0,294903
27°	f0medl9	F0 da unidade VV L9	0,223128
28°	ard	Aumento ou declínio taxa de articulação da janela global localizada após a unidade VV de análise em relação à taxa de articulação da janela global localizada antes da unidade VV de análise	0,208304
29°	df0medd	Diferença entre médias de declínio ou aumento de f0 antes e depois da posição de análise	0,207589
30°	f0medr2	F0 da unidade VV R2	0,197841
31°	pr	Taxa de saliência duracional após a unidade VV de análise	0,197841
32°	df0medl8	Aumento ou declínio de f0 da unidade VV L8 em relação à f0 da unidade VV L7	0,171194
33°	df0meddloc	Indica se a forma de subida ou descida de f0 entre as unidades R1 e L0 é igual ou diferente	0,158073
34°	zr1	Alongamento ou compressão da unidade VV R1	0,14905
35°	zdlloc	Aumento ou declínio da duração normalizada e suavizada da unidade VV de análise em relação à unidade VV imediatamente seguinte	0,140341
36°	arr	Taxa de articulação após a unidade VV de análise	0,139081
37°	sdf0d	Aumento ou declínio da variação de f0 da janela global localizada após a unidade VV de análise em relação à janela global localizada antes da unidade VV de análise	0,131689
38°	zl10	Alongamento ou compressão da unidade VV L10	0,121361
39°	zl9	Alongamento ou compressão da unidade VV L9	0,102132
40°	emphdloc	Aumento ou declínio de intensidade da unidade VV de análise em relação à unidade VV imediatamente seguinte	0,097749
41°	f0rd	Aumento ou declínio da taxa de proeminência da janela global localizada após a unidade VV de análise em relação à janela global localizada antes da unidade VV de análise	0,091367
42°	f0medl10	F0 da unidade VV L10	0,08759
43°	f0medr4	F0 da unidade VV R4	0,076148
44°	zl7	Alongamento ou compressão da unidade VV L7	0,069133
45°	sddf0d	Aumento ou declínio da variação das mudanças no contorno entoacional de f0 (regularidade das mudanças de f0) da janela global localizada após a unidade VV de análise em relação à variação das mudanças no contorno	0,066877

		entoacional de f0 (regularidade das mudanças de f0) da janela global localizada antes da unidade VV de análise	
46°	zl6	Alongamento ou compressão da unidade VV L6	0,060545
47°	zr2	Alongamento ou compressão da unidade VV R2	0,054255
48°	f0rl	Taxa de proeminência de f0 antes da unidade VV de análise	0,049216
49	mzd	Aumento ou declínio da duração normalizada e suavizada da janela global localizada após a unidade VV de análise em relação à duração normalizada e suavizada da janela global localizada antes da unidade VV de análise	0,031199
50°	SDzd	Aumento ou declínio da variação do ritmo da janela global localizada após a unidade VV de análise em relação à janela global localizada antes da unidade VV analisada	0,017637
51°	zl8	Alongamento ou compressão da unidade VV L8	0,016147
52°	f0medr5	F0 média da unidade VV R5	0,007772

3.2.4. Discussão dos resultados obtidos por meio da eliminação de medidas não apontadas na literatura como relevantes para o estabelecimento de fronteiras prosódicas

O modelo mostrou uma concentração de erros para detectar automaticamente os casos de fronteiras não conclusivas localizadas após *retractings*, tomadas de tempo e “né”. Observou-se que 72,5% das fronteiras não terminais, que estavam diante de *retractings*, tomadas de tempo e “né”, não foram detectadas de forma automática pelo modelo. Uma hipótese que justifica a alta porcentagem de erros para detectar fronteiras após *retractings* e tomadas de tempo é a falta de planejamento destes tipos de fronteiras. Possivelmente, elas são fronteiras involuntárias e menos padronizadas no que diz respeito às configurações de parâmetros fonético-acústicos. Para as fronteiras não terminais localizadas antes ou após a interjeição “né”, é necessário desenvolver uma posterior investigação acerca das causas dos erros obtidos.

Esta fase de refinamento do LDA evidencia que é necessário aperfeiçoar a performance do modelo de classificação automática de fronteiras prosódicas, sobretudo a classificação das fronteiras não terminais. Como o modelo alcançou 80% de acerto para detectar fronteiras terminais e não melhorou o desempenho para classificar fronteiras não conclusivas, priorizou-se a busca pelo aumento da capacidade preditiva do modelo para

classificar as fronteiras não terminais. De fato, há outros aspectos sobre o modelo de classificação de fronteiras terminais que podem ser explorados. Porém, como as seções 3.3 e 3.5 mostram, optou-se por refinar o modelo de classificação de fronteiras não terminais de outras formas.

3.3. Refinamento do classificador de fronteiras não terminais

O conjunto de dados utilizado no refinamento sem inclusão dos casos especiais de fronteiras não terminais (*retracting*, “né” e tomada de tempo) compreende um total de 295 fronteiras, das quais 225 são não terminais e 70 terminais. Nesta terceira etapa de treinamento, as medidas não apontadas na literatura como relevantes para o estabelecimento de fronteiras prosódicas foram eliminadas, retirando-se sempre uma medida por vez, até o modelo de classificação de fronteiras não terminais atingir a maior porcentagem de acerto para classificar as categorias presença e ausência de fronteira não terminal.

A seguir, a tabela (23) mostra a frequência bruta de unidades VVs escolhidas aleatoriamente para o refinamento do LDA-NTB desenvolvido nesta fase.

Tabela 23: Refinamento do treinamento LDA-NTB – Frequência bruta de unidades VVs escolhidas aleatoriamente

Fronteira	Presença	Ausência
Não terminal	161	721

Os resultados obtidos revelam que o modelo que melhor explica e prevê as categorias analisadas contém 9 medidas. Abaixo, a tabela (24) mostra o poder de predição alcançado por este modelo.

Tabela 24: Poder de predição alcançado pelo LDA-NTB para prever as categorias presença e ausência de fronteira não terminal usando 9 medidas

Fronteira	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
Não terminal	37,2%	95,1%	62,8%	4,9%

O classificador LDA-NTB acertou 37,2% das ocorrências de fronteiras não terminais. Como consequência disso, 62,8% das fronteiras não terminais não foram identificadas automaticamente pelo modelo. O modelo apresentou uma convergência de 95,1% com os pontos em que os segmentadores indicaram a ausência de fronteira não conclusiva. Com isso, o modelo LDA-NTB obteve 4,9% de falsos alarmes.

Para este modelo de classificação estatística que contém 9 medidas, a tabela (25) mostra as medidas fonético-acústicas que se mostraram relevantes hierarquicamente para prever a realização de fronteiras não terminais.

Tabela 25: Parâmetros relevantes para prever as categorias presença e ausência de fronteira não terminal, excluindo casos especiais de fronteiras não terminais, usando o LDA e 9 medidas

Hierarquia	Sigla	Fenômeno fonético	Peso
1°	zl0	Alongamento ou compressão da unidade VV de análise	24,29213
2°	zr1	Alongamento ou compressão da unidade VV R1	24,18751
3°	zdloc	Aumento ou declínio da duração normalizada e suavizada da unidade VV de análise em relação à duração normalizada e suavizada da unidade VV imediatamente seguinte	24,04666
4°	psp	Interrupção do sinal acústico após a unidade VV de análise	4,013554
5°	psdur	Duração da interrupção do sinal acústico após a unidade VV de análise	2,842826
6°	ard	Aumento ou declínio taxa de articulação da janela global localizada após a unidade VV de análise em relação à taxa de articulação da janela global localizada antes da unidade VV de análise	0,167689
7°	sdf0d	Aumento ou declínio da variação de f0 da janela global localizada após a unidade VV de análise em relação à variação de f0 da janela global localizada antes da unidade VV de análise	0,153763
8°	zl10	Alongamento ou compressão da unidade VV L10	0,137096
9°	srd	Aumento ou declínio da taxa de elocução da janela global localizada após a unidade VV de análise em	0,092402

relação à taxa de elocução da janela global localizada antes da unidade VV de análise

3.4. Teste no subconjunto de dados

O teste no subconjunto correspondente a 30% dos dados não treinados foi realizado para dois dos modelos de classificação automática de fronteiras: O modelo LDA-TB que usou 20 medidas e alcançou 80% de acerto e o modelo LDA-NTB obtido após a eliminação dos casos especiais de fronteiras e que usou 9 medidas, alcançando 37,2% de acerto. A tabela (26) mostra a frequência de unidades VVs usadas nos testes finais dos modelos.

Tabela 26: Teste dos modelos LDA-TB e LDA-NTB – Frequência bruta de unidades VVs escolhidas aleatoriamente

Fronteira	Presença	Ausência
Terminal	25	319
Não terminal	64	277

Abaixo, a tabela (27) mostra os resultados obtidos no teste.

Tabela 27: Poder de predição alcançado pelo teste final dos modelos LDA-TB e LDA-NTB para predizer as categorias presença e ausência de fronteiras

Fronteira	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
Terminal	80%	92%	20%	8%
Não terminal	50%	92,8%	50%	7,2%

Os resultados do teste final mostram que o acerto do modelo para a categoria presença de fronteira terminal é igual ao acerto do treinamento (80%). A porcentagem de falsos

alarmes, contudo, não é semelhante. No teste, a porcentagem de falsos alarmes aumentou 3,2%, totalizando 8%.

Os resultados do teste final indicam que a porcentagem de acerto para identificar fronteiras não terminais é superior à porcentagem de acerto alcançado no treinamento. O teste do modelo final de fronteiras não terminais acertou 50% das ocorrências de fronteiras não terminais e 92,8% das ocorrências de ausência de fronteira não terminal. Deste modo, os acertos para detectar fronteiras não terminais aumentaram 12,8%, já a porcentagem de falsos alarmes aumentou 2,3%.

O aumento de 12,8% para classificar fronteiras não terminais pode gerar discussões sobre a confiabilidade do modelo. Este aumento pode ser justificado pelo fato de que os dados foram coletados em duas partes disjuntas: uma destinada ao treinamento e outra ao teste. Especialmente nas fronteiras não terminais que apresentam uma maior variação em termos de configuração de parâmetros fonético-acústicos, um aumento como esse pode ser atribuído ao fato de que as fronteiras não terminais submetidas ao teste final apresentaram parâmetros mais equivalentes aos parâmetros do modelo.

3.5. Classificação de subtipos de fronteiras não terminais

Como já foi dito, as categorias usadas nas análises são a presença e ausência de fronteira terminal e não terminal. A categoria ausência de fronteira terminal engloba posições às quais os segmentadores não atribuíram nenhuma fronteira prosódica, como também engloba as posições de fronteira não terminal. Semelhantemente, a categoria ausência de fronteira não terminal inclui posições de nenhuma fronteira prosódica, mas também inclui posições de fronteiras terminais.

Para investigar a hipótese de que há subgrupos de fronteiras não terminais, optou-se por utilizar todas as 225 ocorrências de unidades VVs correspondentes às fronteiras não terminais e uma parte das unidades VVs correspondentes a nenhuma fronteira prosódica. Isto é, ao invés de usar 70% das unidades para o treinamento e 30% para o teste final, utilizou-se 100% das unidades VVs correspondentes às fronteiras não terminais e 160 unidades VVs correspondentes a nenhuma fronteira prosódica. Portanto, o classificador LDA realizou apenas

a etapa de treinamento e usou as categorias presença de fronteira não terminal e ausência de fronteira não terminal.

Abaixo, a tabela (28) mostra as unidades VVs usadas.

Tabela 28: Classificação de subtipos de fronteiras não terminais - Frequência bruta de unidades VVs avaliadas

Fronteira	Presença	Ausência
Não terminal	225	160

A fase que investigou a classificação estatística de subtipos de fronteiras não terminais adotou somente o modelo LDA-NTB sem inclusão de casos especiais de fronteiras, usando 9 das 111 medidas extraídas pelo *script*, porque este modelo incluiu o menor número de variáveis preditivas e alcançou 50% de acerto para a categoria presença de fronteira não terminal no teste final do modelo. Nesta fase, este modelo LDA-NTB passará a ser chamado de Modelo 1.

As unidades VVs correspondentes a presença e a ausência de fronteira não terminal, descritas na tabela (28), foram submetidas ao Modelo 1 para identificar os acertos do modelo nos dados adotados nesta fase. O Modelo 1 identificou 152 das 225 fronteiras não terminais analisadas. Em relação à ausência de fronteira não terminal, o Modelo 1 classificou corretamente 125 das 160 unidades VVs correspondentes a ausência de fronteira não terminal. Abaixo, a tabela (29) mostra o poder de predição alcançado pelo Modelo 1 para o conjunto de dados adotado.

