

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOLOGIA E LÍNGUA
PORTUGUESA

DANIEL OLIVEIRA PERES

**O papel da prosódia na identificação das
variedades regionais do português brasileiro**

(versão corrigida)

São Paulo
2011

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOLOGIA E LÍNGUA
PORTUGUESA

**O papel da prosódia na identificação das
variedades regionais do português brasileiro**

(versão corrigida)

Daniel Oliveira Peres

Dissertação apresentada ao Programa de Pós-Graduação em Filologia e Língua Portuguesa do Departamento de Letras Clássicas e Vernáculas da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, para obtenção do título de Mestre em Letras.

Orientador: Prof. Dr. Waldemar Ferreira Netto

São Paulo
2011

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

Peres, Daniel Oliveira

O papel da prosódia na identificação das variedades regionais do português brasileiro / Daniel Oliveira Peres ; orientador Ferreira Netto, Waldemar. – São Paulo, 2011. 152 f. ; il.

Dissertação (Mestrado) – Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo. Departamento de Letras Clássicas e Vernáculas. Área de concentração: Filologia e Língua Portuguesa.

1. Prosódia. 2. Percepção. 3. Português do Brasil. 4. Variação linguística. I. Título. II. Ferreira Netto, Waldemar.

CDD 414

Aprovado em: 14 /10 /2011

Banca Examinadora:

Prof. Dr. Plínio Almeida Barbosa

Instituição: Universidade Estadual de Campinas - UNICAMP

Prof. Dr. Paulo Chagas de Souza

Instituição: Universidade de São Paulo - USP

À Adelina Russo de Oliveira (Vó Dilina)

In memoriam

Agradecimentos

O espaço diminuto que reservo aos agradecimentos é inverso à gratidão que sinto por quem me ajudou de alguma forma neste trabalho.

Agradeço a Deus pela força dada durante meu mestrado. Força que, muitas vezes, parecia se esvaír, mas que sempre retornava nos momentos de maior necessidade.

Aos meus pais, Waldir e Iara, pelo apoio e paciência, pelas condições plenas que tive para desenvolver meu trabalho. Agradeço também a Eduardo O. Peres e Tatiana Bortoleto.

À Fátima Almeida Baia pela inspiração, pelo ar fornecido nos momentos de vontade rarefeita que a vida nos impõe, pelo amor e carinho e pela sabedoria acadêmica que tanto me ajudou pelo caminho. Agradeço também à sua família que me recebeu de braços abertos.

Ao meu orientador, Prof. Dr. Waldemar Ferreira Netto, pela orientação e sua sincera paixão pelo saber. À sua dedicação em busca de explicações sobre o pantanoso mundo da prosódia.

Ao CNPq pela bolsa concedida.

Ao Prof. Dr. Plínio Almeida Barbosa pela leitura detalhada da minha qualificação e pela grande ajuda durante a feitura deste trabalho, seus comentários e auxílio nas análises.

Ao Prof. Dr. Mário Eduardo Viaro pela leitura atenta da minha qualificação e por suas sugestões.

À Profa. Rachel Hayes-Harb pela sugestão de análise dos dados.

À Fernanda Consoni pelo seu altruísmo e enorme ajuda dada durante a minha qualificação. À Gdalva pela ajuda nas gravações.

Aos demais colegas do Grupo *ExProsodia* pelos encontros e sugestões.

Ao André Xavier pela ajuda na análise dos dados e pelo inspirador e genuíno interesse pela pesquisa.

Agradeço a todos que me ajudaram nas viagens que fiz para a aplicação dos experimentos. À Carmem Regina Gonçalves Ferreira pela grande ajuda em Pelotas. Às famílias Almeida e Baia pela hospitalidade com a qual me receberam em Senador Pompeu. Ao Paulo e à Dorinha pela comida na mesa e simpatia sem fim. Ao Cícero, que me cedeu um teto e ótimas conversas durante a minha estadia no sertão cearense. Ao Adauzo, o melhor guia que se pode ter no sertão do Ceará.

Ao Leonardo Borgatto, Alexandre (Aleba), Maurício, Francisco (tio Chico), e demais informantes que cederam suas vozes à feitura deste trabalho. Ao Naldo e demais ouvintes que participaram dos experimentos perceptuais. Ao Prof. Dr. Luís Amaral por ceder o *corpus* VarX, da UFPel.

Às amigas Amanda Osti e Carolina Carnielli.

Aos meus amigos Alexandre Narkunas, Renato Lima de Barros e a todos os amigos que tenho na minha Mooca querida, meu!

RESUMO

Este estudo trata da variação regional em português brasileiro, pautando-se pelo viés perceptivo dos estudos da linguagem. Os trabalhos perceptuais de Atkinsons (1968), Bonte (1975), Maidment (1976), Ohala e Gilbert (1978) e Bezooijen e Gooskens (1999) já apontaram para a importância da prosódia como pista no reconhecimento das línguas e suas variedades regionais. Dessa maneira, este trabalho tem como objetivo verificar se os informantes de português brasileiro são capazes de reconhecer suas próprias variedades regionais apenas pela prosódia, além de buscar, por meio de uma análise de produção, pistas nas variações de F0 que possam justificar o reconhecimento dessas variedades. Trata-se de um estudo experimental sobre 3 variedades do português brasileiro: variedade de Pelotas (RS), a de São Paulo (bairro da Mooca) e a de Senador Pompeu (CE).

Um teste perceptual foi elaborado de maneira a eliminar do sinal acústico os segmentos produzidos (experimento 1), deixando como informação para os participantes somente as características prosódicas. Esse experimento foi subdividido em trechos de fala curtos e longos, e, para a feitura com fala delexicalizada, foi utilizado o *script* PURR (SONNTAG; PORTELE, 1998b). Em contrapartida, o experimento 2 eliminou as variações prosódicas, restando somente a curva melódica monotônica (CMM) e os segmentos. A monotonização dos trechos de fala foi feita por meio de uma senoide com o valor do tom médio de cada estímulo, obtido automaticamente via aplicativo *ExProsodia* (FERREIRA NETTO, 2010).

A análise dos dados foi feita de duas maneiras: via Teoria da Detecção – TD – (MACMILLAN; CREELMAN, 2005) e análise estatística dos dados. Os

resultados obtidos na TD demonstraram que os participantes tiveram um bom desempenho na tarefa de reconhecimento de suas variedades, com índices *d-prime* elevados ($d' > 0$).

O desempenho entre o experimento 1 e 2 foi diverso, apresentando uma maior dificuldade de reconhecimento no experimento 1 com estímulos delexicalizados (PURR). Os informantes tiveram um desempenho melhor no experimento 2 (75% de acertos contra 62% do experimento 1). O mesmo foi observado por Bezooijen e Gooskens (1999), em estudo sobre percepção de características dialetais do inglês e do holandês.

Na análise estatística dos dados, no experimento 1 (PURR), os informantes da variedade de São Paulo tiveram um desempenho diverso dos demais. Da mesma forma, na análise de produção, os estímulos da variedade de São Paulo apresentaram diferenças significativas no tom médio e na dispersão de F0.

A tendência dos alarmes falsos (TD) também corroborou a hipótese de que os ouvintes são capazes de reconhecer sua própria variedade por meio da prosódia.

Palavras-chave: prosódia; percepção; português do Brasil; variação linguística

ABSTRACT

This is a perceptual study on the regional variation in Brazilian Portuguese. Previous perceptual studies conducted by Atkinsons (1968), Bonte (1975), Maidment (1976), Ohala & Gilbert (1978) and Bezooijen & Gooskens (1999) have verified the importance of prosody in languages recognition and their regional varieties. This study aims to verify if Brazilian informants are able to recognize their own regional varieties only by listening to some prosodic information. Moreover, it also aims at searching for cues in F0 variation, which can justify the recognition of varieties.

The experimental studies conducted in this study deal with three varieties of Brazilian Portuguese: the variety of Pelotas (RS), São Paulo (Mooca district) and Senador Pompeu (CE).

The first experiment was designed to eliminate the segments of speech in order to maintain only the prosodic information. It was divided into short and long excerpts, which consisted of delexicalized speech created with PURR script (SONNTAG; PORTELE, 1998b). The experiment 2 eliminated F0 variations from speech in order to maintain only the monotonic melodic curve (MMC) and the segments. The monotonization was done by using a sinusoidal curve with the mean frequency of each stimulus. The mean frequency was obtained automatically by using *ExProsodia* application.

The data were analyzed in two ways: firstly, an analysis was done based on the Detection Theory – DT – (MACMILLAN; CREELMAN, 2005); secondly, a statistical analysis was performed. The results obtained in DT showed that the participants were good at the recognition task of their varieties. The levels of *d-*

prime were elevated ($d' > 0$). The results obtained by conducting experiment 1 and 2 were different as we observed more recognition difficulties in experiment 1 with delexicalized stimuli (PURR). The informants had a better performance in experiment 2 (75% of correct answers compared to 62% of experiment 1). The same was observed by Bezooijen & Gooskens (1999) in their study on the perception of dialectal characteristics of English and Dutch.

In the statistical analysis of experiment 1, the participants, who speak the São Paulo variety, showed different results when compared to others varieties. The same was observed in the analysis of production date as the results showed significant differences in mean frequency and F0 dispersion. The tendency of false alarms (DT) also corroborated the hypothesis which claims that prosody information plays an important role in the recognition of regional varieties.

Keywords: prosody; perception; Brazilian Portuguese; linguistic variation

“Sorprenderse, extrañarse, es comenzar a entender.”
(José Ortega y Gasset)

ÍNDICE

1. Introdução	19
1.1 Apresentação do tema.....	19
1.2 Objetivos e tarefas realizadas neste estudo	23
2. Histórico e formação do português brasileiro	25
2.1 Ocupação e características	25
2.2 Sobre as regiões culturais	32
2.3 Sobre Comunidades de língua (CL) e Comunidades de fala (CF) .	37
2.4 Sobre as regiões estudadas	43
3. Teoria Geral	46
3.1 Estudos de prosódia	46
3.2 Percepção e produção de fala	52
3.3 Abordagens sociofonéticas e categorização dialetal	59
4. Metodologia	67
4.1 Sobre a separação segmental-suprasegmental	67
4.2 Descrição dos experimentos	71
4.2.1 Experimento 1 com técnica PURR	71
4.2.2 Experimento 2 com curva melódica monotônica (CMM)...	74
4.2.3 A Teoria da Detecção	76
5. Análise automática da fala – <i>ExProsodia</i>	80
6. Análise dos dados	87
6.1 Análise com a Teoria da Detecção	87
6.2 Análise geral dos dados do experimento 1	93
6.3 Análise geral dos dados do experimento 2	102
6.4 Análise de produção	106

7. Discussão	111
8. Referências	118
9. Anexos	128

ÍNDICE DE FIGURAS, GRÁFICOS, QUADROS E TABELAS

Figura 1: Esquema das subdivisões de Nascentes (1953) para o português brasileiro	29
Figura 2: Divisão proposta por Nascentes (1953 apud NOLL, 2008)	30
Figura 3: Realização de /s/ final e de /s/ em coda silábica (NOLL, 2008)	30
Figura 4: Realização de /r/ inicial e de /r/ em coda silábica (NOLL, 2008)	31
Figura 5: Africativização de /t/ e /d/ (NOLL, 2008)	31
Figura 6: Mapa das regiões culturais do Brasil (DIEGUES, 1960, encarte) ...	36
Figura 7: Substâncias prosódicas propostas por Vaissière (2002)	51
Figura 8: Ilustração dada em Wertheimer (1938 [1927])	53
Figura 9: Ilustração dada em Wertheimer (1938 [1927])	53
Figura 10: Sobreposição de senoides para obter o sinal PURR	70
Figura 11: Total de estímulos e respostas do experimento 1 (PURR)	73
Figura 12: Resultado do processo de monotonização dos estímulos	75
Figura 13: Total de estímulos e respostas do experimento 2 (CMM)	76
Figura 14: Esquema das componentes de F0	81
Figura 15: Componentes estruturadoras da prosódia	82
Figura 16: Gráfico da escala de cinco tons (FERREIRA NETTO, 2008)	85
Figura 17: Gráfico produzido pela rotina <i>ExProsodia</i> de frequência e intensidade	85
Figura 18: Decomposição da série temporal feita pela rotina <i>ExProsodia</i>	86
Figura 19: Esquema dos alarmes falsos	115
Gráfico 1: Tendência percentual de erros do primeiro experimento de Bezooijen e Gooskens (1999)	65

Gráfico 2: Resultado do informante 1 TC (CE)	90
Gráfico 3: Resultado do informante 9 TL (RS)	90
Gráfico 4: Resultado do informante 11 TC (CE) para $d' = 0$	91
Gráfico 5: Resultado do informante 11 TL (CE) para $d' < 0$	91
Gráfico 6: Resultado do informante 4 (CE)	92
Gráfico 7: Resultado do informante 5 (CE)	93
Gráfico 8: Distribuição de acertos do experimento 1 (PURR) por duração de estímulo (RS)	94
Gráfico 9: Distribuição de acertos do experimento 1 (PURR) por duração de estímulo (SP)	95
Gráfico 10: Distribuição de acertos do experimento 1 (PURR) por duração de estímulo (CE)	96
Gráfico 11: Acertos e erros para cada variedade no experimento 1 (PURR) .	97
Gráfico 12: Boxplot de acertos e duração de estímulo no experimento 1 (PURR)	98
Gráfico 13: Boxplot de acertos e variedades regionais no experimento 1 (PURR)	98
Gráfico 14: Diferenças entre as médias das variedades regionais – Tukey ...	99
Gráfico 15: Boxplot de índices <i>d-prime</i> e duração de estímulos no experimento 1 (PURR)	100
Gráfico 16: Médias de acertos por variedade regional	101
Gráfico 17: Médias de índices <i>d-prime</i> por duração de estímulo	101
Gráfico 18: Desempenho dos 13 informantes de cada variedade no experimento 2 (CMM)	102
Gráfico 19: Total de erros e acertos por variedade no experimento 2 (CMM).....	103

Gráfico 20: Boxplot de acertos e variedades regionais no experimento 2 (CMM)	104
Gráfico 21: Boxplot de índices <i>d-prime</i> e variedades regionais no experimento 2 (CMM)	104
Gráfico 22: Boxplot de acertos e tipos de experimento (PURR vs CMM)	105
Gráfico 23: Médias de acertos nos experimentos 1 e 2	106
Gráfico 24: Boxplot de finalização dos estímulos e variedades regionais	107
Gráfico 25: Boxplot de tom médio dos estímulos e variedades regionais	108
Gráfico 26: Diferenças entre as médias de tom médio e variedades regionais – Tukey	108
Gráfico 27: Tons médios dos estímulos por variedades regionais	109
Gráfico 28: Boxplot de dispersão de F0 e variedades regionais	109
Gráfico 29: Diferenças entre médias de dispersão de F0 entre as variedades regionais – Tukey	110
Quadro 1: Relação de características por informante	72
Quadro 2: Resumo das análises	111
Tabela 1: Dados do Censo Demográfico do IBGE	45
Tabela 2: Exemplo de organização dos dados em TD	78
Tabela 3: Resultados <i>d'</i> obtidos no experimento 1 (PURR) para cada participante de Senador Pompeu (CE)	87
Tabela 4: Resultados <i>d'</i> obtidos no experimento 1 (PURR) para cada participante de São Paulo (SP)	88
Tabela 5: Resultados <i>d'</i> obtidos no experimento 1 (PURR) para cada participante de Pelotas (RS)	88
Tabela 6: Resultados <i>d'</i> para curva melódica monotônica (CMM) obtidos para	

cada participante em cada região	92
Tabela 7: Desempenho – RS	94
Tabela 8: Desempenho – SP	95
Tabela 9: Desempenho – CE	96
Tabela 10: Resultado da contagem de alarmes falsos	116

1. INTRODUÇÃO

1.1 Apresentação do tema

Pretende-se, neste trabalho, tratar da variação regional em português brasileiro, pautando-se pelo viés perceptivo dos estudos da linguagem. A variabilidade é inerente às línguas naturais, seja no âmbito idioletal ou dialetal. Por essa razão, os estudos feitos pela sociolinguística empenharam seus esforços na documentação das variações existentes nas línguas. Por outro lado, os pesquisadores que trabalham com percepção dedicaram-se a eliminar essas variações ou ignorá-las por completo, apesar da possível relação entre os estudos perceptivos da fala e a sociolinguística. Um olhar diferente sobre essa questão é reconhecer que a variabilidade é uma consequência natural das línguas e que se deve buscar maneiras de entender como essa variação é processada na percepção da fala (CLOPPER; PISONI, 2004).