Tabela 29: Poder de predição alcançado pelo Modelo 1 para prever as categorias presença e ausência de fronteira não terminal

Fronteira	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
Não terminal	68%	78%	32%	22%

Os resultados obtidos mostram que o Modelo 1 aumenta muito o poder de predição a partir do momento em que são excluídas as posições de fronteira terminal. Uma hipótese que pode justificar esse resultado é alguma semelhança entre fronteiras terminais e não terminais no que diz respeito às configurações de parâmetros fonético-acústicos. Quando as posições de fronteira terminal são retiradas, o Modelo 1 possivelmente não precisou constatar as diferenças e as semelhanças de parâmetros entre fronteira terminal e não terminal. Por isso, ele não se confundiu e foi capaz de detectar melhor fronteiras não terminais. Optamos por adotar esse procedimento para entender melhor em quantos tipos podem ser subdivididos as fronteiras não terminais, deixando para outro momento a solução do problema de como evitar a confusão para diferenciar fronteiras terminais de não terminais.

Com as 73 fronteiras não terminais não reconhecidas pelo Modelo 1, desenvolveu-se um novo processo de treinamento. Considerando a revisão de literatura da área, as medidas não relevantes para o estabelecimento de fronteiras prosódicas foram excluídas. O treinamento começou com 111 medidas e retirou uma medida por vez até o modelo atingir a maior porcentagem de acerto para classificar a presença de fronteira não terminal. Este novo processo de treinamento indicou um segundo modelo (Modelo 2) para reconhecer as fronteiras não terminais não reconhecidas anteriormente pelo Modelo 1. No total, o Modelo 2 contém 10 medidas.

O Modelo 2 classificou corretamente 57 das 73 fronteiras não terminais não reconhecidas pelo Modelo 1. Em relação à ausência de fronteira não terminal, o Modelo 2 classificou corretamente 128 das 160 unidades VVs correspondentes a ausência de fronteira não terminal. A seguir, a tabela (30) indica o poder de predição alcançado pelo Modelo 2.

Tabela 30: Poder de predição alcançado pelo Modelo 2 para prever as categorias presença e ausência de fronteira não terminal

Fronteira	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
Não terminal	78%	80%	22%	20%

Com as 16 fronteiras não terminais não reconhecidas pelos Modelos 1 e 2, desenvolveu-se outro processo de treinamento. O treinamento começou com 111 medidas e retirou uma medida por vez até o modelo atingir a maior porcentagem de acerto para classificar

fronteiras não terminais. Os resultados deste treinamento geraram um terceiro modelo (Modelo 3) para reconhecer as fronteiras não terminais não detectadas pelos Modelos 1 e 2. No total, o Modelo 3 contém 8 medidas.

O Modelo 3 classificou corretamente 11 das 16 fronteiras não terminais não reconhecidas anteriormente pelos Modelos 1 e 2. Em relação à ausência de fronteira não terminal, o Modelo 3 classificou corretamente 140 das 160 unidades VVs correspondentes a ausência de fronteira não terminal. A seguir, a tabela (31) indica o poder de predição alcançado pelo Modelo 3.

Tabela 31: Poder de predição alcançado pelo Modelo 3 para prever as categorias presença e ausência de fronteira não terminal

Fronteira	Acertos Presença	Acertos Ausência	Erros Presença	Erros Ausência
Não terminal	69%	87,5%	31%	12,5%

Os resultados mostram que os três modelos juntos são capazes de aumentar a convergência do classificador estatístico em relação à categoria presença de fronteira não terminal. Os poderes de predição alcançados mostram que os três modelos juntos classificaram 220 das 225 fronteiras não terminais. Esses valores evidenciam que a adoção de três modelos de classificação diferentes classifica corretamente 220 das 225 fronteiras não terminais. Deste modo, os três modelos LDA-NTB capturam de forma mais fina e gradiente as diferenças entre fronteiras não terminais, corroborando assim a hipótese de subtipos de fronteiras não conclusivas caracterizados por medidas diferentes.

A tabela (25), localizada na seção 3.3, mostra as medidas usadas para prever o estabelecimento de fronteiras não terminais, adotando o Modelo 1. Abaixo, a tabela (32) mostra as medidas usadas para prever o estabelecimento de fronteiras não terminais não classificadas pelo Modelo 1.

Tabela 32: Modelo 2 – Medidas relevantes para a classificação de fronteiras não terminais

Hierarquia	Sigla	Fenômeno fonético	Peso
1°	srl	Taxa de elocução antes da unidade VV de análise	0,717276
2°	sddf0l	Variação das mudanças no contorno entoacional de f0 (regularidade das mudanças de f0) antes da unidade VV de análise	0,625759
3°	sdf0l	Variação de f0 antes da unidade VV de análise	0,467596
4°	ard	Aumento ou declínio da taxa de articulação da janela global localizada após a unidade VV de análise em relação à taxa de articulação da janela global localizada antes da unidade VV de análise	0,45009
5°	f0medl	F0 média antes da unidade VV de análise	0,374038
6°	f0rd	Aumento ou declínio da taxa de proeminência de f0 da janela global localizada após a unidade VV de análise em relação à taxa de proeminência de f0 da janela global localizada antes da unidade VV de análise	0,205596
7°	f0meddloc	Reset de f0	0,098721
8°	f0med0	F0 da unidade VV de análise	0,087294
9°	f0medr1	F0 da unidade VV R1	0,050374
10°	emphl	Intensidade antes da unidade VV de análise	0,012798

A tabela (33) mostra as medidas usadas para predizer o estabelecimento de fronteiras não terminais por meio do Modelo 3.

Tabela 33: Modelo 3 – Medidas relevantes para a classificação de fronteiras não terminais

Hierarquia	Sigla	Medida	Peso
1°	prl	Taxa de saliência duracional antes da unidade VV de análise	151,55576
2°	prd	Aumento ou declínio da taxa de saliência duracional da janela global localizada após a unidade VV de análise em relação à taxa de saliência duracional da janela global localizada antes da unidade VV de análise	150,5872
3°	prr	Taxa de saliência duracional após a unidade VV de análise	149,51503
4°	sdf0r	Variação de f0 após a unidade VV de análise	0,5200116
5°	SDzl	Variação no ritmo antes da unidade VV de análise	0,3132163
6°	df0medr1	Aumento ou declínio de f0 da unidade VV R1 em relação à f0 da unidade VV L0	0,2985421
7°	df0medl10	Aumento ou declínio de f0 da unidade VV L10 em relação à f0 da unidade VV L9	0,1811481

8°	df0meddloc	Indica se a forma de subida ou descida de f0 entre as unidades R1 e L0 é igual ou diferente	0,1425396
----	------------	---	-----------

3.6. Análise de *cluster* hierárquica de fronteiras não terminais

A fim de se investigar melhor subtipos de fronteiras não terminais e possibilitar uma melhor classificação automática deste macrotipo de fronteira prosódica, os Modelos 1, 2 e 3 foram submetidos a uma análise de *cluster* hierárquica. A análise de *cluster* hierárquica é um tipo de técnica estatística utilizada para explorar similaridades entre variáveis, agrupando objetos semelhantes e separando objetos distintos. O agrupamento é realizado levando em consideração as variáveis que apresentam valores próximos entre si.

A análise de *cluster* hierárquica aqui desenvolvida foi utilizada com o objetivo de detectar subtipos de fronteiras não terminais nos Modelos 1, 2 e 3, de acordo com as variáveis de previsão usadas, ou seja, de acordo com os parâmetros fonético-acústicos extraídos automaticamente pelo *BreakDescriptor*. Os *clusters* foram obtidos por meio do método hierárquico aglomerativo, disponível no ambiente R – função `hclust()`. Essa função pertence ao pacote básico incluído no núcleo do R. A seguir, as figuras mostram os *clusters* obtidos para os Modelos 1, 2 e 3.

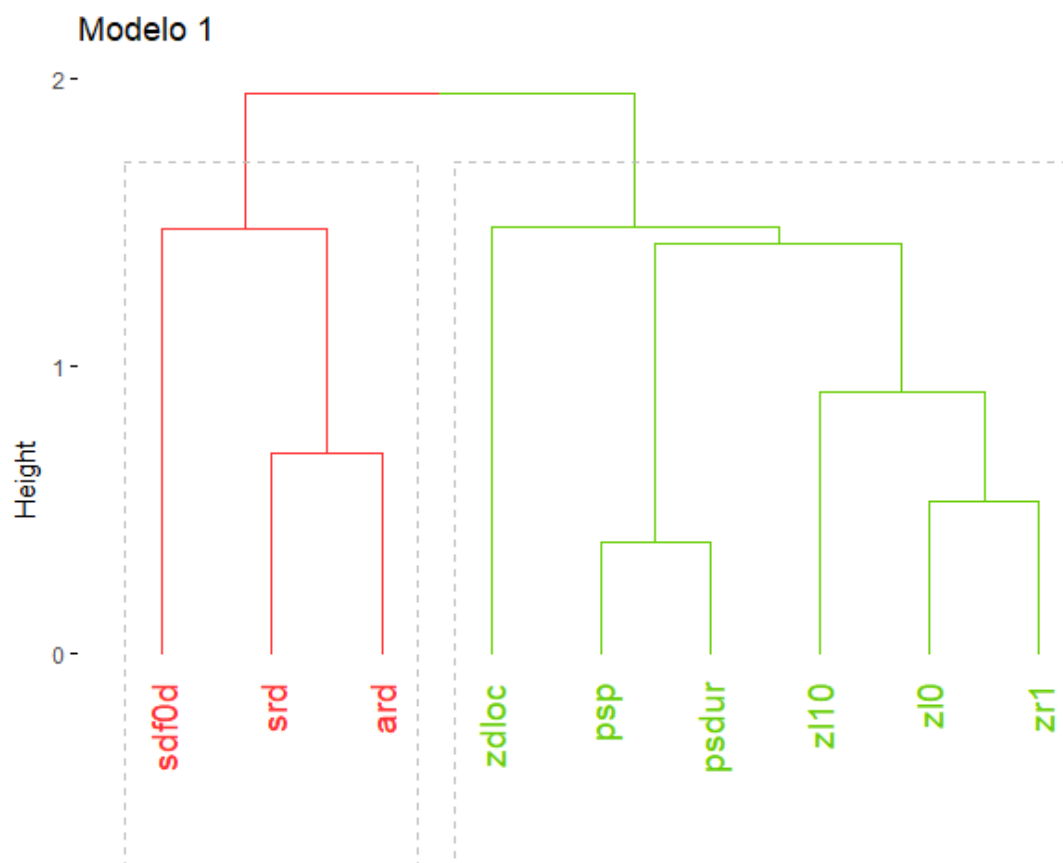


Figura 39: Clusters do Modelo 1

Legenda – Modelo 1

sdf0d – Mudança de variação de f0 (7º / peso 0,153763)

srd – Mudança de taxa de elocução (9º / peso 0,092402)

ard – Mudança de taxa de articulação (6º / peso 0,167689)

zdloc – Aumento ou declínio da duração normalizada e suavizada da unidade VV de análise em relação à duração normalizada e suavizada da unidade VV imediatamente seguinte (3º / peso 24,04666)

Legenda – Modelo 1

psp – Interrupção do sinal acústico (4º / peso 4,013554)

psdur – Duração da interrupção sinal acústico (5º / peso 2,842826)

zl0 – Alongamento na unidade VV correspondente à fronteira (1º / peso 24,29213)

zl10 – Alongamento na unidade VV imediatamente antes da fronteira (8º / peso 0,137096)

zr1 – Alongamento na unidade VV imediatamente após a fronteira (2º / peso 0,137096)