O que está em questão na eventual junção dessas abordagens é a disparidade de pressupostos, já que os pesquisadores da área de percepção de fala, que buscam uma regularidade, um modelo abstrato em que a percepção das categorias seja invariável, não se importam com a variação que essas categorias possam sofrer, sendo esta *“a source of noise [...] irrelevant to the underlying representations”* (CLOPPER; PISONI, 2004, p. 314). Dessa forma, toda e qualquer variação acerca de uma categoria central torna-se dispensável.

Estreitando a abordagem para os interesses desta pesquisa, o papel da prosódia na identificação das variedades regionais do português brasileiro está relacionado à maneira como os ouvintes fazem uso da informação contida no

sinal acústico e do conhecimento adquirido no contato com outros dialetos e discursos, de modo que essas informações sejam relevantes para a identificação da variedade produzida pelo falante.

Este trabalho pretende coadunar uma abordagem a serviço da outra, em outras palavras, trazer para os estudos de percepção da fala, mais especificamente os de percepção de prosódia, os fins encontrados nos estudos sociolinguísticos. Esse encontro tem gerado um vasto número de metodologias aplicadas aos estudos sociolinguísticos que se interessam pela representação mental das variações dialetais.¹

Assim, em meio à aparente impossibilidade de uma intersecção produtiva entre as duas abordagens, por um lado a que trata da variação linguística, por outro aquela que busca regularidades no âmbito perceptual, propõe-se uma pesquisa que esteja inserida na intersecção dos estudos de percepção, especificamente, entre o entendimento de como a fala pode ser percebida e suas implicações no reconhecimento das variedades regionais do português brasileiro. Nota-se, como brevemente exposto acima, que tal tarefa envolve diferentes pontos de vista com relação à abordagem que se dá ao objeto língua, desde seu papel como reveladora de uma informação linguística até o seu papel social como identificadora de diferenças regionais.

A abordagem perceptual, neste trabalho, está restrita ao papel da prosódia como pista para o reconhecimento dialetal. Os trabalhos de Atkinsons (1968), Bonte (1975), Maidment (1976) e Ohala e Gilbert (1978) já apontaram para a importância da prosódia como pista no reconhecimento de línguas. A motivação desses trabalhos foi a observação ordinária de que as línguas diferem umas das outras por seu inventário segmental diverso e por suas

¹ Dentre as metodologias aplicadas, pode-se elencar: *map-drawing task*, *attitude judgments*, *speech caricatures*, *vowel matching* e *matched-guise technique*.

restrições fonotáticas variadas, além de possuírem critérios de formação de palavras e construção de sentenças muito diferentes entre si, mesmo que essas diferenças sejam mais tênues entre algumas línguas e mais salientes entre outras.

Com esse mesmo intuito, outros estudos abordaram o reconhecimento prosódico intralinguístico, como no estudo de Bush (1967), no qual foram estudadas 3 variedades do inglês, a saber, a variedade britânica, a americana e a indiana; e o de Richardson (1973), no qual foram estudados o *Black* e o *White English*.² Esses estudos demonstraram que a busca de pistas prosódicas em estudos perceptuais pode ser estendida também para as variações dialetais. As características contidas nas variações de frequência fundamental, amplitude e duração podem desempenhar o papel de sinalizadores de diversos tipos de variação linguística, sejam elas de caráter diatópico, diafásico ou diastrático.

Outros estudos sobre categorização dialetal foram feitos, principalmente na área de psicologia cognitiva e psicolinguística, com intuito de implementar metodologias que determinem se os ouvintes são capazes de identificar a qual região ou variedade o falante pertence (GROSJEAN; FRAUENFELDER, 1997). É importante salientar que esses estudos não conseguiram dar respostas sobre como os ouvintes usam o seu conhecimento de variação para julgar a variedade à qual o falante pertence, mas somente puderam constatar que, de alguma forma, os ouvintes conseguem determinar de qual variedade o falante faz parte (CLOPPER; PISONI, 2004).

Pode-se concluir, pois, que esses estudos carecem de uma metodologia que possa determinar ou, pelo menos, apontar indícios de quais informações

² Há um terceiro estudo conduzido por Bezooijen e Gooskens (1999) ao qual dar-se-á a devida atenção no subcapítulo 3.3.

contidas no sinal acústico são relevantes para a tomada de decisões em experimentos perceptuais de fala. Uma alternativa para resolver esse tipo de problema seria acrescentar aos trabalhos que lidam com percepção uma análise de produção do sinal de fala utilizado nos experimentos:

In combination with acoustic analyses of the speech signal and/or synthetic manipulation of the speech to highlight certain features, these kinds of studies can also be used to answer basic questions about which acoustic properties of the speech signal are perceptually most salient to listeners in identifying a talker's dialect. (CLOPPER; PISONI, 2004, p. 320).

Outra característica dos estudos de reconhecimento dialetal levantada por Vaissière e Boula de Mareüil (2004) é a diminuta atenção dada à contribuição da entoação para a percepção de uma variedade regional. Os autores defendem que muitas variações dialetais percebidas são frutos da combinação entre a entoação e a informação segmental, mas que podem não ser reconhecidas quando apresentadas separadamente.

Pautado na hipótese do reconhecimento linguístico baseado nas informações prosódicas, como nos trabalhos supracitados, esta pesquisa visa a desenvolver um estudo perceptual no qual se possa testar a habilidade dos falantes de português na tarefa de reconhecimento de sua variedade linguística, e, com o propósito de saber quais informações são relevantes para tal reconhecimento, um estudo de produção se faz necessário para poder avaliar quais informações podem estar interferindo na escolha dos ouvintes durante os experimentos de percepção.

Portanto, esta pesquisa perpassa os estudos de percepção e produção da fala, tendo como motivação a variabilidade encontrada no português brasileiro.

1.2 Objetivos e tarefas realizadas neste estudo

Os objetivos deste trabalho são saber qual o papel da prosódia na delimitação das variedades linguísticas regionais do Brasil e qual seria a pista que os ouvintes procuram, em cada uma das variedades, para reconhecer a sua própria.

Na parte inicial, será abordada a ocupação do território brasileiro e a formação das variedades regionais, buscando justificar a escolha das variedades estudadas neste trabalho.

Após, foi aplicado um teste perceptual (experimento 1) sobre o papel da prosódia no reconhecimento de variedades do português brasileiro. As três variedades estudadas são: a variedade de São Paulo, especificamente a do bairro da Mooca; a variedade do Ceará, da cidade de Senador Pompeu, e a variedade do Rio Grande do Sul, da cidade de Pelotas. No experimento 1, foi empregado o método de delexicalização PURR.³ Os detalhes do método e motivos pelos quais o PURR foi utilizado serão abordados na metodologia deste trabalho (capítulo 4).

Após o experimento 1, foi aplicado um teste perceptual (experimento 2) por meio de uma curva melódica monotônica (CMM) formada a partir do tom médio gerado pela rotina de análise automática da fala — *Exprosodia*®⁴ — e por meio da função *To manipulation*, do Praat (BOERSMA; WEENINK, 1996). Os detalhes da feitura desse experimento serão abordados na metodologia (capítulo 4). Com os resultados dos experimentos perceptuais em mãos, foi

³ PURR - *Prosody Unveiling through Restricted Representation*.

⁴ *ExProsodia* é uma rotina elaborada por Ferreira Netto (2010) em *Visual Basic* e tem por objetivo a análise automática de frequência, intensidade e duração de trechos de enunciados. A Universidade de São Paulo é detentora da patente.

feita uma análise dos dados obtidos via Teoria da Detecção,⁵ bem como a análise estatística dos dados.

Após a etapa perceptiva deste trabalho, foi feita uma análise dos dados de produção obtidos via *ExProsodia*. Com as duas partes analisadas, percepção e produção de fala, foi avaliada a sua relação na identificação das variedades regionais.

⁵ Macmillan e Creelman (2005).

2. HISTÓRICO E FORMAÇÃO DO PORTUGUÊS BRASILEIRO

2.1 Ocupação e características

Abordar a história de qualquer língua, seja ela vernácula ou não, encerra um trabalho exaustivo que não se pode abarcar de uma única vez. Quando anuncia-se, neste capítulo, fazer um histórico, tem-se em mente a possibilidade de traçar linhas gerais que possam justificar as diferenças que, de outros tempos até os dias de hoje, são notadas entre as variedades estudadas neste trabalho.

A língua portuguesa trazida no século XVI e em subsequentes incursões dos portugueses no Brasil já não era um bloco, um todo uniforme, mas trazia variações que estavam presentes no português europeu. Fernão de Oliveira⁶ (1981 apud SILVA NETO, 1986, p. 42) diz que “os homens de Estremadura são diferentes dos D’antre Douro e Minho: porque assi como os tempos, assi também as terras criam diversas condições e conceitos”. Parece, então, natural que essa língua ganhe aqui, em terras brasileiras, novas nuances. A língua vinda com os portugueses começa a ser produzida por novos falantes e em novas situações sociais. Silva Neto muito bem ilustra essa concepção:

Mas a verdade é que a língua, longe de ser um organismo, é um produto social, é uma atividade do espírito humano. Não é, assim, independente da vontade do homem, porque o homem não é uma fôlha sêca ao sabor dos ventos veementes de uma fatalidade desconhecida e cega. (1986, p. 17)

A diversidade linguística encontrada no português brasileiro, do ponto de vista diatópico, é um indício forte de que o português incorporou elementos

⁶ OLIVEIRA, F. Gramática da linguagem portuguesa. 1ª ed. Lisboa : Biblioteca Nacional, 1981. Edição fac-similada.

autóctones. Castilho (2000) sintetiza as duas abordagens acerca da formação do português brasileiro:

1. *A hipótese de mudança natural, ou hipótese internalista* – identificação de tensões no interior da estrutura linguística, ou seja, a língua segue a sua deriva natural, sem motivação externa provocada pelos contatos linguísticos do português europeu.
2. *A hipótese crioulística* – a língua possivelmente sofreu influências fonológicas e gramaticais do substrato indígena e de línguas africanas aqui chegadas durante a escravidão.

De fato, as duas abordagens são profícuas. A segunda delas trabalha em conjunto com a demografia histórica, ou seja, os caminhos que os colonizadores percorreram ao ocupar o território nacional e os contatos linguísticos resultantes dessa ocupação. A união da demografia e do resultado linguístico dos contatos seria o ideal para acessar o que de fato ocorreu, mas há muito pouco interesse da historiografia com relação às características linguísticas: “os historiadores, por via de regra, prestam pouquíssima atenção às maneiras de falar dos povos sobre os quais escrevem” (KIENAN, 1993⁷ apud MATTOS e SILVA, 2001, p. 281).

A região de onde os portugueses provieram é uma das justificativas para a diversidade do português brasileiro, mas, como o acesso a dados demográficos da metrópole naquela época é muito difícil, torna-se complicado precisar qual a sua influência na diversidade linguística brasileira. Mattos e Silva (1998, p. 47) diz que “é muito complexa a relação entre o português

⁷ KIENAN, V. Língua e conquistadores. In: BURKE, F.; PORTER, R. (Orgs.), p. 258-84, 1993.

brasileiro e o europeu, e não se reduz à simplicidade com que tem sido formulada desde Serafim da Silva Neto”, colocando maior peso na origem regional do colonizador. Noll afirma:

As semelhanças evidentes que existem fizeram pensar, conseqüentemente, numa influência atuante, especificamente meridional, na formação do português brasileiro, que se sobressaía em relação à dos dialetos do Norte e do centro de Portugal. (2008, p. 281)

Adiante, Noll afirma que “não há dados que comprovem uma imigração preponderante no Brasil, proveniente do Sul de Portugal” (2008, p. 283). Outras diferenças encontradas entre o português europeu e o brasileiro somente ocorreram no século XIX, partindo de Lisboa (NOLL, 2008, p. 283).

O que há documentado são núcleos primários (RIBEIRO, 1914) estabelecidos pelos portugueses no litoral brasileiro. Silva Neto os enumera da seguinte maneira:

1. Maranhão e Pará, que geram as unidades administrativas do extremo Norte;
2. Pernambuco, que gera os núcleos secundários da Paraíba, Rio Grande do Norte, Ceará e Alagoas;
3. Bahia, que absorve Ilhéus e Porto Seguro e gera Sergipe;
4. Rio de Janeiro, que em 1658 se emancipou da Bahia, englobava a Paraíba do Sul e tinha superior jurisdição sobre o Espírito Santo, São Vicente e Santo Amaro;
5. São Paulo, donde evoluiu todo o Oeste (Goiás emancipado em 1744, Minas em 1720, Mato Grosso em 1748).

(1986, p. 134-5)

A incursão de São Paulo para o Sul (PR, SC e RS) deu-se a partir do século XVIII. Em uma classificação cronológica, Castilho enumera o povoamento e a implantação do português em oito focos litorâneos:

1. 4 focos no século XVI: São Vicente/São Paulo (1532, 1554); Olinda/Recife (1535); Salvador (1549); Rio de Janeiro (1557);
2. 2 focos no século XVII: São Luís do Maranhão (1612) e Belém (1616);
3. 2 focos no século XVIII: Florianópolis (1738) e Porto Alegre (1752).

(2010, p. 174)

Nota-se, pois, que as variedades propostas para a análise deste trabalho estão em núcleos diferentes, na classificação de Silva Neto (1986), e aparecem em épocas distintas nos focos de ocupação cronologicamente organizados por Castilho, principalmente as variedades paulista e gaúcha. Adiante, poder-se-á notar que as regiões culturais propostas por Diegues (1960) também englobam, em regiões culturais distintas, cada uma das variedades estudadas neste trabalho.

Passando para abordagens focadas nas variações linguísticas encontradas em diferentes regiões do Brasil, o século XX teve trabalhos pioneiros como os de Amadeu Amaral (*O dialeto caipira*, 1920) e Antenor Nascentes (*O linguajar carioca*, 1953).

Nascentes (1953)⁸ propôs uma divisão dos falares regionais brasileiros. Esses seriam divididos em norte e sul e subdivididos em seis subfalares. O

⁸ NASCENTES, A. *O linguajar carioca*. Rio de Janeiro: Organização Simões, 1953.

critério adotado por ele foi a cadência e a distribuição de pretônicas abertas. O termo “cadência” adotado por Nascentes faz referência a um aspecto que não é segmental, morfológico ou sintático. Poder-se-ia, então, pensar em algo relacionado com o ritmo, com a taxa de elocução dos falantes ou como uma metáfora musical de encadeamento tonal. Assim, essa divisão pode ser esquematizada da seguinte maneira:

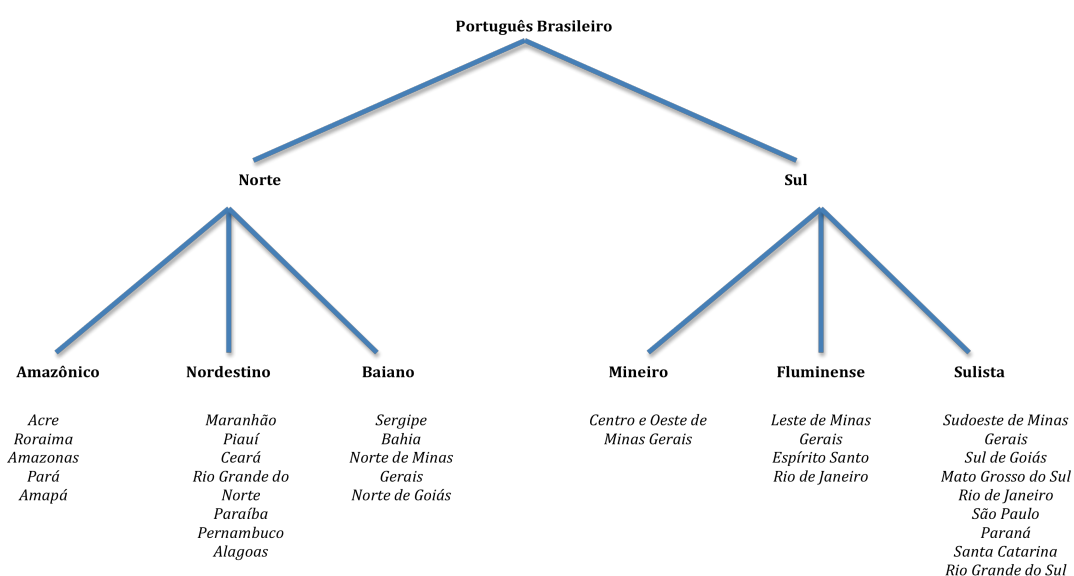


Figura 1: Esquema das subdivisões de Nascentes (1953) para o português brasileiro.

O mapa com as subdivisões propostas por Nascentes foi confrontado com as análises feitas por Noll (2008), o qual tratou, dentre outros fenômenos, da realização de /s/ final e de /s/ em coda silábica, da africativização de /t/ e /d/ e da realização de /r/ inicial e de /r/ em coda silábica. A distribuição dos fenômenos supracitados e a proposta de Nascentes estão nos mapas abaixo (figuras 2, 3, 4 e 5):



Figura 2: Divisão proposta por Nascentes (1953 apud NOLL, 2008).

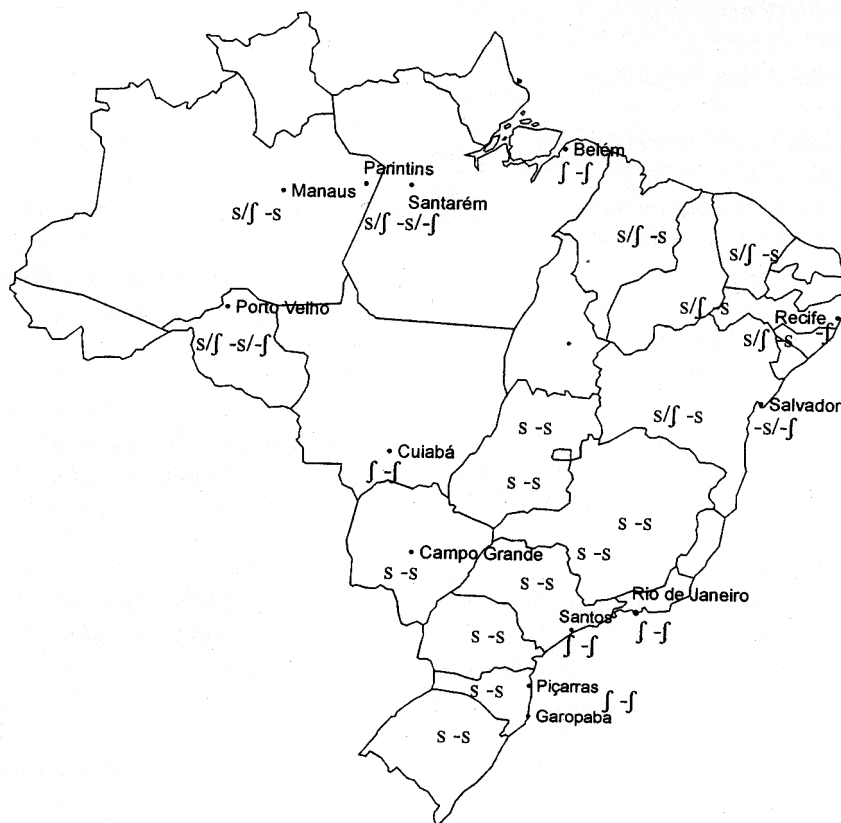


Figura 3: Realização de /s/ final e de /s/ em coda silábica (NOLL, 2008, p. 52)

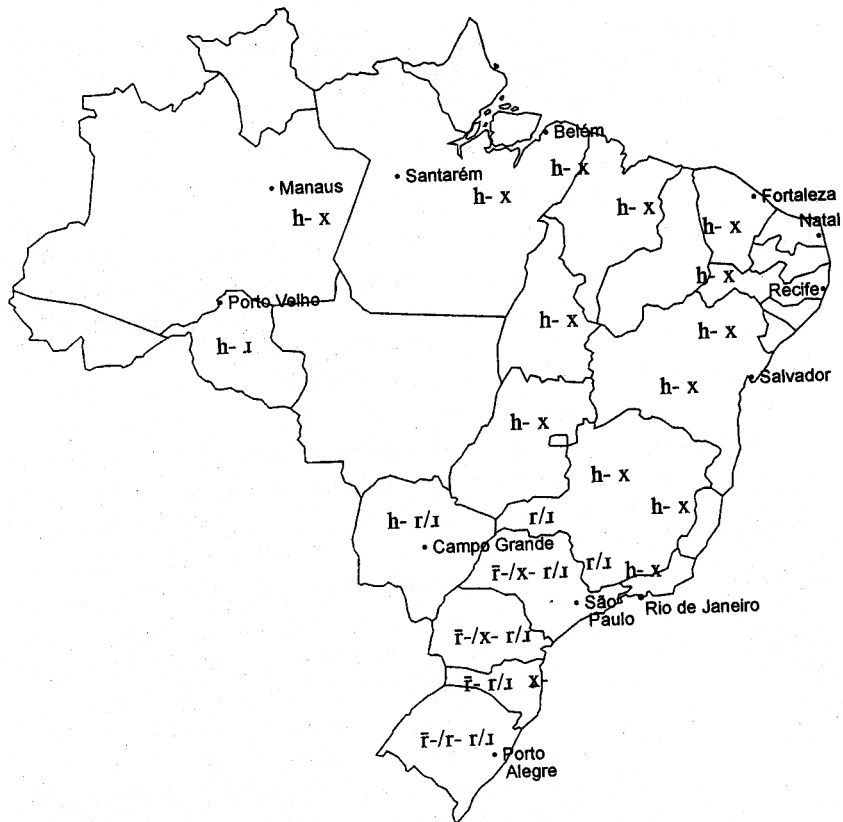


Figura 4: Realização de /r/ inicial e de /r/ em coda silábica (NOLL, 2008, p. 71).

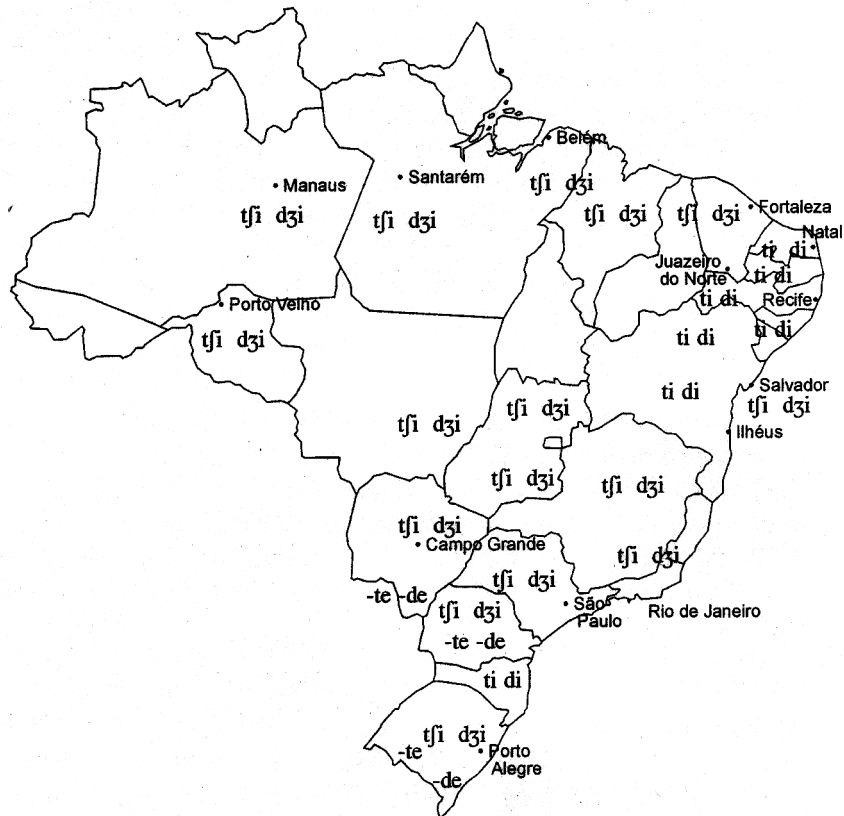


Figura 5: Africativização de /t/ e /d/ (NOLL, 2008, p. 68).

Comparando os mapas, somente a distribuição do /r/ parece obedecer quase que sistematicamente à proposta de Nascentes (1953 apud NOLL, 2008). Os demais fenômenos parecem destoar dessa divisão, a qual é apenas baseada na cadência e nas vogais /e/ e /o/ pretônicas. É perfeitamente compreensível que a divisão tenha sido abrangente dessa maneira, pois, à medida que se tomam outros parâmetros para essa categorização, a divisão acaba tomando outra forma. Quando temos [s] e [ʃ] em posição de coda ou final de palavra, ou a africativização de /d/ diante de /i/, a divisão se torna diferente daquela proposta por Nascentes (1953).

Se a divisão, partindo de análises segmentais, se mostrou não tão pacífica, uma outra distribuição das variedades do português brasileiro baseada na prosódia seria tão problemática quanto a outra, pois não se sabe qual o papel que as línguas trazidas pelos colonizadores e imigrantes, bem como a influência das línguas locais, desempenhou nas diversas prosódias do português brasileiro. Se o português falado no Brasil tem, em geral, um desenho melódico característico (isolando-se somente as variações do parâmetro F0) alguns traços da fusão índio-afro-europeu possivelmente ficaram marcados, os quais prontamente diferenciam o português do Brasil daqueles falados em outros países lusófonos.

2.2 Sobre as regiões culturais

Nesta segunda parte da discussão, será feita uma tentativa para determinar se as variedades regionais abordadas no estudo perceptual são, de fato, distintas.

Em primeiro lugar, será feito um esboço da ocupação territorial do Brasil mediante estudo de Diegues (1960) sobre as regiões culturais brasileiras. A partir desses parâmetros, serão explicitados os conceitos de Comunidades de Língua (CL) e de Fala (CF).

Em seu trabalho, Diegues (1960), antes de traçar as regiões culturais brasileiras, afirma que é necessário pensar o Brasil como parte de um conjunto maior, de uma área cultural mais ampla, na qual estão inclusos Portugal e suas Províncias Ultramarinas. Essa área cultural luso-cristã, assim chamada, possui dois elementos caracterizadores: o elemento lusitano na aglutinação da população portuguesa e o cristianismo como base religiosa. Posto isso, é a integração desses elementos com as características autóctones que permite precisar traços comuns entre os vários países dessa grande área e, por outro lado, mantém a diversidade proveniente do ambiente natural e social de cada lugar.

A ocupação do território brasileiro não foi fruto de uma colonização bem definida, com metas e normas rígidas. Como consequência disso, nasce a possibilidade de pensar o Brasil como um conjunto de regiões culturais, pois cada uma delas é o reflexo das relações estabelecidas entre os povoadores, os povos nativos e a terra ocupada. Diegues afirma que é dessa diversidade que surge a unidade. Em consonância com essa característica, Silva Neto (1963, p. 70) diz que “o extenso território brasileiro é, na verdade, um arquipélago constituído de *ilhas* históricas: Maranhão ou Pará, Pernambuco, Bahia, Rio de Janeiro, São Paulo”. É importante ressaltar que essa delimitação não é discreta, ou seja, as ilhas históricas, nos termos de Silva Neto, e as regiões culturais, nos termos de Diegues, não são justapostas, mas são alteradas ao longo da ocupação.

Considerando-se essa afirmação com fins linguísticos, pode-se dizer que, de fato, lugares que não apresentam muito contato com outras regiões de uso linguístico diferente mantêm ‘estagnadas’ algumas características que, muito possivelmente, poderiam se resolver diante da diversidade a qual seriam expostas em espaços urbanos.

A variação regional parece não depender, em termos espaciais, do isolamento, de linhas bem demarcadas que separam as variedades. A alternância de bolsões linguísticos isolados e espaços urbanos parece ser uma das possíveis razões para a diferenciação gradativa dos falares; os primeiros pela baixa circulação de elementos externos, e os segundos pelo cosmopolitismo ao qual estão expostos e dispostos a vivenciar.

Para Diegues (1960), uma região cultural seria um conjunto ecológico de pessoas, agregadas pela similaridade das relações sociais, pela unidade das relações espaciais da população, somando a essas a estrutura econômica local. Do ponto de vista linguístico, Mufwene (2007) também defende que as características linguísticas são frutos da ecologia na qual elas foram geradas. Mas é importante dizer que o foco de Mufwene é a distinção entre as mudanças linguísticas interna e externamente motivadas, propondo que tal distinção tem bases ideológicas, mormente quando se trata do estabelecimento de um conceito sobre o que seriam as línguas crioulas. No caso de uma possível comparação entre o uso do termo “ecológico” feito por Diegues e a abordagem de Mufwene, haveria uma intersecção no fato de que as características tanto culturais quanto linguísticas são frutos da ecologia interna de cada região.

Dessa forma, as regiões culturais seriam espaços territoriais delimitados por certas características que unificam as ideias, os sentimentos e o estilo de

vida de um grupo populacional (DIEGUES, 1960, p. 7). Num breve parêntese, direcionando a discussão para uma abordagem sociolinguística, tais regiões também compartilhariam traços linguísticos particulares, pois o papel da interação social exerce grande influência na variação e mudança linguísticas. Labov (1972) considera que não há como tratar das diferenças linguísticas, sejam elas numa abordagem diacrônica – em caso de mudança –, ou sincrônica – em caso de variação – sem levar em conta a vida social da comunidade e as pressões sociais existentes nela.

Parece, então, que o pressuposto para a categorização tanto sociocultural como sociolinguística faz uso das relações e dos contatos externos e internos para compor uma região cultural, na concepção de Diegues, e uma comunidade de fala (CF), na visão de Labov. Essa discussão será retomada adiante quando os conceitos de comunidade de língua (CL) e fala (CF) forem abordados.

O estudo da formação de uma área cultural necessita de uma colaboração interdisciplinar que envolva dados históricos, geográficos, antropológicos, sociológicos e linguísticos.

A partir da ocupação do território brasileiro, Diegues (1960) subdivide o Brasil nas seguintes regiões culturais: 1) Nordeste Agrário do Litoral; 2) Nordeste Mediterrâneo; 3) Amazônia; 4) Mineração do Planalto; 5) Centro-Oeste; 6) Pastoril do Extremo Sul; 7) Colonização Estrangeira; 8) Café; 9) Cacau; 10) Sal.