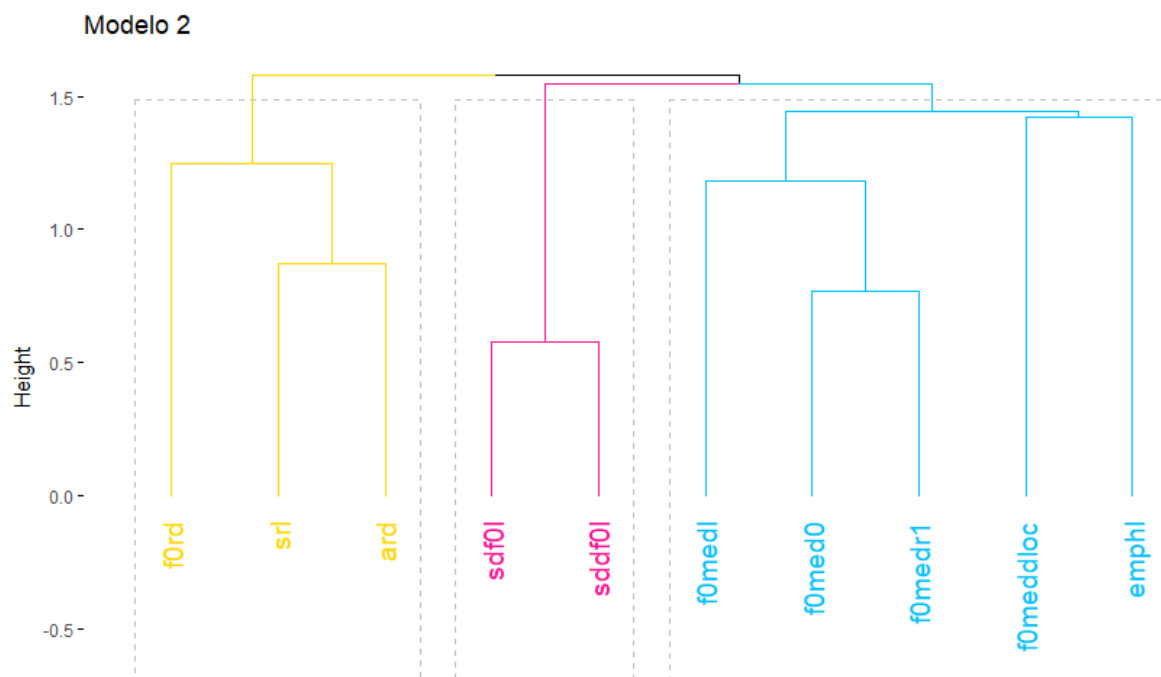


Figura 40: Clusters do Modelo 2

Legenda – Modelo 2

f0rd – Mudança de taxa de proeminência de f0 (6º/ peso 0,205596)

srl – Taxa de elocução antes da fronteira (1º/ peso 0,717276)

ard – Mudança de taxa articulação (4º/ peso 0,45009)

sdf0l – Variação de f0 antes da fronteira (3º/ peso 0,467596)

sddf0l – Variação das mudanças de f0 (regularidade das mudanças de f0) antes da fronteira (2º / peso 0,625759)

Legenda – Modelo 2

f0medl – F0 média antes da fronteira (5º/ peso 0,374038)

f0med0 – F0 na unidade VV correspondente à fronteira (8º/ peso 0,087294)

f0medr1 – F0 na unidade VV imediatamente após a fronteira (9º / peso 0,050374)

f0meddloc – *Reset* de f0 (7º/ peso 0,098721)

emphl – Intensidade antes da fronteira (10º/ peso 0,012798)

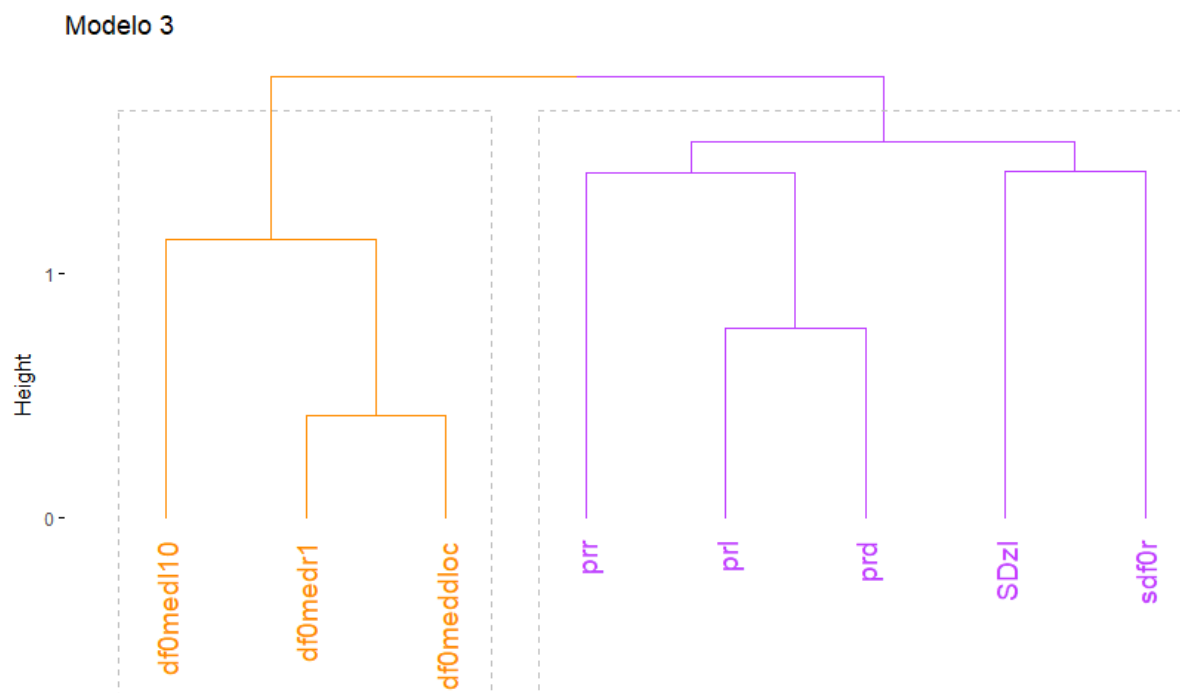


Figura 41: Clusters do Modelo 3

Legenda – Modelo 3

df0medl10 – Mudança no contorno entonacional de f0 na unidade VV imediatamente antes da fronteira (7º/ peso 0,1811481)

df0medr1 – Mudança no contorno entonacional de f0 na unidade VV imediatamente após a fronteira (6º/ peso 0,2985421)

df0meddloc – A mudança da forma de subida ou descida de f0 entre as unidades R1 e L0 (8º / peso 0,1425396)

Legenda – Modelo 3

prr – Taxa de saliência duracional após a fronteira (3º/ peso 149,51503)

prl – Taxa de saliência duracional antes da fronteira (1º/ peso 151,55576)

prd – Mudança de taxa saliência duracional (2º / peso 150,5872)

SDzl – Variação de ritmo antes da fronteira (5º/ peso 0,3132163)

sdf0r – Variação de f0 após a fronteira (4º / peso 0,5200116)

Por meio da análise dos *clusters*, observa-se que os Modelos 1, 2 e 3 apresentam cada um deles dois agrupamentos principais. Cada um dos modelos obtidos também é dividido em subgrupos menores.

4. Discussão sobre os modelos finais de classificação de fronteiras prosódicas

I. Fronteiras terminais

O modelo final de classificação de fronteiras prosódicas terminal indica que são necessárias 20 medidas para classificar automaticamente este tipo de fronteira prosódica. Os três modelos finais obtidos para classificar automaticamente fronteiras não terminais indicam que é necessário mais de um modelo e, na soma dos três, um maior número de variáveis preditivas para classificar fronteiras tradicionalmente indicadas pelos segmentadores como não conclusivas. No total, os três modelos de classificação de fronteiras não terminais utilizam 26 medidas extraídas pelo *BreakDescriptor*.

A diferença entre a quantidade de variáveis preditivas usadas nos modelos para classificar fronteira terminal e não terminal é apenas seis medidas. Contudo, podemos observar outras diferenças entre os modelos decorrentes da natureza dos dois tipos de fronteiras prosódicas.

O resultado alcançado pelo modelo de classificação de fronteiras terminais sugere que a classificação automática deste tipo de fronteira pode ser mais simples, porque o modelo de classificação deste tipo de fronteira alcançou 80% de acerto para identificar a presença de fronteira terminal com o desenvolvimento de uma menor quantidade de treinamentos. A classificação automática de fronteiras não terminais parece ser uma tarefa mais complexa para ser executada com uma ferramenta computacional baseada em variáveis de previsão, porque foram necessárias várias fases de treinamento para reconhecer e diferenciar entre si as fronteiras não terminais.

Os resultados sugerem que as fronteiras terminais são marcadas principalmente pelas medidas relacionadas à pausa. Dentre as 20 medidas incluídas no modelo, a pausa e a sua duração são os parâmetros mais relevantes hierarquicamente para prever o estabelecimento de fronteira prosódica terminal, configurando-se como a pista mais robusta do modelo. As medidas de pausa silenciosa são seguidas pelo *reset* de f_0 , pela mudança no contorno entonacional de f_0 na unidade VV logo após a fronteira e pela tendência geral de mudanças no contorno entonacional de f_0 antes da fronteira (tendência geral de crescimento ou declínio de f_0).

A hierarquia do modelo mostra que o aumento ou o declínio da variação das mudanças no contorno entoacional de f_0 (regularidade das mudanças de f_0) e a mudança da taxa de produção de saliência duracional, associada ao *pitch accent* e ao correlato perceptual de vivacidade (*liveliness*), contribuem moderadamente na classificação fronteiras terminais.

Treze medidas se mostraram menos relevantes hierarquicamente e contribuíram pouco na classificação de fronteiras terminais, pois o peso atribuído a elas é bem mais baixo, se comparado ao peso das demais sete variáveis de previsão. Estas medidas são:

- A variação de f_0 antes da fronteira;
- A mudança no contorno entonacional de f_0 na unidade VV localizada antes da fronteira;
- A taxa de proeminência de f_0 antes da fronteira;
- O aumento ou declínio da f_0 da janela global localizada após a fronteira em relação à f_0 da janela global localizada antes da fronteira;
- O alongamento ou a compressão da unidade VV localizada antes da fronteira;
- A mudança de predominância de valores de f_0 (de graves para agudos ou vice-versa) após a fronteira;
- O aumento ou o declínio da duração normalizada e suavizada da janela global localizada após a fronteira em relação à duração normalizada e suavizada da janela global localizada antes da fronteira;
- A mudança de predominância de contorno entonacional de f_0 (de crescente para decrescente ou vice-versa) após a fronteira;
- A variação no ritmo antes da fronteira;
- A mudança da taxa de articulação;
- O aumento ou o declínio da duração normalizada e suavizada da unidade VV correspondente à fronteira em relação à unidade VV imediatamente após a fronteira;
- A intensidade média antes da fronteira.

II. Fronteiras não terminais

Os resultados mostram que as fronteiras não terminais são marcadas por uma das três configurações de parâmetros fonético-acústicos descritas nos Modelos 1, 2 ou 3. Com nove medidas, o Modelo 1 é aquele que explica a maior parte das fronteiras não terminais marcadas pelos segmentadores, pois a porcentagem de acerto deste modelo em relação ao total de fronteiras não conclusivas é 67,6%. O Modelo 2, com 10 medidas, explica cerca de $\frac{1}{4}$ do total de fronteiras não terminais indicadas na amostra, pois, o modelo alcançou 25,3% de acerto para detectar fronteiras não conclusivas sobre o total. O Modelo 3, com 8 medidas, classifica apenas 4,9% das fronteiras não terminais.

Aparentemente, as medidas incluídas no Modelo 1 são as medidas mais frequentemente usadas para estabelecer fronteiras não terminais ao longo do fluxo da fala em PB. As medidas incluídas nos Modelos 2 e 3, por sua vez, são medidas que caracterizam fronteiras não terminais menos recorrentes em PB.

A hierarquia indicada pelo classificador e a análise de *cluster* são muito importantes para compreender com mais profundidade os diversos tipos de fronteiras não terminais que podem ser produzidos e percebidos pelos falantes. A hierarquia indica a importância de cada medida para o modelo de classificação automática de fronteiras não conclusivas. Os *clusters* complementam os resultados fornecidos pelos modelos, indicando para o pesquisador como as medidas podem ser combinadas para estabelecer as fronteiras.

a) Modelo 1

O Modelo 1 mostra que as medidas mais relevantes hierarquicamente para prever aproximadamente 68% das fronteiras não terminais são medidas temporais relacionadas ao alongamento da unidade VV e à pausa silenciosa, como:

- O alongamento da unidade VV correspondente à fronteira não terminal (z_{l0});
- O alongamento da unidade VV imediatamente após a fronteira não terminal (z_{r1});
- A diferença entre a duração normalizada e suavizada correspondente à fronteira não terminal em relação à duração normalizada e suavizada da unidade VV imediatamente após a fronteira.