No estudo proposto, serão tratadas as regiões 2, 6 e 8, demarcadas no mapa abaixo:

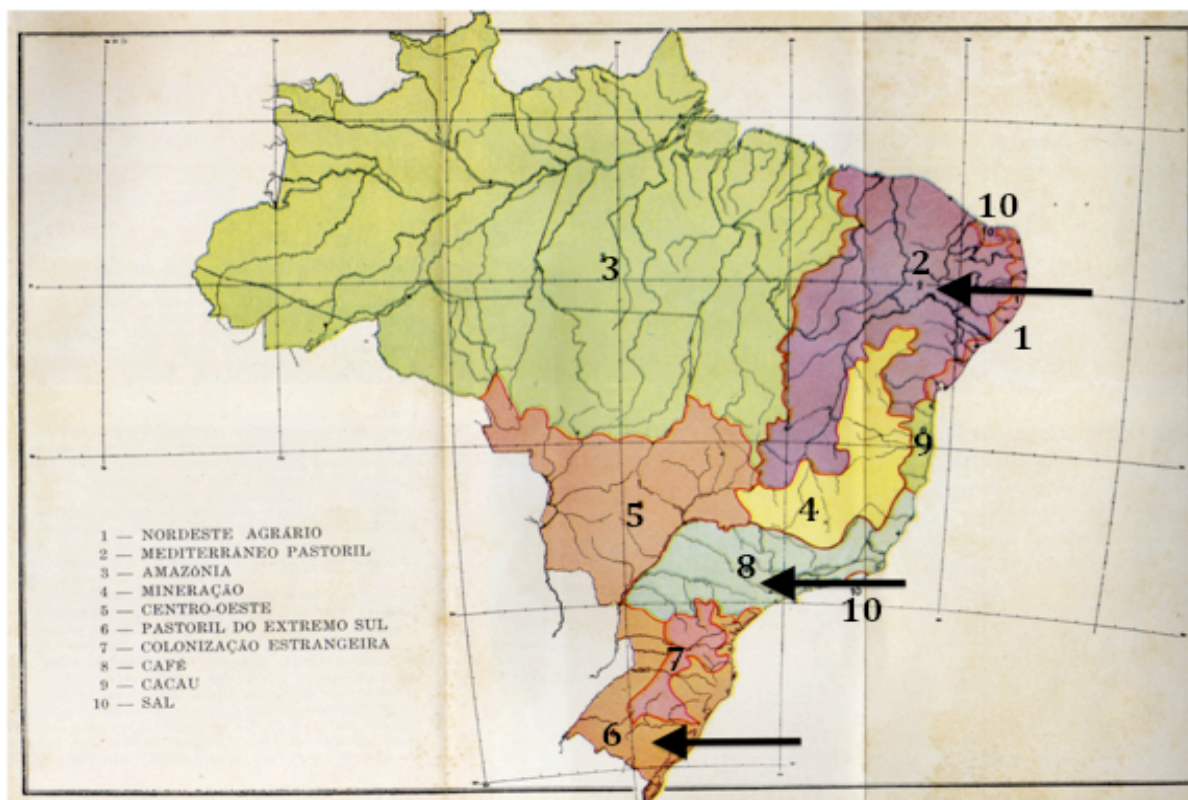


Figura 6: Mapa das regiões culturais do Brasil (DIEGUES, 1960, encarte).

De maneira sucinta, as regiões culturais estudadas têm como traços formadores da sua ocupação as seguintes características:

- i) **Nordeste Mediterrâneo** - é fruto da expansão da ocupação litorânea, com intensa mestiçagem entre brancos e índios. As fazendas de criação e os currais formam o principal centro social da região.
- ii) **Pastoril do Extremo Sul** - a sua formação é decorrente da expansão das correntes paulistas, nordestinas e fluminenses no século XVIII. A pecuária é a principal atividade econômica da região, tendo a estância⁹ como núcleo social. Os

⁹ Fazenda de criação de gado.

espanholismos presentes no falar regional são fruto do contato com as comunidades vizinhas de língua espanhola.

- iii) **Região do Café** - formada pela expansão dos cafezais no século XIX, contemplando duas fases de mão de obra, a primeira baseada na escravidão e a segunda, no trabalho livre dos imigrantes. O centro social e político é a fazenda de café.

Comparando a divisão proposta por Nascentes (1953 apud NOLL, 2008) às regiões culturais de Diegues (1960), estas são mais estritas do ponto de vista territorial, pois, partindo da divisão Norte e Sul de Nascentes, inevitavelmente, ter-se-ia de juntar em uma só área duas variedades diferentes, como é o caso de São Paulo e Rio Grande do Sul. Mesmo baseando-se nos subfalares propostos pelo autor, ainda assim o falar de São Paulo não teria uma classificação, passando ao largo dos subfalares mineiro e sulista.

2.3 Sobre as Comunidades de Língua (CL) e Comunidades de Fala (CF)

Os conceitos de Comunidade de Língua (CL) vs Comunidade de Fala (CF) requerem uma breve retomada da dicotomia saussureana *língua* e *fala*. Para assumir essa bipartição, está pressuposta, por um lado, a visão de língua como um sistema coerente e coeso — no qual suas partes seriam agrupadas em conjuntos por semelhança funcional e formal — e, por outro lado, as infinitas realizações possíveis dentro desse sistema.

A ideia de CL está fundamentada no conceito de língua como um sistema supraindividual, um bem compartilhado socialmente. A CL, pois, compreende o que denominamos largamente de língua. A CL tratada neste estudo é a de língua portuguesa, que corresponde a Portugal, Brasil, Angola, Moçambique, Cabo Verde, Guiné-Bissau, São Tomé e Príncipe, Timor Leste e demais regiões em que a língua portuguesa possa ser utilizada. Cabem aqui algumas considerações, antes de afirmar uma relação unívoca entre o conceito de língua saussureano e CL, pois, Saussure (2003, p. 224) afirma que a separação geográfica é um fator importante na diversificação linguística, ao passo que o conceito de CL não leva em consideração a contiguidade territorial para a sua demarcação. Por outro lado, a delimitação geográfica é preponderante na conceptualização das CF, a qual será abordada com maior atenção adiante.

Ainda sobre o espaço geográfico, Saussure (2003) afirma que, apesar da maior saliência linguística notada nas diversidades geográficas, é o tempo que exerce papel essencial na acentuação das diferenças. Se uma língua 'L' for transportada para um outro território, com o passar do tempo, haverá diferenças de pronúncia, gramática etc., tornando-se outra variedade, a saber 'L*'. Para Saussure é, pois, o tempo que implementa mudanças nas línguas, delegando à distribuição territorial papel secundário:

Que é que cria essas diferenças? Quando se acredita que seja unicamente o espaço, é-se vítima de uma ilusão. Por si só, o espaço não pode exercer nenhuma ação sobre a língua. No dia seguinte ao do seu desembarque em L', os colonos saídos de L falavam exatamente a mesma língua da véspera. Esquece-se o fator tempo, porque é menos concreto que o espaço; na realidade, porém, é dele que releva a diferenciação linguística. A diversidade geográfica deve traduzir-se em diversidade temporal. (2003, p. 229)

Para reforçar esse conceito, Saussure afirma que não se pode medir um volume qualquer pela sua superfície. De maneira análoga, a diferenciação geográfica não se completa quando não se considera a dimensão temporal. Assim, o papel da ocupação territorial proposto por Diegues (1960) só é relevante, numa visão saussureana, quando se projeta o tempo, considerando a acomodação da língua e dos costumes levados para determinada região, bem como as influências que as culturas autóctones exercem sobre eles.

Direcionando essa discussão para um ambiente unilíngue, mesmo sendo do conhecimento geral que tal afirmação não se aplica *stricto sensu* à realidade brasileira (por conta das diversas línguas faladas paralelamente ao português), pode-se vislumbrar os efeitos da ação do tempo num território contínuo como o brasileiro, considerando os dois desdobramentos possíveis propostos por Saussure (2003):

- i) A língua não será mais a mesma, pois não existe imobilidade absoluta em termos de linguagem; ao cabo de um certo tempo, a língua não será mais idêntica a si mesma;
- ii) As mudanças não serão uniformes em todo o território, mas apresentarão variações de acordo com os lugares.

A não uniformidade das mudanças provoca a existência de áreas distintas que justificam as diversas maneiras de falar em todos os pontos do domínio de uma língua (SAUSSURE, 2003, p. 232). Para elucidar essa afirmação, Saussure cita o exemplo do viajante que, ao longo de sua travessia por um país qualquer, não nota diferenças entre localidades vizinhas, mas, ao

passo que avança em sua viagem, começa a perceber que o acúmulo dos limites dialetais dá início a um abismo de ininteligibilidade.

Saussure não leva em conta os fatos sociais para formular sua teoria sobre a ação do tempo e a delimitação de línguas e dialetos. Apesar de afirmar que as semelhanças linguísticas, quando numericamente importantes, “se explicam por fatos sociais, políticos, religiosos etc.”, o autor faz questão de enfatizar a “total abstração” desses fatores (SAUSSURE, 2003, p. 235). Isso se torna um problema do ponto de vista dos estudos de variação linguística, os quais buscam nas minúcias do ambiente social uma parcela importante das explicações para as particularidades de cada variedade linguística.

Ao tratar do conceito de CF, é possível, numa primeira análise, notar que o cunho geográfico do termo é bastante importante para a sua definição. Essa classificação foi usada por Labov (1989) para denominar grandes comunidades linguísticas localizadas em centros urbanos, no caso, a Filadélfia. Para o linguista, uma CF é definida pela participação de seus integrantes num conjunto de normas compartilhadas pautadas pela uniformidade dos padrões abstratos de variação, levando-se em consideração o conceito de língua como um comportamento social (LABOV, 1972).

A dimensão areal das CF também é variável, abrangendo também pequenos agrupamentos urbanos, os quais carregam algum tipo de particularidade em sua fala (FEAGIN, 1996). O termo também foi usado para nomear minorias linguísticas como os falantes de francês de Ontário, no Canadá (MOUGEON; BENIAK, 1996).¹⁰

As raízes da CF, bem como as da CL, estão fixadas na sociolinguística, linguística histórica, filosofia da linguagem, dialetologia, antropologia e início do

¹⁰ Para tratar desse assunto, Patrick (2004) faz uma breve historiografia do conceito de CF, afirmando que há um vazio de análise sobre as aplicações desse conceito, bem como uma grande variação em sua definição.

estruturalismo. Segundo Hymes e Fought (1981, p. 81), Boas diz que as línguas refletem a história e a cultura da comunidade da qual elas têm sido uma parte. De maneira análoga, Sapir (1971) afirma que a fala é uma herança puramente histórica do grupo, o produto a longo prazo de um uso social continuado. Parece, pois, que tais afirmações estão de acordo com a visão de Diegues (1960), ou seja, que o produto cultural de uma determinada área, no qual a língua também está inserida, é um construto histórico de seus ocupantes.

A abordagem clássica trata uma CF como uma distribuição linguística dentro de um espaço social ou geográfico (GUMPERZ, 1968). Outra definição mais recente trata uma CF como “uma área geográfica delimitada por critérios não linguísticos como a demografia ou fronteiras sócio-políticas” (KERSWILL, 1994, p. 23). A antropologia e a geografia dialetal também usam como critério as delimitações territoriais. Já para Bloomfield (1933), dentro de um viés linguístico, uma CF seria um grupo de pessoas que interagem por meio da fala. Neste trabalho, a demarcação das áreas estudadas obedece à distância geográfica, à história, ao povoamento e ao conjunto de características linguísticas vistos em Nascentes (1953), Silva Neto (1963), Noll (2008) e Castilho (2010) e demais trabalhos.

Após perpassar brevemente as considerações saussureanas acerca da distribuição das línguas em território contínuo e unilíngue, bem como as definições de CF, pode-se, então, estabelecer alguns critérios sobre a definição do que seria uma variedade linguística. Até o dado momento, as regiões que carregam algum traço linguístico diferenciador estão sendo tratadas como delimitadoras de variedades de uma mesma língua.

A definição assumida por Chambers e Trudgill (1980) é a de que variedades regionais (dialetos, nos termos dos autores) são tidas como subdivisões de uma determinada língua. Com essa definição, pode-se assumir, por hora, que o português também possui variedades com marcas distintas. Para a manutenção dessa afirmação, é importante a existência de um eixo centralizador que mantenha a inteligibilidade mútua entre as variedades, aliado, concomitantemente, às diferenças que configuram as singularidades entre elas. Portanto, as variedades paulista, gaúcha e cearense, bem como outras não abordadas neste trabalho, partilham em quase sua totalidade essa definição.

Após as sucintas exposições feitas acima, é possível tecer algumas considerações acerca dos conceitos tratados neste subcapítulo. Retomando a visão de Diegues (1960) de que as regiões culturais sejam fruto de sua ocupação, ou seja, que o produto cultural é o reflexo das relações estabelecidas entre povoadores e povos nativos de determinada área ocupada, pode-se considerar que a língua está estreitamente relacionada com esse processo, sendo um construto histórico de seus ocupantes. Como pontos de convergência entre os conceitos de região cultural e CF, a categorização sociocultural e sociolinguística abarca as relações e os contatos internos e externos para a sua composição.

Considera-se, então, que as variedades do português brasileiro inseridas nas regiões culturais propostas por Diegues ou que compartilham um determinado conjunto de normas específicas – que as tornam particulares (com relação à CL portuguesa) ou uniformes (quando dentro dos limites geográficos da CF) – são realidades linguísticas diferentes e frutos de comportamentos sociais distintos (LABOV, 1972).

2.4 Sobre as regiões estudadas

Após as considerações feitas sobre a delimitação das variedades regionais, é necessário um breve histórico de cada uma das regiões estudadas neste trabalho.

A denominação do município de Pelotas (RS) teve origem nas embarcações de varas de corticeira forradas de couro que eram utilizadas para cruzar os rios na época das charqueadas. Em 1758, Gomes Freire de Andrade, Conde de Bobadela, doou ao Coronel Thomaz Luiz Osório as terras que ficavam às margens da Lagoa dos Patos. Em 1780, José Pinto Martins deixou o Ceará por conta da seca e fundou a primeira charqueada do Rio Grande do Sul. O povoado cresceu bastante com a chegada de muitas famílias. A fundação, em 7 de julho de 1812, da Freguesia de São Francisco de Paula ocorreu por provisão eclesiástica do padre Pedro Pereira de Mesquita. A área da freguesia era uma parte do território de São Pedro do Rio Grande. Com o progresso da região, a freguesia foi elevada à categoria de Vila, em 1832. Após a sua criação, a vila de São Francisco de Paula foi elevada à categoria de cidade (Lei Provincial n.º 5, de 27 de junho de 1835). Tempos depois, o deputado Francisco Xavier Pereira sugeriu o nome Pelotas ao recém-criado município.¹¹

Senador Pompeu (CE) começou sua história pela concessão de terras aos desbravadores, os pioneiros do Ceará-Grande, que construíram casas e dominaram os nativos. Nos séculos XVII e XVIII houve várias concessões de

¹¹ Fonte: IBGE.

terras às margens dos rios Banabuiú e Codiá, feitas pelos capitães-mores. A mais importante delas para a formação da cidade foi a outorgada, em 27 de março de 1723, aos desbravadores Thomé Callado Galvão e Nicolau de Souza. Nelas, com três léguas para cada um, às margens do rio Codiá, teve início a cidade de Senador Pompeu. Em 3 de setembro de 1896, foi assinada a Lei nº 332, que criou o município.¹²

No caso de São Paulo, essa tarefa torna-se um pouco mais complicada, pois sabe-se que a sua formação é muito complexa, estendendo-se até os dias atuais. Queiroz¹³ (1992 apud CASTILHO, 2000) sintetiza a ocupação de São Paulo em seis momentos:

1. Século XVI – Fundação da cidade e a ocupação do planalto;
2. De 1600 a 1765 – Período de irradiação do bandeirismo;
3. De 1765 a 1870 – Urbanização de São Paulo;
4. De 1875 a 1930 – Início da economia cafeeira, motivação para transformações políticas e sociais da cidade;
5. De 1930 a 1960 – Industrialização e forte crescimento da cidade.
6. De 1960 a 1990 – Consolidação da São Paulo metrópole.

(2000, p. 341)

Os momentos de 3 a 6 são muito importantes para a formação do português falado em São Paulo. Castilho afirma que:

As profundas alterações na sociedade de São Paulo nos séculos XIX e XX, motivadas pelas migrações externa e interna, pela intensa urbanização do Estado e pela metropolização de sua capital, entre outras causas, individualizaram a variedade paulista do português brasileiro. (2000, p. 335)

¹² Fonte: IBGE.

¹³ QUEIROZ, S. R. R. *São Paulo*. Madrid: Fundacion Mapfre América.