As três medidas acima são seguidas pelas duas medidas relacionadas à pausa silenciosa extraídas pelo *BreakDescriptor*. As medidas de pausa configuram-se como as medidas mais importantes hierarquicamente para classificar fronteiras terminais, mas contribuem um pouco menos na classificação da maior parte das fronteiras não terminais.

Ainda no Modelo 1, de acordo com os pesos atribuídos pelo LDA, outras medidas que contribuem, mas em menor grau, para o estabelecimento da maioria das fronteiras não terminais são a mudança da taxa de articulação, a mudança de variação de f_0 , o alongamento na unidade VV logo antes da fronteira e a mudança da taxa de elocução.

O *cluster* do Modelo 1, mostra que um dos dois principais grupos de fronteiras não terminais é caracterizado pela combinação de medidas relacionadas à taxa de elocução, de articulação e à variação de f_0 nos contextos adjacentes às fronteiras não terminais. Dentro deste grupo, verifica-se dois subgrupos menores de fronteiras: fronteiras caracterizadas pela mudança de variação de f_0 ; fronteiras caracterizadas simultaneamente pela mudança da taxa de elocução e de articulação.

O segundo grupo principal de fronteiras não terminais é caracterizado pela junção de medidas relacionadas às pausas e à duração normalizada e suavizada do segmento silábico. Dentro do segundo grupo principal de fronteiras não terminais, verifica-se outros agrupamentos menores de fronteiras. Um subgrupo é marcado pelo aumento ou declínio da duração normalizada e suavizada da unidade VV correspondente à fronteira em relação à duração normalizada e suavizada da unidade VV após a fronteira. Outro subgrupo é marcado pelas medidas de pausa e pelos alongamentos pré-fronteiriço, fronteiriço e pós-fronteiriço. Este último subgrupo ainda é dividido em três agrupamentos mais específicos: Fronteiras não terminais marcadas pela pausa silenciosa e pela sua duração; Fronteiras não terminais marcadas

pelo alongamento pré-fronteiriço; Fronteiras não terminais marcadas pelos alongamentos fronteiriço e pós-fronteiriço.

b) Modelo 2

O Modelo 2 enfatiza que as primeiras medidas mais relevantes hierarquicamente para prever uma em cada quatro fronteiras não terminais são as medidas de taxa de elocução e de variação das mudanças no contorno entonacional de f_0 (regularidade das mudanças de f_0) antes da fronteira. A variação de f_0 antes da fronteira, a f_0 média antes da fronteira, a mudança das taxas de articulação e de proeminência de f_0 contribuem regularmente com um peso médio na marcação de fronteiras não terminais não estabelecidas pelas configurações de parâmetros fonético-acústicos descritas no Modelo 1. No Modelo 2, quatro medidas contribuem em menor grau na classificação de fronteiras não terminais. Estas medidas são o *reset* de f_0 , a f_0 da unidade VV correspondente à fronteira, a f_0 da unidade imediatamente após a fronteira e a intensidade global antes da fronteira.

O *cluster* do Modelo 2 mostra que um dos principais grupos de fronteiras não terminais é caracterizado pela combinação da taxa de elocução antes da fronteira com a mudança das taxas de articulação e de proeminência de f_0 . O segundo grupo principal é caracterizado pela combinação de várias medidas como: a) medidas melódicas de variação de f_0 antes da fronteira; b) a intensidade média antes da fronteira; c) a f_0 global antes da fronteira; d) a f_0 na unidade VV correspondente à fronteira; e) a f_0 na unidade VV após a fronteira; f) o *reset* de f_0 .

O primeiro principal agrupamento do Modelo 2 pode ser subdividido em dois agrupamentos menores. Um agrupamento em que as fronteiras não terminais são marcadas exclusivamente pela mudança da taxa de proeminência de f_0 , outro em que as fronteiras são marcadas pela taxa de elocução antes da fronteira e pela mudança da taxa de articulação.

O segundo principal agrupamento do Modelo 2 pode ser subdividido primeiramente em dois agrupamentos menores de fronteiras. Um deles é marcado pela combinação da variação de f_0 com a regularidade das mudanças no contorno entonacional de f_0 , ambos antes da fronteira. O segundo é marcado pela f_0 geral antes da fronteira, pela f_0 na

unidade VV correspondente à fronteira, pela f_0 na unidade VV imediatamente após a fronteira, pelo *reset* de f_0 e pela intensidade global antes da fronteira.

Dentro do último agrupamento citado no parágrafo anterior, verifica-se ainda outros agrupamentos mais específicos. Um subgrupo marcado pelo *reset* de f_0 e pela intensidade global antes da fronteira. Um subgrupo marcado pela f_0 na unidade VV correspondente à fronteira, pela f_0 na unidade VV imediatamente após a fronteira e pela f_0 global antes da fronteira. Este último subgrupo ainda é subdividido em dois agrupamentos mais específicos: fronteiras não terminais marcadas pelo valor de f_0 global antes da fronteira; fronteiras não terminais marcadas simultaneamente pelo valor de f_0 na unidade VV correspondente à fronteira e pelo valor de f_0 na unidade VV imediatamente após a fronteira.

c) Modelo 3

O Modelo 3 mostra que as primeiras medidas mais relevantes hierarquicamente para prever em torno de 5% das fronteiras não terminais são as três medidas globais de taxa saliência duracional. Apesar de serem incluídas no Modelo 3, algumas medidas contribuem pouco na classificação automática de fronteiras não terminais. As medidas são: a) a variação de f_0 após a fronteira; b) a variação no ritmo antes da fronteira; c) a mudança no contorno entonacional de f_0 na posição imediatamente após a fronteira; d) a mudança no contorno entonacional de f_0 na posição precedente à fronteira.

O *cluster* do Modelo 3, mostra que um dos dois principais grupos de fronteiras não terminais é caracterizado pela combinação de três medidas de f_0 :

- A mudança no contorno entonacional de f_0 na unidade VV antes da fronteira;
- A mudança no contorno entonacional de f_0 na unidade VV após a fronteira;
- A mudança da forma de subida ou descida de f_0 entre as unidades R1 e L0.

Dentro deste grupo, verifica-se dois subgrupos menores de fronteiras: fronteiras caracterizadas pela mudança no contorno entonacional de f_0 na unidade VV antes da fronteira; fronteiras caracterizadas simultaneamente pela mudança no contorno entonacional de f_0 na unidade VV após a fronteira e pela mudança da forma de subida ou descida de f_0 entre as unidades R1 e L0.

O segundo grupo principal de fronteiras não terminais é caracterizado pela: a) taxa de saliência duracional antes da fronteira; b) taxa de saliência duracional após a fronteira; c) mudança da taxa de saliência duracional; c) variação de f_0 após a fronteira; d) variação de ritmo antes da fronteira. Neste segundo grupo principal, ainda é possível observar agrupamentos menores. Um deles é marcado pela variação de f_0 após a fronteira e pela variação de ritmo antes da fronteira. O outro é marcado pela taxa de saliência duracional antes da fronteira, pela taxa de saliência duracional após a fronteira e pela mudança da taxa de saliência duracional. Este último agrupamento pode ser subdividido em outros mais específicos: Um que indica o estabelecimento de fronteiras não terminais por meio da taxa de saliência duracional após a fronteira; outro que indica a marcação de fronteiras não terminais por meio da mudança da taxa de saliência duracional e pela taxa de saliência duracional antes da fronteira.

III. Parâmetros importantes para prever fronteiras prosódicas

Os modelos finais de classificação de fronteiras prosódicas indicam que treze medidas são relevantes hierarquicamente para classificar fronteiras prosódicas terminais e não terminais. As medidas são:

- A duração da interrupção do sinal acústico após fronteira (psdur);
- A interrupção do sinal acústico após a fronteira (psp);
- O alongamento na unidade VV imediatamente antes da fronteira (z110);
- O aumento ou o declínio da duração normalizada e suavizada da unidade VV de análise em relação à duração normalizada e suavizada da unidade VV imediatamente seguinte (zdloc);

- O *reset* de f_0 ($f_0meddloc$);
- A variação de f_0 antes da fronteira (sdf_0l);
- A intensidade relativa antes da fronteira ($emphl$);
- A mudança no contorno entonacional de f_0 na unidade VV imediatamente antes da fronteira ($df_0medl10$);
- A mudança da forma de subida ou descida de f_0 entre as unidades R1 e L0 ($df_0meddloc$);
- A variação de ritmo antes da fronteira ($SDz1$);
- A mudança no contorno entonacional de f_0 na unidade VV imediatamente após a fronteira (df_0medr1);
- A mudança de taxa saliência duracional (prd);
- A mudança de taxa de articulação (ard).

Estes resultados mostram que medidas relacionadas à pausa, à duração normalizada do segmento silábico, à frequência fundamental e ao ritmo são importantes para estabelecer a presença de fronteira prosódica sem distinção entre fronteira terminal e não terminal, porque elas foram relevantes hierarquicamente para a classificação automática dos dois macrotipos de fronteiras prosódicas analisadas.

Abaixo, a tabela (34) mostra a distribuição destas medidas em cada um dos quatro modelos finais, indicando como os parâmetros que parecem ser mais relevantes para a percepção da fronteira prosódica terminal apresentam-se nos modelos destinados à classificação de fronteira prosódica não terminal. Dentre elas, quatro medidas se mostraram importantes para o modelo de fronteiras terminais e para o Modelo 1 de fronteiras não terminais. Duas se mostraram relevantes para o modelo de fronteiras terminais e para o Modelo 2 de fronteiras não terminais. Cinco foram relevantes para o estabelecimento de fronteiras terminais, mas também foram relevantes para classificar fronteiras não terminais através do Modelo 3. Somente uma medida apresentou-se relevante para prever fronteiras prosódicas terminais e não terminais classificadas por meio dos Modelos 1 e 2.

Tabela 34: Medidas em comum aos modelos

Medidas	Modelo TB		Modelo NTB 1		Modelo NTB 2		Modelo NTB 3	
	Rank	Peso	Rank	Peso	Rank	Peso	Rank	Peso
psdur	1 ^a	2,641517	5 ^a	2,842826				
psp	2 ^a	1,948023	4 ^a	4,013554				
zl10	13 ^a	0,027885	8 ^a	0,137096				
zdloc	19 ^a	0,001426	3 ^a	24,04666				
f0meddloc	3 ^a	0,328589			7 ^a	0,098721		
sdf0l	8 ^a	0,091317			3 ^a	0,467596		
emphl	20 ^a	0,000584			10 ^a	0,012798		
df0medl10	9 ^a	0,065812					7 ^a	0,1811481
df0meddloc	11 ^a	0,032238					8 ^a	0,1425396
SDzl	17 ^a	0,009925					5 ^a	0,3132163
df0medr1	4 ^a	0,264072					6 ^a	0,2985421
prd	7 ^a	0,003481					2 ^a	150,5872
ard	18 ^a	0,003481	6 ^a	0,167689	4 ^a	0,45009		

A tabela (34) mostra um aspecto muito importante sobre as medidas que são importantes para os dois tipos de fronteiras prosódicas. Apesar das medidas descritas serem importantes para estabelecer fronteiras não terminais e terminais, não há correspondência entre a relevância hierárquica da medida nos modelos. As medidas de pausa, por exemplo, são as duas medidas mais importantes para assinalar fronteiras terminais, mas são a 5^a e 4^a medidas relevantes hierarquicamente no Modelo 1 de classificação de fronteiras não terminais. O alongamento pós-fronteiriço e a diferença entre a duração normalizada e suavizada correspondente à fronteira não terminal em relação à duração normalizada e suavizada da unidade VV imediatamente após a fronteira são respectivamente a 13^a e 19^a medidas mais relevantes do modelo de fronteiras terminais, mas são a 8^a e 3^a medidas do Modelo 1 destinado à classificação de fronteiras não terminais. Assim, podemos observar que as medidas são importantes para classificar simultaneamente fronteiras terminais e não terminais, mas a contribuição de cada uma delas é diferente conforme seja o modelo.

Os três modelos destinados à classificação de fronteiras não terminais evidenciam que há vários subgrupos de fronteiras não terminais marcados por diferentes configurações de parâmetros fonético-acústicos. Mesmo sendo de conhecimento geral que as fronteiras não terminais indicam a continuidade do enunciado, é possível observar que não há parâmetros em comum entre os Modelos 1, 2 e 3 destinados à classificação automática de fronteiras não

conclusivas. Estes resultados corroboram a hipótese de que as fronteiras não terminais são realizadas por meio de um sistema fonético-acústico muito refinado e gradiente.

Provavelmente, os parâmetros fonético-acústicos mais importantes para especificar o tipo de fronteira prosódica (não terminal ou terminal) são aqueles que apresentam relevância hierárquica e peso altos apenas no modelo de um dos tipos de fronteiras. O modelo final destinado à classificação de fronteiras terminais indica que sete medidas especificam a terminalidade de uma fronteira prosódica, porque elas se mostraram relevantes hierarquicamente somente para a classificação estatística deste tipo de fronteira. As medidas são:

1. A tendência geral de mudanças no contorno entonacional de f_0 antes da fronteira terminal (df_0medl);
2. O aumento ou o declínio da variação das mudanças no contorno entoacional de f_0 (regularidade das mudanças de f_0) da janela global localizada após a fronteira terminal em relação à variação das mudanças no contorno entoacional de f_0 (regularidade das mudanças de f_0) da janela global localizada antes da fronteira terminal ($sddf_0d$);
3. A taxa de proeminência de f_0 antes da fronteira terminal (f_0rl);
4. O aumento ou o declínio da f_0 da janela global localizada após a fronteira terminal em relação à f_0 da janela global localizada antes da fronteira terminal (f_0medd);
5. A mudança de predominância de valores de f_0 (de graves para agudos ou vice-versa) após a fronteira terminal (skf_0d);
6. O aumento ou o declínio da duração normalizada e suavizada da janela global localizada após a fronteira terminal em relação à duração normalizada e suavizada da janela global localizada antes da fronteira terminal (mzd);
7. A mudança de predominância de contorno entonacional de f_0 (de crescente para decrescente ou vice-versa) após a fronteira terminal ($skdf_0d$).

Dentre as sete medidas relacionadas acima, os resultados sugerem que as medidas 1 e 2 são as medidas mais relevantes hierarquicamente para estabelecer a terminalidade de fronteiras prosódicas, porque o peso atribuído a elas pelo LDA é maior e essas medidas não são relevantes para a marcação de fronteiras não terminais. A pausa e a sua duração não podem ser vistas como medidas responsáveis por estabelecer a percepção de conclusão ou continuidade da fronteira, porque elas são importantes para predizer tanto fronteiras terminais, quanto não terminais. As medidas 3, 4, 5, 6 e 7 auxiliam na especificação da terminalidade de fronteiras de forma mais reduzida.

Os três modelos destinados à classificação de fronteiras não terminais sugerem que 11 medidas podem estabelecer a continuidade de uma fronteira prosódica, porque elas se mostraram como medidas relevantes hierarquicamente para classificar apenas fronteiras não terminais. As medidas são:

8. O alongamento fronteiro (z_{l0} – Modelo NTB 1);
9. O alongamento pós-fronteiro (z_{r1} – Modelo NTB 1);
10. A taxa de elocução antes fronteira (s_{r1} – Modelo NTB 2);
11. A variação das mudanças de f_0 (regularidade das mudanças de f_0) antes da fronteira (s_{ddf0l} – Modelo NTB 2);
12. A f_0 média antes da fronteira (f_{0medl} – Modelo NTB 2);
13. A mudança da taxa de proeminência de f_0 (f_{0rd} – Modelo NTB 2);
14. A f_0 da unidade VV correspondente à fronteira (f_{0med0} – Modelo NTB 2);
15. A f_0 da unidade VV após a fronteira (f_{0medr1} – Modelo NTB 2);
16. A taxa de saliência duracional antes da fronteira (p_{r1} – Modelo NTB 3);
17. A taxa de saliência duracional após a fronteira (p_{rr} – Modelo NTB 3);
18. A variação de f_0 após a fronteira (s_{df0r} – Modelo NTB 3).

Dentre as onze medidas relacionadas acima, os resultados sugerem que as medidas 8 e 9 são as medidas mais relevante hierarquicamente para especificar fronteiras prosódicas não terminais classificadas por meio do Modelo NTB 1, pois, o peso atribuído a cada uma delas pelo LDA é maior e elas não são relevantes para estabelecer fronteiras terminais. Os resultados também indicam que as medidas 10 e 11 são as duas mais relevantes para especificar fronteiras

não terminais classificadas por meio do Modelo NTB 2. As medidas 16 e 17 parecem ser as medidas mais importantes para especificar fronteiras não terminais classificadas por meio do Modelo NTB 3.

IV. Diferença entre os parâmetros para estabelecer fronteiras terminais e não terminais

Verificou-se que as medidas responsáveis por estabelecer a terminalidade ou a continuidade de uma fronteira prosódica são diferentes. As medidas de pausa e f_0 parecem ser mais importantes para classificar fronteiras terminais. As medidas de *speech rate* e de duração normalizada e suavizada do segmento silábico parecem ser muito importantes para classificar fronteiras não terminais. As medidas de intensidade, por sua vez, se mostraram pouco relevantes para classificar fronteiras terminais e não terminais.

a) Medidas de pausa

As medidas de pausa silenciosa configuram-se como as medidas mais importantes para assinalar fronteiras terminais, mas contribuem também na classificação de aproximadamente 68% das fronteiras não terminais, porque são a 4^a e 5^a medidas mais relevantes hierarquicamente no Modelo 1. Em aproximadamente 32% das fronteiras não terminais, a pausa e a sua duração não são importantes para assinalar fronteiras.

b) Medidas de duração normalizada e suavizada do segmento silábico

Em geral, os resultados indicam que as medidas de duração normalizada e suavizada do segmento silábico configuram-se como medidas muito relevantes para predizer

principalmente a realização de fronteiras não terminais. O Modelo 1 destinado à classificação de fronteiras não conclusivas mostra que o alongamento pré-fronteiriço, fronteiriço, pós-fronteiriço e a diferença entre a duração normalizada e suavizada correspondente à fronteira não terminal em relação à duração normalizada e suavizada da unidade VV imediatamente após a fronteira são variáveis muito importantes para classificar a maior parte das fronteiras não terminais. Apesar de identificar somente 5% das fronteiras não terminais, o Modelo 3 mostra que as taxas de saliência duracional e a variação da duração normalizada e suavizada antes da fronteira contribuem bastante na classificação estatística da menor parcela das fronteiras não terminais.

Os resultados mostram que as medidas de duração normalizada e suavizada do segmento não contribuem na classificação de aproximadamente 25% das fronteiras não terminais classificadas com auxílio do Modelo 2, porque neste modelo as medidas de duração contribuem com pesos muito baixos, sempre menores que 0,1.

c) Medidas de frequência fundamental

Os resultados sugerem que as medidas de frequência fundamental desempenham um papel muito importante no modelo de classificação de fronteiras terminais. Depois das medidas de pausa, o *reset* de f_0 , a mudança no contorno entonacional de f_0 na unidade VV após a fronteira, a tendência geral de mudanças no contorno entonacional de f_0 antes da fronteira e o aumento ou o declínio da variação das mudanças no contorno entonacional de f_0 (regularidade das mudanças de f_0) são as medidas mais relevantes hierarquicamente para prever fronteiras conclusivas.

Com exceção do Modelo 2, nos modelos destinados à classificação de fronteiras não terminais, as medidas de frequência fundamental têm uma importância reduzida. O Modelo 1 classifica 68% das fronteiras não terminais atribuindo um baixo peso à mudança de variação de f_0 .

O Modelo 2 é o único modelo destinado à classificação de fronteiras não terminais que requer uma grande gama de medidas de f_0 . No Modelo 2, são incluídas sete medidas de f_0 . Duas são muito relevantes hierarquicamente e cinco apresentam pouca contribuição na classificação estatística. A variação de f_0 e a regularidade das mudanças no contorno

entonacional de f_0 antes da fronteira são as duas medidas relacionadas à f_0 que se mostraram importantes. O valor geral de f_0 antes da fronteira, a mudança da taxa de proeminência de f_0 , o *reset* de f_0 , o valor de f_0 na unidade VV correspondente à fronteira e o valor de f_0 na unidade VV localizada logo após a fronteira são as medidas relacionadas à f_0 que contribuem pouco na classificação de fronteiras não terminais classificadas com auxílio do Modelo 2.

No Modelo 3, a variação de f_0 após a fronteira, a mudança no contorno entonacional de f_0 na unidade VV após a fronteira e a mudança no contorno entonacional de f_0 na unidade VV antes da fronteira também contribuem pouco na marcação de fronteiras, porque no Modelo 3 as medidas de taxa de saliência duracional são as mais relevantes hierarquicamente.

d) Medidas de *speech rate* e ritmo

Os resultados indicam que as medidas de *speech rate* e ritmo contribuem bastante na classificação estatística de fronteiras não terminais identificadas com auxílio dos Modelos 1 e 2, porque a mudança da taxa de articulação, a mudança da taxa de elocução e a taxa de elocução antes da fronteira são muito relevantes hierarquicamente nos dois modelos em questão. Apesar de ser incluída no modelo de classificação de fronteiras terminais, a mudança da taxa de articulação contribui pouco na classificação deste tipo de fronteira, pois o peso atribuído a ela é baixo. Nenhuma medida de *speech rate* e ritmo contribui na classificação de fronteiras não terminais identificadas por meio do Modelo 3.

e) Medidas de intensidade

Os resultados sugerem que as medidas de intensidade contribuem pouco na classificação estatística das fronteiras terminais e não terminais. Nos Modelos 1 e 3 destinados à classificação de fronteiras não terminais, nenhum parâmetro de intensidade é relevante hierarquicamente. No Modelo 2, a intensidade global média antes da fronteira não terminal contribui com o menor peso do modelo. No modelo destinado à classificação de fronteiras terminais, a intensidade global média antes da fronteira também contribui com o menor peso

do modelo em questão. Ainda sobre estes dois últimos modelos citados, é importante enfatizar que ambos precisam pouco das medidas de intensidade. Em termos comparativos, especialmente as fronteiras terminais parecem precisar menos do parâmetro de intensidade.

5. Conclusões e perspectivas futuras

Este trabalho teve como objetivo investigar os parâmetros fonético-acústicos relevantes para produção e percepção de fronteiras de natureza prosódica em dados de fala espontânea monológica em português brasileiro (PB). A análise aqui desenvolvida utilizou uma metodologia de base fonético-acústica e estatística. Foram utilizados sete trechos de fala espontânea monológica masculina. Os trechos são compostos por aproximadamente 200 palavras e foram extraídos dos *corpora* C-ORAL-BRASIL (Raso e Mello, 2012; Raso e Mello, em preparação). Cada trecho foi segmentado autonomamente em unidades entonacionais por quatorze segmentadores *experts*. As fronteiras prosódicas percebidas pelos segmentadores foram marcadas como fronteiras prosódicas terminais (/) ou não terminais (/).

Os trechos foram segmentados em cinco camadas de anotação usando o *software* Praat (Boersma e Weenink, 2015): 1) Segmentação em unidades VV utilizando transcrição fonética larga em caracteres ASCII; 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 do intervalo referente a pausas silenciosas; 5) Transcrição ortográfica da unidade entonacional.

Desenvolveu-se uma versão estendida do *script ProsodyDescriptor* (Barbosa, 2013) para extrair uma série de parâmetros acústicos ao longo do sinal de fala. A versão estendida, denominada *BreakDescriptor*, extrai os parâmetros acústicos em todas as unidades VVs em uma janela centrada em toda fronteira de palavra fonológica, o que inclui as posições percebidas pelos segmentadores como fronteiras e também não fronteiras. A janela inclui 10 unidades VVs à esquerda e 10 unidades VVs à direita de cada unidade VV. O *BreakDescriptor* considera como fronteira prosódica as posições em que pelo menos sete segmentadores perceberam uma fronteira, ou seja, pelo menos 50% de acordo. As demais posições, também localizadas em limites de palavras fonológicas, são consideradas pelo *script* como não fronteiras.

No total, para cada unidade VV, são calculadas 111 medidas fonético-acústicas globais e locais. As propriedades físicas relativas à produção dos sons extraídas automaticamente pelo *script* compreendem cinco parâmetros fonético-acústicos: 1) Medidas de *speech rate* e ritmo; 2) Medidas de duração normalizada e suavizada do segmento silábico;

3) Medidas de frequência fundamental (f_0); 4) Medidas de intensidade; 5) Medidas de pausa silenciosa.