Não obstante, individualizar uma variedade da cidade de São Paulo é uma tarefa problemática, pois, em uma cidade com quase 12 milhões habitantes e destino de muitas migrações, é muito provável que a heterogeneidade no seu falar seja muito presente. Para este trabalho, a variedade da Mooca foi escolhida por ter sido o foco das migrações, por ser um dos primeiros bairros de São Paulo, com fundação em 17 de agosto de 1556, e por ter uma colônia forte de italianos, seguidos de espanhóis, portugueses, lituanos entre outros. A migração interna, especialmente do nordeste brasileiro, também é bem representativa no bairro. A tabela abaixo mostra os dados de área, população e densidade demográfica das três regiões:

	Senador Pompeu (CE)	Pelotas (RS)	Mooca – São Paulo (SP)
População	26.469	328.275	63.133
Área	1 002,127 km ²	1 608,768 km ²	7,7 km ²
Densidade demográfica	26,44 hab./km ²	203,74 hab./km ²	81,99 hab./km ²

Tabela 1: Dados do Censo Demográfico do IBGE de 2010.

Pode-se notar, pois, que as regiões estudadas possuem diferenças no tocante à sua ocupação e história. Espera-se, na análise proposta para este trabalho, verificar se essas diferenças se mantêm na percepção da prosódia de cada região.

3. TEORIA GERAL

3.1 Estudos de prosódia

O interesse gerado pela prosódia remonta às raízes da versificação grega e latina. Pelo termo *προσῳδία* entende-se a melodia e a cadência que acompanha os sons da fala. No Renascimento, o termo *prosódia* passa a fazer parte da versificação como disciplina. No século XVIII, Steele (1969 [1775]), partindo da análise musical, descrevera a prosódia como variações tonais semelhantes aos desenhos melódicos encontrados na música, utilizando o quarto de tom para marcar as variações da entoação da fala. Com o advento do estruturalismo, algumas análises fizeram o caminho inverso, ou seja, o da linguística para a música, levando em conta os níveis de análise linguística para análise musical (NATTIEZ, 2004).

No século XX, as diferentes teorias linguísticas atribuíram às questões prosódicas importância científica. Jakobson (1972), apesar de não ter a atenção especialmente voltada para a prosódia, já trata a frase como “unidade de sentido que se superpõe às palavras, e todo processo fônico que indica seu limite, sua divisão ou a hierarquia de suas partes” (1972, p. 25). O acento vocálico é o delimitador lexical e determina quantos itens há na frase. A hierarquia entre o léxico da frase é dada pela diferença de força empregada em cada item lexical, sendo o grau mais alto a palavra de maior importância na frase.

Exemplificando a frase como unidade superposta às palavras, o autor menciona a entoação descendente da enunciação, ou cadência, a qual delimita o final de uma frase; e a entoação ascendente, a qual indica que a mensagem

que está sendo veiculada pela frase ainda não terminou. Para ilustrar o papel que a prosódia desempenha nas línguas, mesmo que o termo não esteja explícito em suas colocações, Jakobson afirma que:

Os elementos fônicos que caracterizam uma frase - entoação frasal, acento frasal, pausa frasal etc. – só podem ter no âmbito da função representativa da linguagem um papel de dividir e de subordinar, ou ainda de coordenar. (1972, p. 26)

Jakobson (1972) afirma que esses elementos fônicos só servem para “limitar, articular e hierarquizar as significações, mas nunca para estabelecer relações semânticas”, papel atribuído por ele às palavras. Assim, há duas funções atribuídas aos processos fônicos:

1. *função distintiva* – processos fônicos distintivos ou fonemáticos.
2. *função articulativa* – processos fônicos articulativos – subdividida em:
 - a. *apical* – correspondente aos ápices.
 - b. *delimitativa* – correspondente às delimitações.

Na segunda função supracitada insere-se o papel da prosódia. Mas, poder-se-ia pensar o contrário de Jakobson, ou seja, na prosódia como integrante da função distintiva também, tomando-se, por exemplo, a distinção que algumas línguas fazem entre questões sim/não e asserções, as quais diferenciam-se somente pelo seu desenho melódico, dentre elas o português.

Quanto à tripartição dos signos feita por Jakobson, à luz de Saussure, os fatos prosódicos poderiam entrar nos signos de primeira classe.¹⁴

¹⁴ Jakobson (1972) estabeleceu uma escala dividida em três classes básicas de signos de acordo com o significado. Os signos de primeira classe funcionam na expressão como conteúdo. Dentre eles estão as frases, as palavras e os morfemas. Jakobson diz que “*deve-se incluir igualmente nessa mesma classe todos os meios que servem para o apelo e a manifestação psíquica. A plenitude afetiva, a expressão pessoal de quem manda a mensagem*”. Os signos de segunda classe são de natureza distintiva. Dentre eles estão as unidades de amplitude mais variável, como os fonemas. Os signos de terceira classe

Contrariamente à visão exclusivamente articulativa da prosódia, Julius de Laziczius (apud JAKOBSON, 1932, p. 37) chama de “emphatica” os processos vocais para a manifestação psíquica e o apelo. Os “elementos enfáticos”, segundo o autor, são os “signos providos de conteúdo”, ao contrário dos processos vocais tratados pela fonologia que “são exclusivamente signos de signos”. Para o linguista, a forma sonora dos “elementos enfáticos” designa diretamente o conteúdo.

A ideia de sílaba como portadora das particularidades prosódicas foi empreendida por Trubetzkoy (1970), afirmando que essas particularidades não dizem respeito somente às vogais em si, mas também às sílabas. À unidade mínima da prosódia Trubetzkoy chamou de “prosodema”, subdividindo-a em sílabas para línguas que contam sílabas e moras para as que contam moras. As línguas que contam sílabas diferenciam seus prosodemas por meio da intensidade e as que contam moras, por meio da altura musical (TRUBETZKOY, 1970, p. 180).

O linguista trata a entoação como sendo uma constante oposição de direções tonais, mas tal característica é aplicada somente a línguas que não possuem esse correlato para distinção lexical. Dessa maneira, haveria duas direções tonais: *ascendente* – função de continuidade; *descendente* – função conclusiva. A prosódia estaria subdividida em *entoação da frase, diferenças de registro na frase, acento frasal e pausa*.

Segundo Martinet (1970), a prosódia não tem valor pela sua presença ou ausência, mas pelas suas variações ao longo da fala. Essa abordagem torna o fato prosódico impossibilitado de ser analisado no plano paradigmático, causando um problema para a abordagem funcionalista, o qual tem por base a

funcionam como signo de outro signo, chamados por Jakobson de “signos de terceiro grau”. Essa divisão poderia ser dada da seguinte forma: [Letra (3º)] → [fonema (2º)] → [morfema (1º)] → [conteúdo].

comutação e a segmentação. Martinet aborda os signos prosódicos como marginais, pois “escapam à segunda articulação”, fato que impossibilita a inclusão deles na primeira articulação.

Incluem-se na prosódia todos os fatos de discurso que não entram na fonemática, quer dizer, os que de qualquer modo escapam à segunda articulação [...] nestas condições, não é difícil compreender que tais fatos não possam valer linguisticamente pela sua presença ou ausência em determinado ponto, mas sim pelas suas modalidades, variáveis de parte para parte dos enunciados, por consequência, não podem caracterizar unidades discretas. (MARTINET, 1970, p. 82)

Em seu estudo, Fónagy (2003) analisa a prosódia do ponto de vista das funções por ela desempenhadas. A contribuição de Fónagy, no tocante às funções exercidas pela entoação, foi a quebra dos sistemas ternários atribuídos à prosódia anteriormente. Dentre vários defensores de três funções entoacionais referidos pelo autor, podem-se citar Karl Bühler e Faure.

Fónagy afirma que a prosódia é a condição fundamental da interpretação da fala. O autor ilustra sua afirmação pela leitura de textos escritos que não tenham nenhum tipo de pontuação e intervalos, constatando que a falta da marcação prosódica nesse tipo de escrita demonstra a importância da prosódia pela sua ausência.

Fónagy (2003), então, atribui aos signos prosódicos algumas funções exercidas nas línguas. Dentre elas estão: *função demarcadora*, *função enfática*, *função sintática*, *função modal*, *função imitativa*, *função de chamamento*, *função lógica*, *função preditiva*, *função alusiva*, *função identificadora*, *função estética*, *função expressiva*, *função exploratória* e *função preparatória*. As funções postuladas por Fónagy estariam vinculadas às modalidades discursivas, obedecendo a uma hierarquia.

Dentre as funções expostas acima, nenhuma delas trata das características ou da possível relação que a prosódia possa ter com as variedades linguísticas. Talvez, num esforço de adaptação, coubesse na função identificadora esse tipo de função. Fónagy afirma que, por meio da função identificadora da prosódia, os informantes são capazes de determinar apenas por pistas prosódicas se se trata de um texto lido ou fala espontânea. Para tal, esses experimentos foram feitos com sinais laringográficos.¹⁵

Cabe maior discussão acerca da função de reconhecimento das variedades regionais que a prosódia pode desempenhar, não só em ambientes totalmente manipulados e incomuns, como os estímulos delexicalizados apresentados aos ouvintes, mas em fala ordinária, sem a intervenção mínima ou ostensiva do pesquisador.

Em categorização semelhante à feita por Fónagy, Vaissière (2004) também caracteriza a prosódia como portadora de múltiplas funções. Em seu estudo, a autora focaliza a entoação, mas não deixa de considerar a prosódia como um fenômeno complexo.

As funções citadas são: *sintática* – segmentação da fala em palavras prosódicas, sintagmas, sentenças; *informacional* – tema, rema, foco; *interativa* – mudança de turnos, continuação, final de tópico discursivo, atração de atenção; *modal* – asserção, questão, ordem; *atitudinal* – dúvida, descrença, polidez, ironia; *emocional* – prazer, raiva, tristeza; *outros* – identidade do falante, sexo, idade, variedades regionais, arcabouço sociocultural etc. (VAISSIÈRE, 2004). Nota-se que a autora atenta para o fato de que a prosódia desempenha a função discriminadora de variedades regionais.

¹⁵ Esse e outros métodos utilizados para isolar a prosódia serão tratados na metodologia deste trabalho.

Vaissière usa para a sua categorização as variações locais de F0, as quais são modificadas pelos órgãos da fala, subdivididos em subglotais, glotais e supraglotais. Em seu artigo “Les universaux de substance prosodique”, Vaissière (2007), ao tentar estabelecer quais seriam os universais de substância prosódica, define substância prosódica como um conjunto de variações na produção dos fonemas que transmitem informações diversas dos segmentos, sendo essas informações reiteradas ou modificadas. A figura abaixo resume a maneira como a autora organiza as substâncias prosódicas:

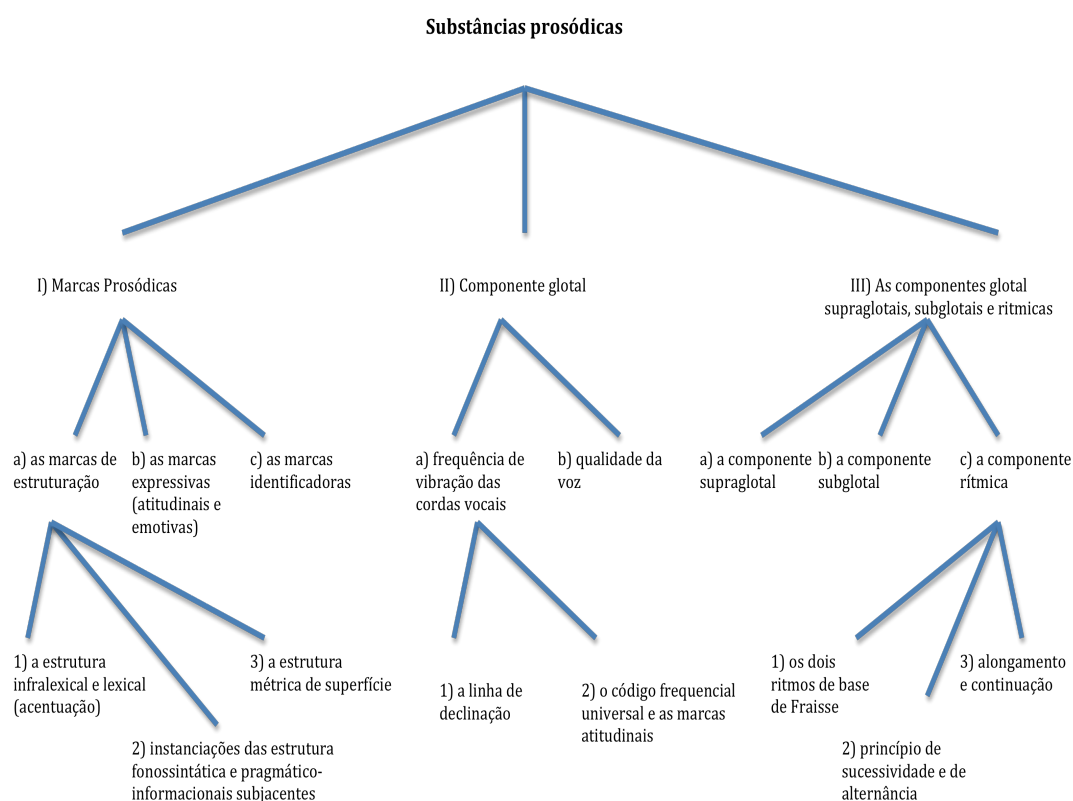


Figura 7: Substâncias prosódicas propostas por Vaissière (2002).

As línguas integram em sua gramática alguns desses itens prosódicos e os tornam parte da fonologia. Vaissière exemplifica com o golpe de glote, em alemão. Outros itens prosódicos não entram na fonologia da língua, mas

exercem influência nos esquemas fonológicos, perturbando-os ou reforçando-os. Como exemplo, há a palatalização como índice de fala infantilizada ou a faringalização como índice de rejeição. Segundo Vaissière (2002), essas são funções que são compreendidas sem aprendizagem.

Parece razoável, pois, que a determinação das variedades regionais seja dada pelos itens prosódicos que não entram na fonologia da língua. Dentre as marcas prosódicas expostas pela autora, há as marcas identificadoras, definidas como não controladas pelos falantes, tais como a informação sobre idade, sexo e origem regional, dentre outras.

Resta, para uma tentativa de esclarecer o que está envolvido nessa função, buscar saber qual seria a informação contida no sinal acústico que, inequivocamente, faz com que pessoas que compartilhem a mesma língua percebam as variedades regionais ou que, ao menos, não se reconheçam nelas.

3.2 Percepção e produção da fala

A importância dos trabalhos de percepção parte do pressuposto da fala como ato que possui um mecanismo duplo, envolvendo a produção e, no seu fim, a percepção. “Falar é um ato exclusivamente humano e, na qualidade de ato, um duplo mecanismo está envolvido: produção e percepção” (BARBOSA, 1995, p. 1). Este capítulo tratará dos trabalhos de percepção, desde a *gestalt* a trabalhos que tratam da percepção da fala em geral e da prosódia.