As medidas extraídas automaticamente pelo *BreakDescriptor* foram submetidas a dois classificadores estatísticos, a saber *Random Forest* e *Linear Discriminant Analysis*, para identificar a combinação de medidas que melhor explicam a segmentação realizada pelos segmentadores perceptualmente. Consideramos, para ambos os modelos de classificação, a presença de fronteira e a ausência de fronteira, tanto para as fronteiras terminais, quanto para as fronteiras não terminais. Consideramos também o poder de predição alcançado pelos modelos obtidos. Uma análise deste tipo tem como propósito gerar modelos com parâmetros fonético-acústicos, que poderiam funcionar como preditores do estabelecimento de fronteiras prosódicas.

Os resultados mostram sucesso relativo de ambos os classificadores para detectar automaticamente fronteiras prosódicas percebidas pelos segmentadores humanos. Em termos gerais, observamos que o modelo estatístico gerado por meio do LDA aproximou-se mais às decisões tomadas pela maioria dos segmentadores. Por isso, ele foi submetido a várias fases de treinamento.

As diversas fases de treinamento tinham como objetivo reduzir o "ruído" nos modelos gerado por parâmetros não relevantes para o estabelecimento de fronteiras, aumentar a porcentagem de acertos para a presença de fronteira e reduzir aquela de falsos alarmes para a ausência de fronteira. Após a realização de várias fases de treinamento, os modelos obtidos para classificar fronteiras terminais e não terminais indicam que há uma diferença clara entre os dois tipos de fronteiras. As fronteiras terminais são mais facilmente identificadas por meio de um modelo que contém 20 medidas. As fronteiras não terminais, de fato, são mais difíceis de serem detectadas automaticamente por meio das configurações de parâmetros fonético-acústicos.

Como o modelo destinado à classificação de fronteiras terminais alcançou um acerto de 80%, a atenção do trabalho se concentrou nas fronteiras não terminais. Para tentar melhorar o desempenho do modelo destinado à classificação de fronteiras não terminais, optou-se por excluir do conjunto de dados fronteiras não terminais localizadas diante de *retractings*, tomadas de tempo e "né", porque foi observada uma alta porcentagem de erro para classificar essas fronteiras. Depois disso, foi desenvolvida uma análise estatística que buscava investigar a hipótese de que há subgrupos de fronteiras não conclusivas marcados por diferentes configurações de parâmetros fonético-acústicos. Para esta última fase, foram consideradas 225

posições de fronteiras não terminais e 160 de não fronteira, excluindo as posições de fronteiras terminais para tentar reduzir o ruído gerado no modelo por outro tipo de fronteira prosódica.

Os resultados mostram que as fronteiras não terminais podem ser classificadas automaticamente por meio de três modelos distintos. O Modelo 1 contém 9 medidas e identifica 152 das 225 fronteiras não terminais indicadas pelos segmentadores. O Modelo 2 contém 10 medidas e identifica 57 fronteiras não terminais. O Modelo 3 contém 8 medidas e classifica automaticamente 11 fronteiras não terminais indicadas na amostra. Os três modelos juntos identificam 220 das 225 fronteiras não terminais indicadas nos textos de fala espontânea monológica masculina. Os três modelos destinados à classificação de fronteiras não terminais ainda podem ser subdivididos em grupos menores de fronteiras não terminais marcados por diferentes configurações de parâmetros fonético-acústicos através de uma análise de cluster hierárquica que mostra diversas possibilidades de agrupamentos, de acordo com as configurações de parâmetros.

As principais diferenças entre fronteiras terminais e não terminais no que diz respeito aos parâmetros fonético-acústicos estão relacionadas com a relevância hierárquica das medidas para prever os dois tipos de fronteiras prosódicas. As medidas mais relevantes hierarquicamente para prever o estabelecimento de fronteiras terminais são a duração da pausa, a presença da pausa, o *reset* de f_0 , a mudança no contorno entonacional de f_0 na unidade VV posterior à fronteira e a tendência geral de mudanças no contorno entonacional de f_0 antes da fronteira. As medidas de *speech rate* e de duração normalizada e suavizada do segmento silábico parecem desempenhar um papel maior na classificação de fronteiras não terminais, apesar de algumas medidas de frequência fundamental serem importantes também para o modelo de fronteiras não terminais. As medidas de intensidade se mostraram pouco relevantes para classificar fronteiras terminais e não terminais.

As medidas mais relevantes hierarquicamente para prever 68% das fronteiras não terminais são o alongamento fronteiro e o alongamento pós-fronteiro (Modelo 1). Para prever aproximadamente 25% das fronteiras não terminais (Modelo 2), as medidas mais importantes são a taxa de elocução antes da fronteira, a regularidade das mudanças no contorno entonacional de f_0 antes da fronteira, a variação de f_0 antes da fronteira e a mudança da taxa de articulação. Para prever a realização de aproximadamente 5% das fronteiras não terminais (Modelo 3), as medidas mais relevantes hierarquicamente são a taxa de saliência duracional antes da fronteira, a mudança da taxa de saliência duracional e a taxa de saliência duracional após a fronteira.

As principais semelhanças entre fronteiras terminais e não terminais no que diz respeito às configurações de parâmetros fonético-acústicos são devidas às medidas hierarquicamente relevantes para predizer ambos os tipos de fronteiras. Trata-se principalmente das medidas relacionadas à pausa, de algumas medidas relacionadas à frequência fundamental, à *speech rate* e à duração normalizada do segmento silábico. Contudo, a contribuição de cada uma delas na classificação estatística é diferente, conforme seja o tipo de fronteira.

A tarefa de detecção automática de fronteiras prosódicas apresentada neste trabalho é, em geral, comparável com os resultados recorrentes na literatura da área. Porém, apresenta como vantagem a possibilidade desenvolver modelos capazes de distinguir fronteiras terminais de não terminais. Além disso, os modelos são orientados para a fala espontânea e parecem ser capazes de distinguir também subtipos de fronteiras não terminais.

Futuramente, será importante:

1. Aumentar a base de dados para desenvolver análises que possam alcançar um maior grau de refinamento;
2. Verificar em detalhes os casos de desacordo entre segmentação perceptual realizada pelos segmentadores e segmentação automática realizada pelos modelos;
3. Aprofundar a análise dos casos de fronteiras localizadas antes ou após “né”;
4. Verificar o desempenho dos modelos diferenciando os casos de fronteiras marcadas com pausa daqueles marcados sem pausa, já que a pausa sozinha é um marcador de fronteira;
5. Aprofundar o modelo destinado à classificação estatística de fronteiras terminais, porque possivelmente ele ainda pode ser melhorado;
6. Verificar as posições de falsos alarmes, observando o comportamento dos segmentadores e as configurações de parâmetros fonético-acústicos;
7. Verificar melhor a causa do Modelo 1 de classificação de fronteiras não terminais aumentar muito o seu desempenho quando são retiradas do conjunto de dados as posições de fronteiras terminais, de modo a reduzir o ruído provocado por estas;
8. Verificar se os resultados obtidos com as vozes masculinas se confirmam com as vozes femininas. Se isso ocorrer, poderemos trabalhar com modelos

únicos destinados às duas vozes; caso contrário, as mesmas operações deverão ser realizadas para obter modelos destinados à classificação de fronteiras em vozes femininas;

9. Analisar dados de fala dialógica, que certamente apresentarão problemas de natureza diferente daqueles identificados na fala monológica.

REFERÊNCIAS

- 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: Japan, 2004.
- AUER, P. *Zum Segmentierungsproblem in der gesprochenen Sprache*. 2010. Disponível em: Acesso em: 18/01/2017.
- AUSTIN, J. *How to do things with words*. London: Oxford University Press, 1962.
- AVANZI, M., LACHERET-DUJOUR, A., VICTORRI, B. ANALOR: A Tool for Semi-Automatic Annotation of French Prosodic Structure: ANALOR, Campinas, Brasil, 2008, p.119-122.
- BARBOSA, P. Caractérisation et génération automatique de la structuration rythmique du français. Tese (Doutorado) – Institut National Polytechnique de Grenoble, França, 1994.
- BARBOSA, P. At least two macrorhythmic units are necessary for Brazilian Portuguese duration. In: *Proceedings of the First ESCA Tutorial Research Workshop on Speech Production Modeling and Fourth Speech Production Seminar*. Autrans, França: 1996a, p. 85-88.
- BARBOSA, P. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration: emphasis on segmental duration generation. *Cadernos de Estudos Linguísticos*, n. 31, p. 33-53, 1996b.
- BARBOSA, P. Elementos para uma tipologia do ritmo (lingüístico) da fala à luz de um modelo de osciladores acoplados. In *Cognito-Cadernos Românicos em Ciências Cognitivas*. 2 (1), 2004, p. 31-58.
- BARBOSA, P. A. *Incursões em torno do ritmo da fala*. São Paulo: Pontes Editores, 2006.
- BARBOSA, P. Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação, 06/2012, *Revista de Estudos da Linguagem*, Vol. 20, Faculdade de Letras, Belo Horizonte, Brasil, 2012, p.11-27.
- BARBOSA, P. Semi-automatic and automatic tools for generating prosodic descriptors for prosody research. 2013. In Brigitte Bigi & Daniel Hirst (eds.), *Proceedings of the Tools and Resources for the Analysis of Speech Prosody*, vol. 13, 86–89. Aix-en-Provence: Laboratoire Parole et Language. Disponível em: <http://www.lpl-aix.fr/~trasp/Proceedings/19874-trasp2013.pdf>.
- BARBOSA, P.; SILVA, W. A New Methodology for Comparing Speech Rhythm Structure between Utterances: Beyond Typological Approaches. In: *Computational Processing of the Portuguese Language 10th International Conference*, PROPOR 2012, Coimbra, Portugal, Abril, 17-20, 2012.
- BARBOSA, P. *BreakDescriptor* (Versão 1.0) [Programa de computador]. 2016. Disponível com o autor.
- BARTH-WEINGARTEN, D. *Intonation Units Revised: Cesuras in talk-in-interaction*. Philadelphia: John Benjamins Publishing Company, 2016.
- BATLINER, A. et al. M = Syntax + Prosody. A syntactic-prosodic labelling scheme for large spontaneous speech databases. *Speech communication* 25, 1998, p. 193-222.
- BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. *The Longman Grammar of Spoken and Written English*. London and New York: Longman, 1999.
- BIRKNER, K. Konstruktionen zur Personenattribution. In: GÜNTHER, S., WOLFGANG, I. *Konstruktionen in der Interaktion*. Berlin: Mouton de Gruyter, 2006, p. 205-238.

- BLAAUW, E. The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech communication*, 14, Elsevier Science Publishers, 1994, p. 359-375.
- BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer*. 2015. Programa de computador. Disponível: <<http://www.praat.org/>>. Acesso em: 16 jan. 2015.
- BOLINGER, D. Around the edges of language. In: BOLINGER, D. (ed.), *Intonation: Selected Readings*. Harmondsworth. Penguin: 1972, p. 19-29.
- BROWN, G. et al. *Questions of Intonation*. Londres: Croom Helm, 1980.
- BUHMANN, Jeska et al. Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the Spoken Dutch Corpus. In: *International Conference on Language Resources and evaluation*, 3, 2002, Las Palmas. Proceedings. Las Palmas: Elra, Universidad de Las Palmas de Gran Canaria, 2002. p. 1 - 7. Disponível em: <http://www.lrec-conf.org/proceedings/lrec2002/pdf/96.pdf>. Acesso em: Janeiro, 2018.
- BYBEE, J. *Language, Usage and Cognition*. Cambridge: CUP, 2010.
- CAMPBELL, N. Automatic detection of prosodic boundaries in speech. *Speech communication*. North-Holland: Elsevier Science Publishers, 1993, p. 343-354.
- CAMBIER-LANGEVELD, T. The domain of final lengthening in the production of Dutch. In COERTS, J., HOOP, H. (Eds.), *Linguistics in the Netherlands*. Amsterdam: John Benjamins, 1997, p. 13-24,
- CHAFE, W. The Deployment of Consciousness in the production of a Narrative. In: _____. (Org.). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood: Ablex, 1980. p. 9-50.
- CHAFE, W. *Discourse, consciousness and time: The Flow and displacement of Conscious Experience in Speaking and writing*. Chicago: University of Chicago Press, 1994.
- CHISTOVICH E, L.; OGORODNIKOVA, E. Temporal processing of spectral data in vowel perception. *Speech communication*, v.1, 1982, p. 45-54.
- CHO, T., KEATING, P. Articulatory strengthening at the onset of prosodic domains. In: *Korean Journal of Phonetics*, 28, 2001, p.155-190.
- CHOMSKY, N. Remarks on nominalization. In: R. JACOBS; ROSENBAUM, P. (eds.) *Reading in English Transformational Grammar*. Waltham: Ginn, 1970, p. 184-221.
- 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 in spontaneous speech. *Experimental and Theoretical Advances in Prosody*. Cornell University: 2008, p. 11-13.
- COOPER, W., PACCIA COOPER, J. *Syntax and Speech*. Cambridge: Harvard University Press, 1980.
- COOPER, W., PACCIA, J., LAPOINTE, S. Hierarchical coding in speech timing. *Cognitive Psychology*. Vol. 10, Elsevier, 1978.
- 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.

- CRESTI E., GRAMIGNI P. Per una linguistica corpus based dell'italiano parlato: le unità di riferimento. In: F. Albano Leoni, F. Cutugno, F. Pettorino, M. Savy (a cura di). *Il parlato italiano. Atti del Convegno internazionale*. Napoli: D'Auria Editore, 2004.
- CRESTI, E. *Corpus di Italiano parlato*. v. 1. Firenze: Accademia della Crusca, 2000.
- CRESTI, E.; MONEGLIA, M. (Orgs.). *C-ORAL-ROM: Integrated reference corpora for spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins, 2005.
- CROFT, W. Intonation Units and Grammatical Structure. *Linguistics* 33 (5), 839-882, 1995.
- 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, p. 141-201, 1997.
- DOGIL, G. Prototypical speech events and speech perception. In: *Proceedings of the XI International Congress of Phonetic Sciences*. Tallin, Estônia, 1-7 de agosto, 1987, v.3, p. 360-365.
- DU BOIS, J. et al. Discourse Transcription, Santa Barbara Papers in Linguistics 4. Santa Barbara: Department of Linguistics, University of California, 1992.
- DU BOIS, J. W., CHAFE, W. L.; MEYER, C.; THOMPSON, S. A.; ENGLEBRETSON, R.; MARTEY, N. *Santa Barbara corpus of spoken American English, Parts 1-4*. Philadelphia: Linguistic Data Consortium, 2000-2005. Disponível: <http://www.linguistics.ucsb.edu/research/santa-barbara-corpus>. Acesso em: Janeiro, 2018.
- DU BOIS, J. Rhythm and Tunes: The notation Unit in the Structure of Dialogic Engagement. *Conferência Prosody and Interaction*, University of Potsdam, 2008.
- DUANMU, S. Pre-juncture lengthening and foot binarity. *Studies in the Chinese Linguistic*. 26.1/2, 1996, p. 95-115.
- EDWARDS, J. A.; LAMPERT, M. (Eds.). *Talking Data: Transcription and Coding Methods for Language Research*. NJ: Lawrence Erlbaum, 1993.
- FLEISS, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, Vol. 76, No. 5, p. 378-382, 1971.
- FLETCHER, J. The prosody of speech: Timing and rhythm. In: HARDCASTLE, W. J.; LAVER, J.; GIBBON, F. E. *The handbook of phonetic sciences*, 2ª ed. West Sussex: WileyBlackwell, 2010, p. 523-602.
- FÓNAGY, I. *Des fonctions de l'intonation : Essai de synthèse*. Flambeau, 29, p. 1-20, 2003.
- FRAZIER, L., C. CLIFTON, C., CARLSON, K. Don't break, or do: Prosodic boundary preferences. *Lingua*, vol. 114, no 1, 2004, p. 3-27.
- GAMER, M.; LEMON, J.; SINGH, I. irr: Various Coefficients of Interrater, Reliability and Agreement, 2012. R package version 0.84.

- GEISLER, C. Representation of speech sounds in the auditory nerve. *Journal of Phonetics*, v.16, 1988, p. 19-35.
- HALLIDAY, M.A.K. *Speech and Situation*. Londres: University College, 1965.
- HARRIS, Z. *String Analysis of Sentence Structure*. The Hague: Mouton, 1962.
- HIRST, D. Peak, boundary and cohesion characteristics of prosodic grouping, *Prosody*, 1993, p. 32-37.
- IZRE'EL, S. et al. *The corpus of spoken Israeli Hebrew (CoSIH)*, 2009. Disponível em: <http://cosih.com/english/index.html>. Acesso em: Janeiro, 2018.
- KELLY, J.; LOCAL, J. On the Use of General Phonetic Techniques in Handling Conversational Material. In: ROGER et al. *Conversation: An Interdisciplinary Perspective*. Clevedon: Multilingual Matters, 1989.
- KIM, J. *Automatic detection of sentence boundaries, disfluencies, and conversational fillers in spontaneous speech*. 103 f. Tese (Doutorado). University of Washington, 2004.
- KNOWLES, G. Prosodic labelling: the problem of tone group boundaries. In Johansson, S. & Stenstrom, A.-B. (eds.), *English Computer Corpora: Selected Papers and Research Guide*. Berlin: Mouton de Gruyter, 1991.
- LADD, R. *Intonational phonology*. 2ed. Cambridge: CUP, 2008.
- LIAW, A. WIENER, M. Classification and Regression by randomForest. *R News* 2(3), 2002, p. 18-22.
- LINDBLOM, B., RAPP, K. Some Temporal Regularities of Spoken Swedish. *Papers from the Institute of Linguistics (PILUS)* 21. University of Stockholm, Sweden, 1973.
- LIU, Y. et al. Enriching speech recognition with automatic detection of sentence boundaries and disfluencies, *IEEE Trans. Audio, Speech Lang. Process.*, vol. 14, no 5, 2006, p. 1526–1539.
- MARCUS, S. Acoustic Determinants of Perceptual-Center (P-Center) Location. *Perception and Psychophysics* 30(3). 1981, p. 247-256.
- MARTIN, P. Phonetic realizations of prosodic contours in French. *Speech Communication*, 1(3,4), p. 283-294, 1982.
- MARUYAMA, T. Speech segmentation by clausal and non-clausal boundaries in Japanese. International Conference on Cognitive Science, 5, Kaliningrad. Abstracts... V. 2: *Workshop Spoken discourse corpora as a window on cognitive mechanisms of speech production*. Kaliningrad, Junho 2012. p. 787-783
- MARUYAMA, T. Two-Level Utterance Units: Cognitive and Communicative Aspects of Spontaneous Speech. LABLITA, 9; LEEL, 4, International Workshop, Belo Horizonte, 2015. Conferências. *Units of Reference for Spontaneous Speech Analysis and their correlations across languages*, Belo Horizonte, 6 de agosto de 2015.
- MELLO, H. R. et al. Transcrição e segmentação prosódica do corpus C-ORAL-BRASIL: critérios de implementação e validação. In: RASO, T.; MELLO, H. R. (Ed.) *C-ORAL – Brasil I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG, 2012. p. 125-176.
- MELLO, H.; RASO, T. Para a transcrição da fala espontânea: o caso do C-ORAL-BRASIL. *Revista Portuguesa de Humanidades*, v. 13, n. 1, 2009, p. 153-178.
- MERLO, S.; BARBOSA, P. A. Séries temporais de pausas e de hesitações na fala espontânea. *Caderno de Estudos Linguísticos*. v. 1, n. 54, p. 11-24, 2012.

- MERTENS, P.; SIMON, A. C. Towards Automatic Detection of Prosodic Boundaries in Spoken French. In: MERTENS, P.; SIMON, A. C. *Proceedings of the Prosody-Discourse Interface Conference 2013 (IDP 2013)*, Leuven, p. 81-87, 2013.
- METTOUCHI, A.; CHANARD, C. From Fieldwork to Annotated Corpora: the CorpAfroAs Project. *Faits de Langue-Les Cahiers*. n. 2, p. 255-265, 2010.
- MITTMANN, M. M.; BARBOSA, P. An automatic speech segmentation tool based on multiple acoustic parameters. *CHIMERA: Romance Corpora and Linguistic Studies*, v. 3, p. 133-147, 2016.
- 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.
- NI, C.J., ZHANG, A.Y., LIU, W.J., XU, B. Automatic prosodic break detection and feature analysis. *J. Comput. Sci. Tchol.*, 27, 2012, p. 1184-1196.
- PETERS et al. Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In: KLAUS et al. (eds), *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)*, Kiel: IPdS, 2005, p. 143-184.
- PIERREHUMBERT, J. *Phonetics and phonology of English intonation*. 1980. Dissertação. Massachusetts Institute of Technology. 1980.
- PIKE, L. *The intonation of American English*. Ann Arbor: University of Michigan Press, 1945.
- POMPINO-MARSCHALL, B. On the Psychoacoustic Nature of the P-Center Phenomenon. *Journal of Phonetics* (17), 1989, p. 175-192.
- PORTAL ACTION. *Coeficiente de assimetria*. Disponível em: <http://www.portalaction.com.br/estatistica-basica/25-coeficiente-de-assimetria>. Acesso: Jan 2018.
- 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, p. 883-922, 2015.
- RASO, T. Fala e escrita: meio, canal, consequências pragmáticas e linguísticas. *Domínios de Lingu@gem*, v. 7, p. 12-46, 2013.
- RASO, T.; MELLO, H. (Org.). *C-ORAL-BRASIL I: corpus de referência do português brasileiro falado informal*. 1ed. Belo Horizonte: UFMG, 2012.
- RASO, T.; MITTMANN, M. M.. As principais medidas da fala. In: Tommaso Raso; Heliana Mello (Org.). *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. 1ed. Belo Horizonte: UFMG, 2012, p.178-221.
- ROELOF DE PIJPER, J., SANDEMAN, A. A. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America* (96) p. 2037-2047, 1994.
- SCHEGLOFF, E. Reflections on Studying Prosody in Talk-in-interaction. *Language and Speech* 41 (3-4): 235-263, 1998.
- SCHUBIGER, M. *English Intonation: Its Form and Function*. Tübingen: Niemeyer, 1958.

- SCHUETZE-COBURN, S. *Prosody, syntax, and discourse pragmatics: Assessing information flow in German conversation*. Dissertação. Ph.D. Universidade da Califórnia, Los Angeles. 1994.
- SCHUETZE-COBURN, S; SHAPLEY, M; WEBER, E. G. Units of intonation in discourse: a comparison of acoustic and auditory analyses. *Language and speech*. v. 34, n. 3, p. 207-234, 1991.
- SCOTT, S. WISE, R. PET and fMRI studies of the neural basis of speech perception. *Speech communication*, v.41, 2003, p. 23-34.
- SELKIRK, E. Comments on Intonational Phrasing in English. In: FROTA,S.; VIGÁRIO, M. & FREITAS, M.J. (eds.) *Prosodies*, Berlim: Mouton de Gruyter, p. 11-58, 2005.
- SELMINI, A. M. ; VIOLARO, F. Segmentação Automática de Fala para o Português Brasileiro. In: *XXVI Simpósio Brasileiro de Telecomunicações*, 2008, Rio de Janeiro - RJ. Anais do XXVI Simpósio Brasileiro de Telecomunicações - Biblioteca Digital, 2008.
- SHATTUCK-HUFNAGEL, S., TURK, A. A Prosody Tutorial for Investigators of Auditory Sentence Processing. *Journal of Psycholinguistic Research*, Vol. 25, No. 2, 1996.
- SHEN, X. A pilot study on the relation between the temporal and syntactic structures in Mandarin. *Journal of the International Phonetic Association*, 22: p. 35- 43, 1992.
- SHRIBERG, E. Spontaneous speech: How people really talk and why engineers should care, Interspeech 2005, 9th Eur. Conf. Speech Commun. Technol. Lisboa, p. 1781–1784, 2005.
- SORIANELLO, P. *Prosodia*. Roma: Carocci, 2006.
- STEEDMAN, M. Structure and intonation in spoken language understanding. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*. Pittsburgh, 1990.
- STETSON, R. *Motor phonetics: A Study of Speech Movements in Action*. Oberlin: Springer Netherlands, 1928.
- 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. Disponível em: <http://dx.doi.org/10.1109/ICASSP.1987.1169628>. Acesso em: 16 jan. 2018.
- SWERTS, M, GELUYKENS, R., TERKEN, J. , Prosodic correlates of discourse units in spontaneous speech, in *Proceedings of the International Conference on Spoken Language Processing*, 1992, p. 421–424.
- SWERTS, M., R., COLLIER, R., TERKEN, J. Prosodic predictors of discourse finality in spontaneous monologues, *Speech Commun.*, vol. 15, no 1–2, 1994, p. 79–90.
- SWERTS, M. Prosodic features at discourse boundaries of different strength, *J. Acoust. Soc. Am.*, vol. 101, no 1, p. 514–521, 1997.
- SZCZEPEK REED, B. Prosody, Syntax and Action Formation: Intonation Phrases and Action Components. In: BERGMANN, P. et al. (eds), *Prosody and Embodiment in Interactional Grammar*, Berlim: Mouton de Gruyter, 2012, p. 142-169.
- 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.
- TRAUNMÜLLER, H., ERIKSSON, A. Acoustic effects of variation in vocal effort by men, women, and children. In: *The Journal of the Acoustical Society of America*, v. 107, n. 6, 2000, p. 3438-3451.