A psicologia da *gestalt* tratou da maneira como os humanos percebem os eventos no mundo, arranjando-os e dividindo-os, por meio de alguns princípios, de forma que sejam inteligíveis. Wertheimer afirma que:

When we are presented with a number of stimuli we do not as a rule experience “a number” of individual things, this one and that and that. Instead larger wholes separated from and related to one another are given in experience: their arrangement and division are concrete and definite. (1938 [1923], p. 1)

Os princípios propostos por Wertheimer foram a proximidade e a similaridade. Esses também abrangem os eventos auditivos, como o ritmo, tido como intervalo regular de pulsos,¹⁶ e a melodia, tida como uma coleção de semitons arranjados de maneira subsequente. Abaixo, tem-se duas figuras que ilustram tais princípios:

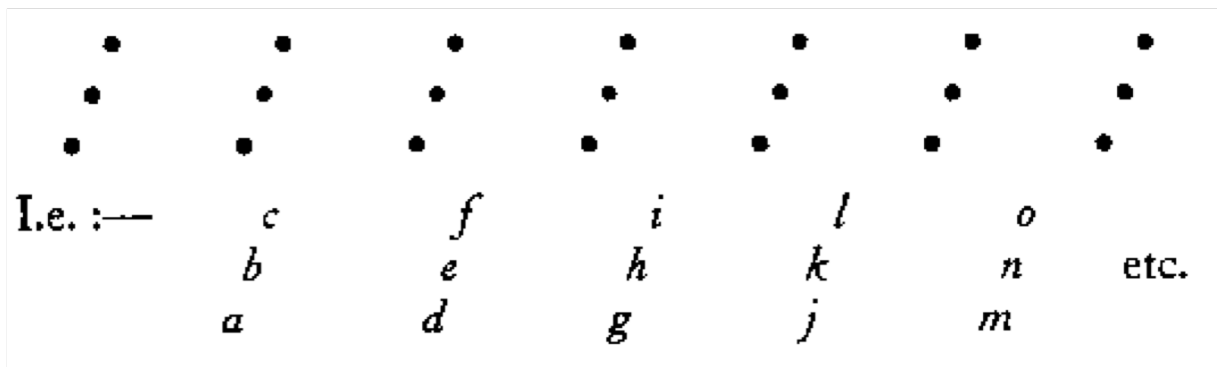


Figura 8: Ilustração dada em Wertheimer (1938 [1927], p. 2)

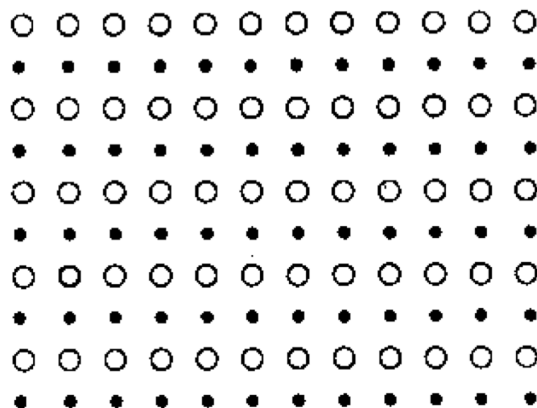


Figura 9: Ilustração dada em Wertheimer (1938 [1927], p. 4).

¹⁶ Na fala, a duração é o principal fator do ritmo, mas os demais parâmetros prosódicos também influenciam na percepção rítmica. Para Barbosa (1995), o ritmo é a variação a longo termo da duração percebida.

O fator de proximidade assume que dois objetos ou eventos que estejam próximos uns dos outros sejam percebidos mais facilmente como um conjunto só, ao passo que objetos dispostos a uma distância maior sejam percebidos como distintos uns dos outros. Assim, é possível notar, no exemplo com letras (figura 8), que a sequência mais natural a ser percebida seria *cba/fed*, ao contrário de *aei/dhl*. O fator de similaridade assume que dois objetos ou eventos similares entre si serão percebidos como equivalentes, ao passo que a dessemelhança entre eles faz com sejam percebidos como distintos. Na figura 9, nota-se mais naturalmente as fileiras horizontais, formadas ora por círculos, ora por pontos. Por outro lado, as fileiras verticais tornam-se não tão perceptíveis por serem constituídas por figuras diferentes.

Em relação aos estudos de percepção da fala, a partir da década de 1950, os interesses estavam concentrados em como a fala é percebida, apesar de haver uma enorme variabilidade em sua realização (HAWKINS, 2004).

As hipóteses sobre a percepção da fala, assumindo que as categorias percebidas são discretas, partem da invariância, ora do sinal acústico, ora da maneira como a audição processa os sons ou como a cognição categoriza a informação.

O trabalho de Cherry (1953) teve a atenção voltada para o que seria relevante ou irrelevante para a percepção da fala ou, em outros termos, o que seria sinal ou ruído. Os humanos teriam a capacidade de juntar um grupo de sons semelhantes como sendo de um único trato vocal em meio a um ambiente ruidoso, descartando os demais. Essa tarefa fica mais árdua quando os sons concorrentes são de mesma natureza, ou seja, trechos de fala superpostos. A esse tipo de evento, Cherry (1953) deu o nome de *cocktail party effect*.

Da mesma forma que os humanos possuem a capacidade de separar o que é importante de um conjunto de estímulos, possuem também a capacidade de somar outras fontes de estímulo para perceber um sinal único. Os experimentos de Sumbly e Pollack (1954) demonstraram que a acuidade por palavras em ambientes ruidosos pode ser aumentada se ao sinal acústico se somar o estímulo visual.

Outras questões foram abordadas além do contexto de produção, tais como a influência das propriedades fonéticas do sinal para a percepção da fala. Em experimentos com fala conectada, Pickett e Pollack (1963) demonstraram que os trechos de fala conectada devem ter pelo menos 800ms para serem inteligíveis, dependendo da taxa de elocução.

Outros trabalhos demonstraram que o ambiente fonético que precede o estímulo também desempenha um papel importante na percepção, como o estudo empreendido por Ladefoged e Broadbent (1957), no qual a diferença de F1 da sentença veículo interfere na percepção das vogais da palavra subsequente. O estudo foi feito com um par mínimo que possivelmente pudesse se tornar ambíguo por sua proximidade vocálica. O par *bit* e *bet* foi veiculado na sentença "Please says what this word is...". Quando o F1 da sentença veículo é alto (380-660 Hz), 97% dos participantes identificaram *bit*; quando o F1 da sentença veículo é baixo (200-380Hz), os participantes identificaram *bet*. Esse efeito auditivo foi chamado de "Information conveyed by vowels".

Num segundo momento das pesquisas de percepção da fala, o interesse dos estudos voltou-se para a não linearidade entre as variações do sinal acústico e a resposta perceptual (HAWKINS, 2004). A MT – "*motor theory of speech perception*" (LIBERMAN; MATTINGLY, 1985) – afirma que o processo

de decodificação da fala está no acesso às articulações equivalentes que produziram o sinal, em outras palavras, o discurso é reconhecido através da comparação entre a articulação inferida e um formato articulatorio armazenado. A teoria quântica (STEVENS, 1972) – *quantal theory* – afirma que a associação entre a articulação e o sinal acústico é não linear, ou seja, pequenas mudanças de parâmetros articulatorios podem resultar em grandes mudanças e vice-versa, o mesmo válido para as mudanças do sinal acústico e a percepção.

Esse segundo período dos estudos de percepção especificou as unidades da fala como dinâmicas, as quais funcionam em conjunto com o seu contexto. Hawkins (2004, p. 18) afirma que *“the context and the current signal together determine whether the speech sounds coherent, and hence what each unit is”*.

Os estudos desenvolvidos desde a década de 1990 abordaram memória, aprendizado de segunda língua e funções cerebrais. Estudos que trabalharam com detalhes fonéticos finos deram evidências de que o detalhe fonético não só reflete alguns atributos da língua, mas também pode ser usado pelos ouvintes para reconhecer a fala (HAWKINS, 2003).

Ao tratar exclusivamente da percepção de prosódia, nota-se que há uma dificuldade em aplicar os métodos dos estudos de percepção de fala de maneira análoga aos estudos de percepção de prosódia. Vaissière (2004, p. 240) afirma que *“the well-established experimental methods developed in psychoacoustics do not actually apply in the field of intonation”*. A distinção fonológica utilizada nos estudos segmentais e a diferenciação semântica não se aplicam aos estudos de entoação. As variações de F0 que carregam significação são gradientes, envolvendo elementos discretos e não discretos

que tornam os estudos de percepção da entoação difíceis de serem realizados (VAISSIÈRE, 2004, p. 241).

Aqui cabe uma breve elucidação sobre a coexistência dos termos prosódia e entoação. Vaissière e tantos autores referem-se à visão estrita da entoação, definindo-a como o uso das variações de F0 para transmitir informação. Nos estudos de prosódia, pode-se encontrar o termo “entoação” para designar os fenômenos suprasegmentais com valores estritamente prosódicos. Nesse viés, toda e qualquer manifestação prosódica como variações de F0, duração e amplitude estão arrolados no termo entoação. Numa abordagem mais estrita, a entoação seria a variação melódica, isto é, modulações do parâmetro F0 que ocorrem durante a fala (HIRST; Di CRISTO, 1998).

Hart, Collier e Cohen (1990) entendem a entoação como basicamente variações de F0, trabalhando com uma abordagem perceptual que visa a dar conta da percepção de *pitch* na fala. Para tal, tentam estabelecer a relevância do papel psicofonético dessas variações, partindo de experimentos psicoacústicos baseados nas *JNDs*¹⁷ (*just noticeable difference*) para que se possa chegar no que há de relevante linguisticamente.

A variedade de definições e parâmetros que podem ser incluídos ou excluídos da análise prosódica, mesmo tratando de produção, está refletida também na definição dos objetos e objetivos desse tipo de estudo. Os interesses de pesquisa variam entre os que estudam as funções gramaticais

¹⁷ Quando se tenta aferir se dois objetos são iguais ou diferentes, referente ao peso, às tonalidades de cor, ou, no âmbito acústico, sensação de altura ou intensidade, chama-se de *JNDs* a mínima diferença perceptível entre os objetos. A percepção das diferenças também é mutável de acordo com a medida dos objetos. Por exemplo, uma diferença de 5 gramas entre dois objetos de 10 e 15 gramas é muito mais perceptível do que em dois objetos de 100 e 105 gramas.

em primeiro plano e os que colocam as funções gramaticais da entoação em segundo plano, dando maior importância às funções emotivas.

Para o estudo experimental da prosódia, há alguns métodos que são usados para isolar os parâmetros envolvidos em sua produção, de maneira que cada um deles possa ser manipulado separadamente ou em conjunto. Essa é uma característica positiva e necessária, pois possibilita ao pesquisador um maior controle dos parâmetros envolvidos na produção. Em muitos desses trabalhos, utilizam-se trechos de síntese de fala em que o parâmetro F0 foi isolado. Mas Vaissière (2004, p. 242) argumenta que esse tipo de metodologia talvez não seja a melhor maneira de se trabalhar percepção quando o interesse da pesquisa for a fala do dia a dia.

Para este trabalho, não há um modelo de análise que possa ser utilizado de maneira direta. Nota-se que as funções prosódicas propostas por Fónagy (2003) não abordam as características regionais passíveis de serem marcadas pela prosódia. A função identificadora, homônima a uma das funções propostas por Vaissière, trata do reconhecimento da prosódia de textos lidos e espontâneos, e do nível tonal que os interlocutores de uma conversa podem adotar a partir do gênero de discurso. Para Vaissière, a função identificadora informa o sexo, idade, personalidade, local de origem e estado mental patológico. Em texto publicado pela autora (VAISSIÈRE, 2004, p. 237), essa função diluiu-se em uma categoria geral chamada *others*.

Uma explicação possível para essa ausência pode estar no caráter multifacetado que a variação dialetal pode assumir em termos prosódicos. Não está em jogo a substituição de um fonema ou traço fonológico por outro, mas a características que cada uma das variedades possui, para as quais não há,

pelo menos até os estudos atuais, características relevantes do ponto de vista perceptual.

3.3 Abordagens sociofonéticas e categorização dialetal

O termo sociofonética refere-se a um tratamento descritivo das variações fonéticas e fonológicas presentes nos dialetos, estabelecendo seus estudos nos limites dos estudos sociolinguísticos e fonéticos. Os estudos sociofonéticos levantaram hipóteses e questões sobre as teorias fonéticas e fonológicas, deixando de distinguir, muitas vezes, *fonética* de *fonologia*. Ohala (1990) contesta a expressão “interface entre fonética e fonologia” argumentando que fonética e fonologia são dependentes entre si e não precisam constituir duas disciplinas autônomas.¹⁸

Os trabalhos na área têm interesse sobre a relação entre as formas fonético-fonológicas e os fatores sociais, em busca das origens e da transmissão de mudanças linguísticas (FOULKES, 2005).

Esta subárea da fonética e da sociolinguística tem dado contribuições para além da teoria linguística, levando em consideração as características das variações regionais, sociais e contextuais para a síntese de fala. Além disso, tem auxiliado na descrição de como as variações acontecem e quão sistemáticas elas podem ser, tornando possível gerar sistemas de reconhecimento de fala muito mais robustos (FOULKES, 2005).

O primeiro estudo autointitulado como sociofonético foi o trabalho de Deshaies-Lafontaine (1974), o qual abordou o francês falado no Canadá. Os trabalhos nessa área abarcam vários interesses relacionados às línguas e suas

¹⁸ Jakobson (1972, p. 17), apesar de autonomizar a fonologia, deixa claro que “quanto mais o fonólogo se voltar para a abundância dos dados da fonética, tanto mais frutuoso será o seu trabalho; quanto mais substância fonética a fonologia experimentar, tanto melhor”.

variações, mas têm suas publicações, em sua maioria, direcionadas ao estudo de produção de fala. Embora os estudos de percepção sejam largamente desenvolvidos em estudos nas áreas da fonética e da psicologia, ainda não é tão profícuo na área da sociolinguística.

Segundo Thomas (2002), uma das razões pela qual isso ocorre é a tradição impressionística que subjaz os estudos de variação, baseada no desenvolvimento do Alfabeto Fonético Internacional e do sistema de Vogais Cardinais por Daniel Jones. O problema desse tipo de abordagem é que, mesmo que se acredite na fidelidade dos dados transcritos, estes não escapam ao crivo perceptual do pesquisador. Um tratamento mais profundo sobre a validade da transcrição fonética pode ser encontrado no trabalho de Kerswill e Wright (1990).

Os trabalhos sociofonéticos que tratam de percepção de fala utilizaram, de maneira geral, palavras ou sentenças as quais foram apresentadas a ouvintes leigos para que pudessem julgar de acordo com o questionamento feito pelo pesquisador, tanto em experimentos de discriminação quanto de identificação. Os experimentos desenvolvidos nesses estudos tinham como objetivo testar um dos tópicos expostos abaixo (THOMAS, 2002):

- i) A habilidade dos ouvintes em identificar uma variedade regional, etnia ou nível socioeconômico dos falantes;
- ii) Como os estereótipos podem influenciar na percepção de sons;
- iii) Como diferenças dialetais afetam a categorização dos fones;

iv) Avaliação de traços pessoais baseada na voz.

Tendo como objetivo a questão *i*, foram feitos estudos em que os falantes de diferentes variedades foram gravados e apresentados a ouvintes para que pudessem identificar qual variedade estava em questão. Os estímulos apresentados incluíram logotomas, léxico real, sentenças manipuladas por filtros passa-baixa e passa-alta ou com seu sinal original preservado. Os experimentos feitos com áudios filtrados focalizaram seu interesse nos fatores prosódicos como pista na discriminação de dialetos (BUSH, 1967; RICHARDSON, 1973).

Bezooijen e Gooskens (1999) investigaram a importância da cadeia segmental e da prosódia em dialetos do holandês e do inglês britânico. Eles trataram os estímulos de três formas: a) eliminaram a maior parte da informação segmental usando filtro passa-baixa até 400Hz; b) eliminaram as variações de F0 por meio de monotonização da curva melódica; e c) forneceram os sinais originais. A partir dos estímulos, foram propostas duas tarefas: identificar de qual localidade os falantes eram e avaliar o grau de divergência dialetal dos estímulos com relação ao holandês padrão e o inglês britânico padrão.

Os resultados demonstraram que os ouvintes ingleses foram melhores em reconhecer sua prosódia do que os ouvintes holandeses. A questão não respondida nesse trabalho é o porquê desse resultado. Qual informação contida no sinal acústico foi importante para o resultado obtido?

Outra finalidade encontrada nos estudos sociofonéticos contida na questão *i* é a identificação da etnia dos falantes, avaliada por meio de experimentos de percepção desenvolvidos para testar essa habilidade nos

ouvintes. O estudo de Baugh¹⁹ (1996 apud THOMAS, 2002) demonstrou que os ouvintes distinguiram falantes afro-americanos de falantes euro-americanos e mexicanos por meio de gravações de textos lidos e espontâneos.

A questão *i* também trata da identificação de estratos sociais por meio de experimentos perceptuais. Muitos trabalhos que abordam a classe econômica dos falantes são conduzidos, muitas vezes, paralelamente aos estudos que tratam de identificação de etnias.

Os estudos referentes à questão *ii* verificam se os estereótipos podem alterar a percepção dos ouvintes. Strand (1999) trabalhou com a influência do sexo do falante na percepção das sibilantes /s/ e /ʃ/ – em sílabas como [sq] e [[a] – com experimento multimodal baseado no efeito *McGurk*²⁰ (MCGURK; MACDONALD, 1976). As fricativas /s/ e /ʃ/ foram sintetizadas e seguidas pela vogal /a/. Os resultados demonstraram que a percepção não se dá somente pelos atributos físicos da fala, mas também pelos estereótipos de gênero. Os participantes perceberam diferenças na produção induzidos pela visão dos rostos femininos e masculinos, mesmo quando a contraparte acústica dos estímulos permanecia idêntica.

Tratando da questão *iii*, alguns experimentos envolvendo falantes de diferentes dialetos com relação à percepção de fonemas demonstraram que as escolhas correspondem diretamente às diferenças encontradas entre os dialetos (WILLIS, 1972).

¹⁹ Baugh, J. *Perceptions within a Variable Paradigm: Black and White Racial Detection and Identification Based on Speech*. In: Focus on the USA, Ed. Edgar W. Schneider, 169–82. Amsterdam: Benjamins, 1996.

²⁰ Nos testes tradicionais que envolvem esse tipo de técnica, há a permutação áudio-visual dos segmentos linguísticos. Como exemplo, os informantes desse tipo de experimento afirmam ter ouvido a sílaba /na/ como resultado da sobreposição das sílabas /da/ - *estímulo visual* e /ma/ - *estímulo acústico*. O resultado perceptual é fruto da interpretação de traços visuais e acústicos de cada sílaba. Assim, os informantes uniram perceptualmente o modo de articulação do estímulo acústico (nasalidade) e o ponto de articulação do estímulo visual (alveolar).

E, finalmente, na questão *iv*, encontram-se trabalhos que tratam das características pessoais baseadas na voz. O estudo de Lambert et al. (1960) entre bilíngues demonstrou que os ouvintes avaliaram diferentemente os falantes dependendo da língua falada, no caso inglês e francês, mesmo quando se tratava do mesmo falante. É importante salientar que os ouvintes não sabiam que se tratava do mesmo falante ora falando inglês, ora falando francês.

Após essa breve exposição sobre a abrangência dos estudos sociofonéticos, é possível atestar que o estudo aqui pretendido está baseado na questão *i*, especificamente, no reconhecimento de variedades regionais. A questão *i* exposta por Thomas (2002) tem como referência um trabalho já citado, o de Bezooijen e Gooskens (1999), o qual será mais detalhadamente descrito e comparado aos intentos deste estudo. No artigo intitulado "*Identification of language varieties*", os autores realizaram três experimentos (como descritos anteriormente) com o intuito de delimitar qual o papel dos diferentes estímulos na identificação das variedades linguísticas.

Para o presente estudo, somente os dois primeiros experimentos feitos pelos autores têm relação estreita com os experimentos que serão descritos na metodologia deste trabalho. O terceiro experimento não será abordado aqui, pois trata-se de um teste feito com leitura. Mesmo não tendo relação direta, avaliar variedades regionais de forma perceptual por meio de textos lidos pode ser uma boa maneira de trabalhar somente com informações segmentais e prosódicas, excluindo do experimento pistas marcadas lexical, morfológica e sintaticamente.

Entretanto, é importante também notar que a prosódia de textos lidos pode ser mais padronizada, prejudicando posteriores avaliações sobre a

entoação. Fónagy (2003) afirma que os movimentos tonais são mais expressivos e menos previsíveis na conversação do que em leituras de texto, demonstrando que o desvio padrão das alturas intersilábicas em fala espontânea é maior do que em textos lidos.

Para uma abordagem sobre o tema, os trabalhos de Romano e Moutinho (2004), Barbosa (2008) e Ferreira Netto e Peres (2009) analisaram algumas características presentes na prosódia de textos lidos e as suas diferenças com relação à fala espontânea.

O primeiro experimento aplicado por Bezooijen e Gooskens (1999) para o holandês tentou buscar qual nível seria mais importante para o reconhecimento dialetal. Os estímulos foram divididos em *prosodic*, *verbal* e *integral*, os quais podem ser definidos como fala com filtro passa-baixa, fala monotônica e fala sem manipulação, respectivamente. Os informantes tiveram de discriminar de onde os falantes eram, preenchendo um formulário no qual deveriam colocar o país, região e província de onde provinham as gravações.

Os resultados apontaram para um baixo papel da prosódia, aumentando o nível de acertos em direção à fala sem manipulação. É esperado que, na medida em que se retira informação da fala, a tarefa de reconhecimento se torne mais árdua.

Outro ponto analisado no primeiro experimento de Bezooijen e Gooskens (1999) foi a precisão com a qual os ouvintes indicavam a proveniência dos falantes. A porcentagem de acertos diminuiu quando a discriminação era mais específica como, por exemplo, a província de onde os falantes eram oriundos. O índice de erros em porcentagem está ilustrado abaixo:

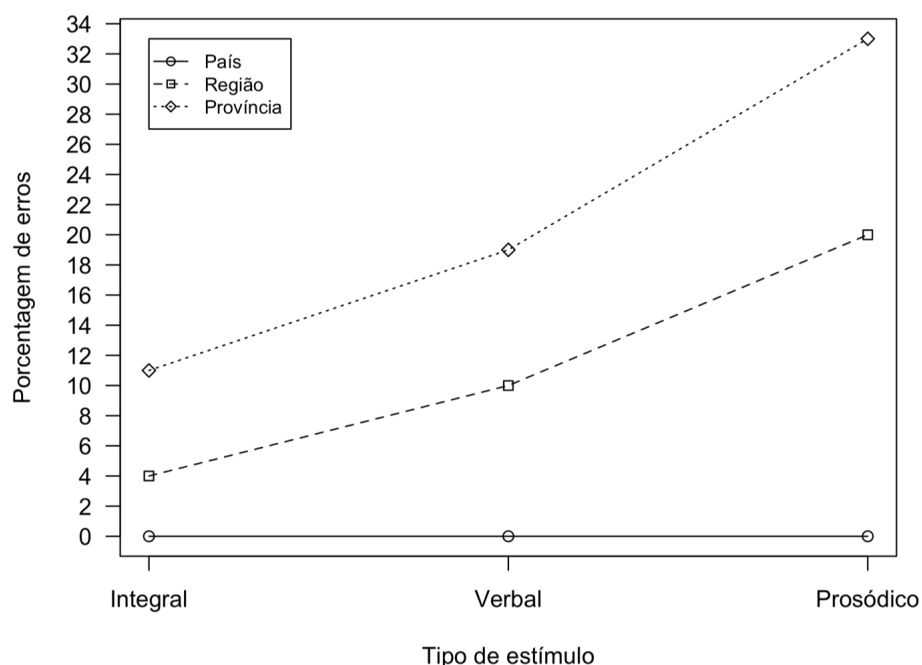


Gráfico 1: Tendência percentual de erros do primeiro experimento de Bezooijen e Gooskens (1999).

O experimento 2 contou com a mesma metodologia, diferindo do primeiro apenas em relação ao lugar da sua aplicação – Reino Unido – e a adição de um estímulo chamado *prosodic special*, definido pelos autores como uma prosódia típica e específica avaliada por linguistas.²¹ O resultado obtido foi semelhante ao encontrado no primeiro experimento. A adição de *prosodic special* não influenciou os resultados, exceto para os obtidos em Belfast nos níveis de país e região.

Uma característica do estudo de Bezooijen e Gooskens (1999) que difere essencialmente dos intuítos aqui pretendidos é que os informantes dos experimentos perceptuais não eram provenientes das regiões dos falantes, ou seja, não foi pedido para os informantes reconhecerem sua própria variedade, mas uma outra. Como conclusão, os autores atestaram que a prosódia desempenhou um papel menor no reconhecimento dialetal dos holandeses e

²¹ Os autores não deram detalhes sobre os critérios adotados pelos linguistas para selecionar o estímulo *prosodic special*, somente deixam claro que se trata de “*dialect-specific prosody [...] speech utterances with marked regional prosodic characteristics*”.

dos ingleses. Não há como medir, até então, a influência dessa decisão nos resultados com estímulos puramente prosódicos.

Essa decisão metodológica afasta significativamente os objetivos de Bezooijen e Gooskens (1999) dos pretendidos neste trabalho, mas, por outro lado, abre uma possibilidade de diálogo entre os resultados, possibilitando medir qual o peso da decisão de submeter aos testes perceptuais ouvintes provenientes ou não das regiões de onde os falantes são nativos.

Tratando especificamente do experimento de fala monotônica feito pelos autores, estipulou-se um tom médio padrão de 110Hz para todos os estímulos, não havendo nenhuma diferenciação quanto às regiões dos falantes, tampouco alguma relação com o tom médio dos falantes. Os resultados apresentaram um índice de acerto próximo ao obtido com a fala integral, superando os resultados obtidos somente com os estímulos prosódicos.

Neste trabalho, ao contrário da adoção de um tom médio padrão para os falantes, foi possível extrair automaticamente os valores de F0 médio de cada falante para cada trecho de fala selecionado, utilizando para isso o aplicativo *ExProsodia*. Adotou-se essa medida porque não está clara qual seria a relação entre a média de F0 dos falantes e a percepção das variedades regionais. Ao estabelecer um único valor para o F0 médio, como no estudo envolvendo o holandês e o inglês, elimina-se por completo a possibilidade de se avaliar qual o seu papel. A importância desse parâmetro vai ser tratada na discussão dos resultados deste estudo.

Outra diferença entre os dois estudos é a metodologia adotada para a eliminação das informações lexicais, pois Bezooijen e Gooskens (1999) utilizaram um filtro passa-baixa de 400Hz. Neste trabalho, foi usado um método de delexicalização (PURR) que será descrito no próximo capítulo.

4. METODOLOGIA

4.1 Sobre a separação segmental-suprasegmental

Com o intuito de avaliar qual informação os ouvintes podem perceber sem que estejam presentes os segmentos originais de fala, vários métodos foram criados para que pudesse haver uma separação entre segmentos e prosódia. É, porém, importante deter-se ao fato de que esses métodos promovem uma “distorção” em um dos níveis que se pretende reconhecer. Os trabalhos que lidam com esses tipos de técnica experimental têm como objetivo, em sua maioria, avaliar a identidade do falante e o reconhecimento de línguas (MOFTAH; ROACH, 1988).

Nos estudos de Bush (1967) e de Ohala e Gilbert (1978), as pistas segmentais foram modificadas para que os ouvintes não conseguissem ter acesso aos segmentos que foram produzidos pelos falantes. Da mesma forma, Bezooijen e Gooskens (1999) também modificaram o estímulo que foi dado aos informantes, mas adicionando ao seu trabalho a sua contraparte, ou seja, estímulos que tiveram a entoação monotônica. As técnicas utilizadas para esse tipo de experimento são chamadas, em termos mais amplos, de fala distorcida.

De acordo com os estudos que trabalham com experimentos perceptuais com fala “sem palavras”, há alguns métodos distintos que têm como objetivo retirar a informação segmental da cadeia sonora da fala. Um deles é o sinal laringográfico,²² que capta a onda diretamente da fonte, isto é, produzida pela laringe. A vantagem desse tipo de método é que o som captado não passou pelas ressonâncias do trato vocal e que provavelmente não sofreu

²² *Laryngograph signal* (abreviado como Lx).

nenhuma interferência do filtro. A desvantagem do sinal laringográfico é o seu caráter invasivo, pois é coletado por um conjunto de eletrodos colocados na garganta do informante. Estudos que lidam com fala espontânea podem sofrer algum tipo de interferência por conta desse método.

Outro método utilizado é a filtragem passa-baixa na qual é utilizado um filtro que permite passar somente as frequências pré-determinadas pelo pesquisador, abaixo de um limiar previamente estabelecido. O método de filtragem muitas vezes não consegue eliminar por completo os resquícios que os segmentos possam deixar no sinal acústico. Por essa razão, optou-se, neste trabalho, pela utilização de um outro método de delexicalização.

A delexicalização consiste em retirar do sinal acústico da fala toda a informação segmental relevante, tornando-o ininteligível do ponto de vista segmental. Ao mesmo tempo, o resultado desse processo preserva as características prosódicas como variações de F0, amplitude e duração. Vainio et al. (2009) propuseram um método de delexicalização que possibilita a manipulação de todos os parâmetros presentes na prosódia de maneira autônoma, adicionando a esses a qualidade vocal. O princípio desse método é separar a fonte do filtro, interpretando a fala como fluxo linear constituído de três partes: a fonte, o trato vocal e os lábios.

Vainio e seus colegas (2009) propuseram uma filtragem glotal inversa²³ em que a fala passa por um procedimento repetitivo de cancelamento dos efeitos causados pelo trato vocal e lábios ao sinal produzido pela glote. Assim, é possível estimar a velocidade da onda gerada pela glote. A dificuldade de uso desse método é o difícil acesso e aplicação, pois os arquivos têm de ser enviados para processamento pelos próprios autores.

²³ *Glottal inverse filtering.*

Outro método de delexicalização é o de Sonntag e Portele (1998a; 1998b). O método chama-se PURR (*Prosody Unveiling Restricted Representation*), o qual deixa como informação para os ouvintes apenas a intensidade, variações de F0 e a estrutura temporal. Em estudo comparativo, os autores compararam seis métodos de delexicalização antes de avaliar os resultados com PURR. Os métodos avaliados foram descritos da seguinte forma:

- **CCITT** – Marcas de *pitch* extraídas do sinal original e preenchidas com um sinal proposto por CCITT (1989), e depois filtrados por passa-baixa.
- **F0 filt** – o som original é filtrado com passa-baixa utilizando-se um filtro com corte de frequência logo acima de F0. Os segmentos desvozeados foram automaticamente modificados para frequência nula.
- **inv** – Combinação entre inversão espectral e filtragem proposta por Kreimann (1982). Depois de aplicar um filtro passa-alta de 600Hz, o sinal resultante passou por inversão espectral. Após isso, o som foi filtrado com passa-baixa de 4000Hz e somado ao som original filtrado por passa-baixa de 200Hz. Este método mantém as diferenças entre vozeado e desvozeado.
- **lfm** – Marcas de *pitch* extraídas do sinal original e preenchidas como proposto pelo modelo do fluxo glotal de Fant, Liljencrants e Lin (1985).
- **saw** – Marcas de *pitch* extraídas do sinal original e preenchidas com uma onda dente de serra.
- **sin** – Marcas de *pitch* extraídas do sinal original e preenchidas com uma senoide com o primeiro harmônico com 1/4 da amplitude e o segundo harmônico com 1/16 da amplitude (KLASMEYER, 1997).

Os métodos foram avaliados segundo os critérios de funcionalidade (detecção de sílabas, acento frasal e fronteiras) e naturalidade. Os métodos julgados como menos naturais foram o **inv** e o **saw**. O **sin** obteve os melhores resultados com relação a funcionalidade e naturalidade.

Com relação ao PURR, os autores demonstraram em experimento de reconhecimento rítmico que esse método foi muito bem avaliado. Os informantes conseguiram relacionar frases escritas de mesma estrutura sintática com o estímulo gerado por PURR, mesmo que os segmentos fossem diferentes dos representados na escrita. Os resultados validam a hipótese de que a percepção do estímulo PURR é totalmente separada das qualidades segmentais e que sentenças de estruturas sintáticas semelhantes podem ser comparadas apesar das diferenças lexicais (SONNTAG; PORTELE, 1998b).

Para a delexicalização das frases, o método PURR determina marcas de *pitch* ao longo da sentença e três ondas senoidais são sobrepostas: a primeira refletindo a amplitude e frequência originais do som captado; a segunda com o dobro da frequência e 1/4 da amplitude; e a terceira com o triplo da frequência e 1/16 da amplitude. Na figura 10, pode-se visualizar a sobreposição das senoides com as diferentes frequências e amplitudes.