- TSENG, C., CHANG, C. Pause or no pause?: Prosodic phrase boundaries revisited. *Tsinghua Sci. Technol.*, vol. 13, no 4, 2008, p. 500–509.
- TURK, E., SHATTUCK-HUFNAGEL, S. Phrase-final lengthening in American English. *Journal of Phonetics*, 35(4), 2007, p. 445-472.
- TYLER, J. Prosodic correlates of discourse boundaries and hierarchy in discourse production. *Lingua*, vol. 133, 2013, p. 101–126.
- VAN HEMERT, J. P. Automatic segmentation of speech. *IEEE Transactions On Signal Processing, Piscataway*, v. 39, n. 4, Institute of Electrical & Electronics Engineers (IEEE), p. 1008-1012, 1991.
- VENABLES, W. N. RIPLEY, B. D. *Modern Applied Statistics with S*. 4ed. New York: Springer, 2002.
- 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, 1707-1717, 1992.
- XIE, L. Discovering salient prosodic cues and their interactions for automatic story segmentation in Mandarin broadcast news, *Multimed. Syst.*, vol. 14, no 4, 2008, p. 237–253.
- YANG, Y., WANG, B. Acoustic correlates of hierarchical prosodic boundary in Mandarin. *Proceedings of Speech Prosody*. Aix-en-Provence: 2002.
- 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.
- ZHANG, X. *A Comparison of Cue-Weighting in the Perception of Prosodic Phrase Boundaries in English and Chinese*. 128 f. Tese (Doutorado em Linguística) – Faculdade de Filosofia, University of Michigan, Ann Arbor, 2012.

ANEXOS

Script do Praat *BreakDescriptor* (Barbosa, 2016)

```
# BreakDescriptor.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, new version (x.y):
form File acquisition
word FileOut OutPutParameters.txt
word AudiofileExtension *.wav
integer VVTier 1
integer NTBTier 2
integer TBTier 3
integer PauseTier 4
integer F0Thresholdleft 75
integer F0Thresholdright 300
positive Smthf0Thr 5
positive F0step 0.05
positive Njudges 14
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 ctb ntb cntb psp psdur
srl srr srd arl arr ard mzl mzm mzd zl1 zl2 zl3 zl4 zl5 zl6 zl7 zl8 zl9 zl10
zl0 zr1 zr2 zr3 zr4 zr5 zr6 zr7 zr8 zr9 zr10 zdloc SDzl SDzr SDzd skzl skzr
```

```

skzd prl prr prd f0medl f0medr f0medd f0medl1 f0medl2 f0medl3 f0medl4 f0medl5
f0medl6 f0medl7 f0medl8 f0medl9 f0medl10 f0med0 f0medr1 f0medr2 f0medr3
f0medr4 f0medr5 f0medr6 f0medr7 f0medr8 f0medr9 f0medr10 f0meddloc sdf0l
sdf0r sdf0d skf0l skf0r skf0d df0medl df0medr df0medd df0medl1 df0medl2
df0medl3 df0medl4 df0medl5 df0medl6 df0medl7 df0medl8 df0medl9 df0medl10
df0med0 df0medr1 df0medr2 df0medr3 df0medr4 df0medr5 df0medr6 df0medr7
df0medr8 df0medr9 df0medr10 df0meddloc sddf0l sddf0r sddf0d skdf0l skdf0r
skdf0d f0rl f0rr f0rd emphl emphr emphd emphdloc 'newline$'
for ifile from 1 to numberOfFiles
  select Strings list
  audiofile$ = 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$'
begin = Get starting time
end = Get finishing time
totaldur = end - begin
nselected = Get number of intervals... 'vVTier'
nptobreaks = Get number of points... 'tBTier'
#
##### df0, VVwise
tl = Get starting point... 'vVTier' 2
tr = Get end point... 'vVTier' 2
select Pitch 'filename$'
#Down to PitchTier
k = nselected -1
df0'k' =0
f0ant = Get quantile... 'tl' 'tr' 0.5 semitones re 1 Hz
f0medl = f0ant
df0medl = 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
    dell = i - 1
    if smz'i' > smz'dell'
      boundary'i' = 1
      boundcount = boundcount + 1
      btime'i' = time
      bctime'boundcount' = time
    else
      boundary'i' = 0
    endif
  endif
endifor
# end loop for 2
# Computing parameters in the two 10-VV-long half windows
endVV = nselected-9
for x from 1 to npotbreaks
  select TextGrid 'filename$'
  timeVV = Get time of point... 'tBTier' 'x'
  indexVV = Get interval at time... 'VVTier' 'timeVV'
  iVV = indexVV - 1
  if iVV > 10 and iVV < endVV
# Agreement judges
  tbtime = Get time of point... 'tBTier' 'x'
  tb$ = Get label of point... 'tBTier' 'x'
  tb = number(tb$)
  ntb$ = Get label of point... 'nBTier' 'x'
  tb20 = tbtime + 0.02
  ntb = number(ntb$)
  indpause = Get interval at time... 'pauseTier' 'tb20'
  pause$ = 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 - 10
  lefttime = Get starting point... 'VVTier' 'indleft'

```

```

indright = iVV
righttime = Get starting point... 'VVTier' 'indright'
ileftlastwindowunit = iVV-1
leftlastwindowunit = Get starting point... 'VVTier'
'ileftlastwindowunit'
srl = 10/(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/10
f0medl = f0medl/10
df0medl = df0medl/10
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/9)
sdf0l = sqrt(sdf0l/9)
sddf0l = sqrt(sddf0l/9)
# 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 = (10/((10-1)*(10-2)))*skzl
skf0l = (10/((10-1)*(10-2)))*skf0l
skdf0l = (10/((10-1)*(10-2)))*skdf0l
# Computing left f0rate
select PointProcess temp
f0rl = Get mean period... 'lefttime' 'righttime' 0.3 3 5

```

```

f0r1 = 1/f0r1
# 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'
emph1 = 1000*(allenergy - filtenergy)
emphlast1 = allenergylast - filtenergylast
# Computing SR/AR/mz in the right half-window
select TextGrid 'filename$'
indleft = iVV
lefttime = Get end point... 'VVTier' 'indleft'
indright = iVV + 10
righttime = Get end point... 'VVTier' 'indright'
irightlastwindowunit = iVV+1
rightlastwindowunit = Get end point... 'VVTier' 'irightlastwindowunit'
srr = 10/(righttime - lefttime)
srd = srr - srl
sdur = 0
sz = 0
f0medr = 0
df0medr = 0
pr = 0
nunstr = 10
for l from iVV+1 to indright
  if boundary'l' == 0
    sdur = sdur + dur'l'
  else
    nunstr = nunstr - 1
    pr = pr + 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/10
f0medr = f0medr/10
df0medr = df0medr/10
pr = 1000*pr/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/9)
sdf0r = sqrt(sdf0r/9)
sddf0r = sqrt(sddf0r/9)
# 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 = (10/((10-1)*(10-2)))*skzr
skf0r = (10/((10-1)*(10-2)))*skf0r
skdf0r = (10/((10-1)*(10-2)))*skdf0r
ard = arr - arl
mzd = mzr - mzl
zdloc = zrl - 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 = 1000*(allenergy - filtenergy)
emphfirstright = 1000*(allenergyfirst - filtenergyfirst)
emphd = emphr - emphl
emphdloc = emphfirstright - emphlastl
ctb = njudges - tb
cntb = njudges - ntb
printcond = 1
anyundefined =
tb*ntb*mzl*mzr*mzd*zl1*zl2*zl3*zl4*zl5*zl6*zl7*zl8*zl9*zl10*zl11*zr1*zr2*zr
3*zr4*zr5*zr6*zr7*zr8*zr9*zr10*zdloc*sdzl*sdzr*sdzd*skzl*skzr*skzd*prl*pr
r*pr
d*pr
d*f0medl*f0medr*f0medd*f0medl1*f0medl2*f0medl3*f0medl4*f0medl5*f0medl6*f0
medl7*f0medl8*f0medl9*f0medl10*f0medl11*f0medr1*f0medr2*f0medr3*f0medr4*f0
medr5*f0medr6*f0medr7*f0medr8*f0medr9*f0medr10*f0meddloc*sdf0l*sdf0r*sdf0d*s
kf0l*skf0r*skf0d*df0medl*df0medr*df0medd*df0medl1*df0medl2*df0medl3*df0medl
4*df0medl5*df0medl6*df0medl7*df0medl8*df0medl9*df0medl10*df0medl11*df0medr1
*df0medr2*df0medr3*df0medr4*df0medr5*df0medr6*df0medr7*df0medr8*df0medr9*df
0medr10*df0meddloc*sddf0l*sddf0r*sddf0d*skdf0l*skdf0r*skdf0d*f0rl*f0rr*f0rd
*emphl*emphr*emphd*emphdloc
if anyundefined == undefined
    printcond = 0
else
    if tb >= njudges/2
        tb$ = "STR"
    else
        tb$ = "NO"
    endif
    if ntb >= njudges/2
        ntb$ = "STR"
    endif

```

```

else
    ntb$ = "NO"
endif
endif
endif
if printcond
    fileappend 'fileOut$' 'filename$' 'tbtime:3' 'x' 'tb$' 'ctb' 'ntb$'
'cntb' 'prespause' 'durpause:3' 'srl:2' 'srr:2' 'srd:2' 'arl:2' 'arr:2'
'ard:2' 'mzl:2' 'mzr:2' 'mzd:2' 'zll:2' 'zlr:2' 'zld:2' 'zll:2' 'zlr:2' 'zld:2' 'zll:2' 'zlr:2' 'zld:2' 'zll:2' 'zlr:2' 'zld:2'
'zr4:2' 'zr5:2' 'zr6:2' 'zr7:2' 'zr8:2' 'zr9:2' 'zr10:2' 'zldloc:2' 'sdzl:2'
'sdzm:2' 'sdzm:2' 'skzl:2' 'skzm:2' 'skzd:2' 'prl:2' 'prz:2' 'prd:2'
'f0medl:0' 'f0medr:0' 'f0medd:0' 'f0medl1:0' 'f0medl2:0' 'f0medl3:0'
'f0medl4:0' 'f0medl5:0' 'f0medl6:0' 'f0medl7:0' 'f0medl8:0' 'f0medl9:0'
'f0medl10:0' 'f0medl11:0' 'f0medr1:0' 'f0medr2:0' 'f0medr3:0' 'f0medr4:0'
'f0medr5:0' 'f0medr6:0' 'f0medr7:0' 'f0medr8:0' 'f0medr9:0' 'f0medr10:0'
'f0meddloc:0' 'sdf0l:2' 'sdf0r:2' 'sdf0d:2' 'skf0l:2' 'skf0r:2' 'skf0d:2'
'df0medl:2' 'df0medr:2' 'df0medd:2' 'df0medl1:2' 'df0medl2:2' 'df0medl3:2'
'df0medl4:2' 'df0medl5:2' 'df0medl6:2' 'df0medl7:2' 'df0medl8:2'
'df0medl9:2' 'df0medl10:2' 'df0medl11:0' 'df0medr1:2' 'df0medr2:2'
'df0medr3:2' 'df0medr4:2' 'df0medr5:2' 'df0medr6:2' 'df0medr7:2'
'df0medr8:2' 'df0medr9:2' 'df0medr10:2' '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' 'newline$'
endif
endif
endif
# Loop: for x from 1 to npotbreaks
endifor
# All files
select all
Remove
##
procedure zscorecomp nome$ 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
            endif
        endif
    endwhile
endprocedure

```

```

        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
endp

```