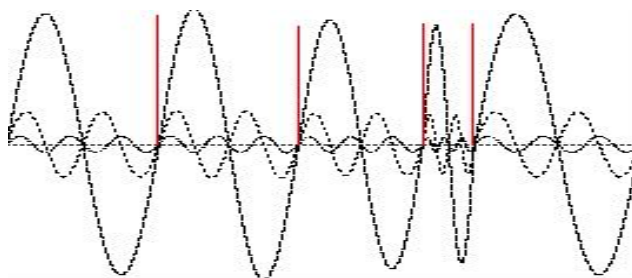


Figura 10: Sobreposição de senoides para obter o sinal PURR.²⁴

²⁴ Figura obtida em: <http://www.ikp.uni-bonn.de/forschung/phonetik/prosodiegenerierung-fur-die-sprachsynthese/purr/?searchterm=PURR>. Acesso em: 10 jan. 2011.

Os resultados obtidos por Sonntag e Portele (1998b) atestam que o método pode ser usado para avaliações de diferentes realizações prosódicas, não sendo necessário o controle do conteúdo pragmático e lexical das sentenças. A escolha desse método de delexicalização para este trabalho está baseada na avaliação positiva de sua funcionalidade e naturalidade, além de poder ser aplicado em trechos de fala que muitas vezes não são captados em ambientes ideais de gravação, tal a sua natureza espontânea.

Outro argumento a favor é sua fácil aplicabilidade, pois, para a sua implementação, há um *script* para Praat, criado por Petra Wagner (2004). Assim, todos os trechos de fala utilizados neste estudo passaram pelos mesmos critérios de manipulação.

4.2 Descrição dos experimentos

4.2.1 Experimento 1 com técnica PURR

No experimento 1 — enunciados delexicalizados com a técnica PURR — foram utilizadas amostras de fala adulta do sexo masculino divididas entre nove informantes (distribuídas igualmente entre as três variedades escolhidas [SP, RS e CE]). Os informantes gravados nasceram e residem em suas respectivas cidades, sendo muito breves e raras as suas ausências dos locais de nascimento. Os detalhes de cada sujeito estão expostos no quadro abaixo:

Informantes	Idade	Estado	Escolaridade
1	32	SP	2º Grau completo
2	34	SP	Superior completo
3	62	SP	Superior completo
4	31	RS	2º Grau completo
5	34	RS	Superior completo
6	43	RS	2º Grau completo
7	30	CE	2º Grau completo
8	58	CE	2º Grau incompleto
9	59	CE	2º Grau incompleto

Quadro 1: Relação de características por informante.

Ao todo, foram selecionados seis trechos fluentes de cada sujeito (três trechos curtos [TC] e três trechos longos [TL]). As frases foram obtidas por meio de gravações de fala espontânea. A duração dos trechos curtos escolhidos para o teste foi de até 10 segundos, e a dos trechos longos entre 15 e 25 segundos. As gravações de São Paulo (Mooca) e do Ceará (Senador Pompeu) foram feitas com gravador digital ZOOM H2, com taxa de 44.1 kHz. As gravações do Rio Grande do Sul (Pelotas) fazem parte do banco de dados VARX (AMARAL; BORGES, 2001), também com taxa de 44.1 kHz.

No experimento 1 (PURR), houve um breve treino em que os ouvintes familiarizaram-se com o que é entoação. Todas as instruções para a feitura do experimento foram gravadas para que os informantes ouvissem sempre a mesma explicação, evitando, dessa maneira, algum tipo de enviesamento na aplicação dos experimentos. Os informantes também puderam acompanhar as explicações por meio de uma versão escrita.²⁵

Para a fase de treinamento, uma frase de cada variedade foi apresentada em sua versão não manipulada e manipulada com PURR, com o

²⁵ A versão escrita dos testes está nos anexos.

intuito de fazer o ouvinte entender o procedimento. Após o treinamento, 13 ouvintes de cada uma das localidades ouviram duas sequências de estímulos, a primeira constituída de 27 trechos curtos (TC) das 3 variedades em questão, a segunda constituída de 27 trechos longos (TL) das 3 variedades, ambos dispostos aleatoriamente.

As duas sequências somaram 30 minutos de experimento. O esquema abaixo resume a totalidade de estímulos e o número de respostas dadas pelos ouvintes no experimento 1 (PURR):

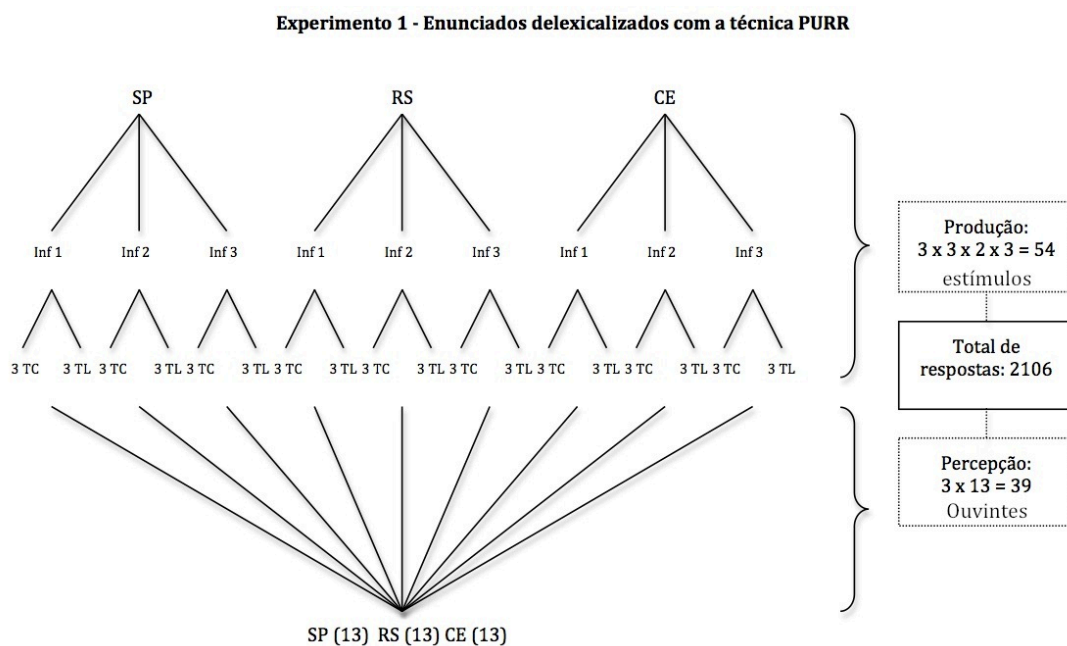


Figura 11: Total de estímulos e respostas do experimento 1 (PURR).

Os ouvintes que participaram do experimento possuem idade entre 20 e 55 anos, de ambos os sexos, sem problemas auditivos e alfabetizados. O problema de se controlar uma variante como a escolaridade está na natureza heterogênea presente em estudos que lidam com variedades tão distantes do ponto de vista territorial. A seleção de ouvintes leigos que possa suprir uma categoria estrita de nível escolar, como, por exemplo, nível superior completo,

torna-se uma tarefa árdua, levando-se em conta as diferenças entre campo, cidade e regiões brasileiras.

Dessa maneira, optou-se por adotar uma categoria mais ampla como *alfabetizados*. Com essa delimitação, tornou-se possível o controle dos níveis de escolaridade, os quais variaram entre o ensino médio incompleto ou completo.

4.2.2 Experimento 2 com curva melódica monotônica (CMM)

No experimento 2 (CMM) — enunciados com curva melódica monotônica — foram utilizadas amostras de fala adulta do sexo masculino divididas entre 9 informantes (distribuídas entre as 3 variedades escolhidas [SP, RS e CE]). Os informantes foram os mesmos que participaram do experimento 1.

Para eliminar as variações de F0 presentes nos estímulos originais, foi necessário pensar uma maneira que automatizasse algumas partes desse processo.

Como referido na seção 3.3, Bezooijen e Gooskens (1999) haviam monotonizado sentenças utilizando um tom médio geral de 110 Hz para os falantes, ignorando uma provável diferenciação no tom médio dos falantes e de suas regiões. Por isso, optou-se por obter o tom médio de cada falante via *ExProsodia*. O aplicativo é capaz de gerar automaticamente os valores de tom médio para cada trecho de fala. As demais funções do *ExProsodia* serão exploradas no capítulo 5.

Após gerar os valores individuais do tom médio, as camadas de linha melódica dos trechos originais de fala foram substituídas pelo tom médio obtido

pelo *ExProsodia*. A substituição das camadas tonais foi feita com a utilização do *software* Praat. O primeiro passo foi gerar uma senoide no valor do tom médio e na mesma duração da frase original. O segundo passo foi gerar um arquivo no comando *To manipulation*, (com tempo de 0.01, mínimo e máximo de 75Hz e 600Hz) para a senoide e para o som original.

O terceiro passo foi extrair as camadas tonais do som original e da senoide utilizando o comando *extract pitch tier*. O quarto e último passo foi sobrepor a camada tonal da senoide ao arquivo original, utilizando o comando *replace pitch tier*. Na figura abaixo, tem-se o arquivo já devidamente monotonizado:

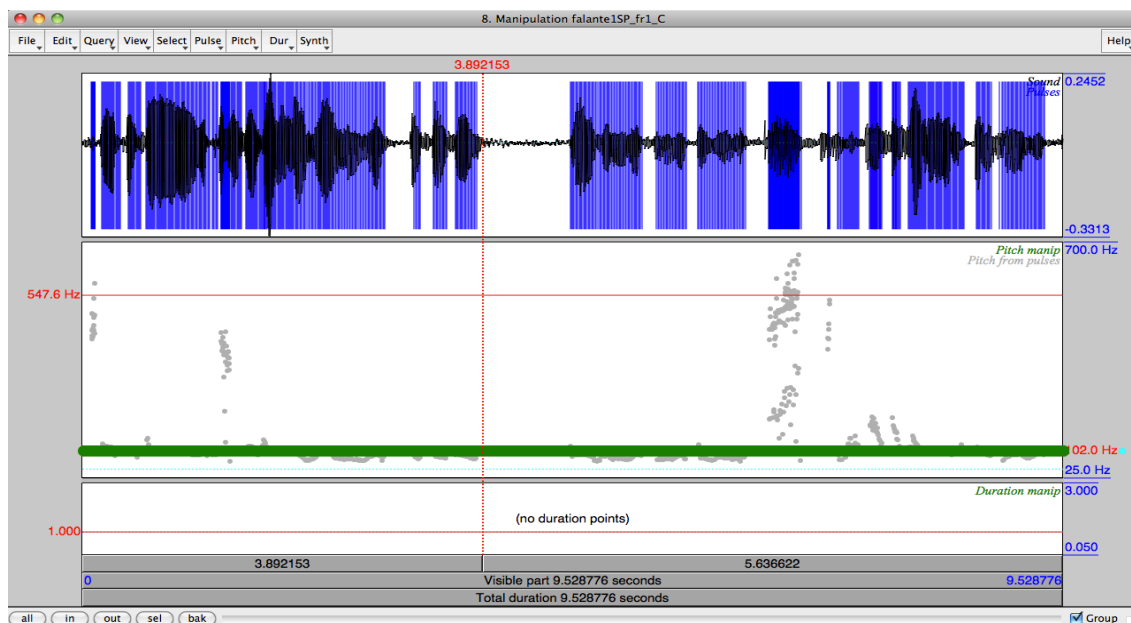


Figura 12: Resultado do processo de monotonização dos estímulos.

Para a fase de treinamento, apresentou-se uma frase de cada variedade, apresentada em sua versão não manipulada e monotonizada. Após o treinamento, 13 ouvintes de cada uma das localidades ouviram uma sequência de estímulos constituída de 18 trechos das 3 variedades em questão dispostos aleatoriamente. O experimento durou 5 minutos. O esquema abaixo

resume a totalidade de estímulos e respostas dadas pelos ouvintes no experimento 2 (CMM):

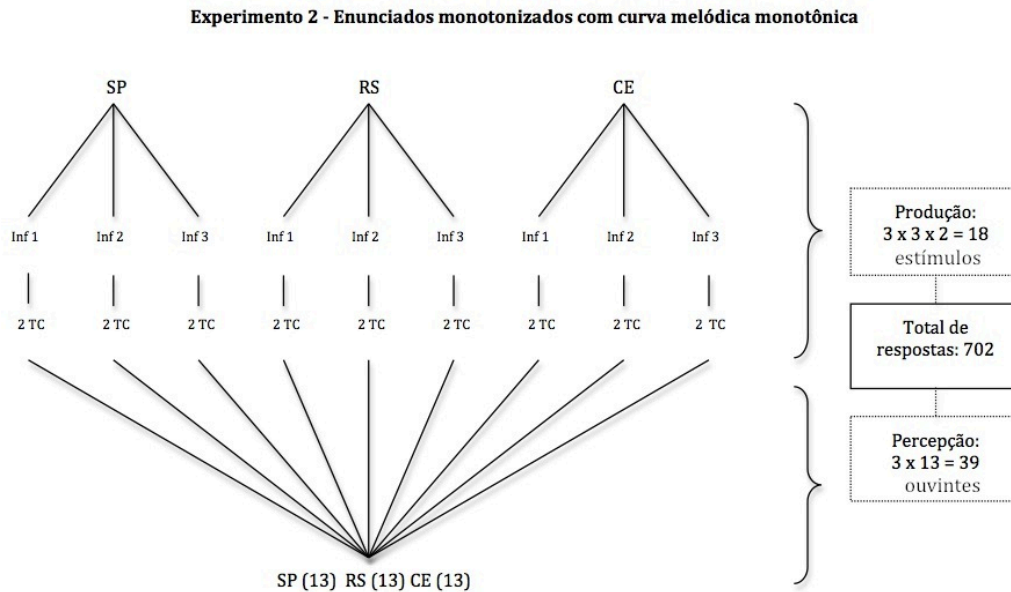


Figura 13: Total de estímulos e respostas do experimento 2 (CMM).

Os ouvintes que participaram do experimento 2 são os mesmos que participaram do experimento 1, obedecendo ao mesmo critério descrito no experimento com PURR.

4.2.3 A Teoria da Detecção

A Teoria da Detecção – TD – (MACMILLAN; CREELMAN, 2005) é baseada na Teoria da Detecção do Sinal, estudo pioneiro feito por Green e Swets (1966). A TD é uma abordagem psicofísica usada para medir o desempenho dos informantes em tarefas de identificação, nas quais há somente um estímulo, ou em tarefas de discriminação com pares de estímulos.

A TD foi introduzida na psicologia como um método para explicar a *performance* de informantes em experimentos nos quais os sinais fracos (auditivos ou visuais) fossem reconhecidos em meio a um ambiente ruidoso. Dentro da terminologia da TD, os informantes são avaliados quanto a sua habilidade em extrair um determinado sinal do ruído.

É importante esclarecer que os conceitos de sinal e ruído são metafóricos para o tipo de experimento desenvolvido neste trabalho (ABDI, 2007), no qual figuram como sinal as variedades pertencentes aos ouvintes do experimento e como ruído os estímulos que não fazem parte da variedade dos informantes.

Para os fins deste trabalho, a teoria em questão será utilizada para mensurar o quanto os informantes estão aptos a reconhecer a sua variedade, mediante estímulos manipulados com a técnica PURR e com a curva melódica monotônica (CMM).

A Teoria da Detecção trabalha com quatro tipos de respostas dadas pelos informantes: *hit*, *miss*, *false alarm* e *correct rejection*. Neste estudo, *hit* (acerto) é o reconhecimento de uma variedade pelo participante (cearense reconhecendo a variedade cearense etc.); *miss* (erro) é o não reconhecimento da variedade pelo participante (cearense dizendo não para a variedade cearense); *false alarm* (alarme falso) é o reconhecimento indevido de uma variedade (cearense reconhecendo a variedade paulista como sendo a sua); e *correct rejection* (rejeição correta) é o não reconhecimento de uma variedade da qual o participante não faz parte (cearense não reconhecendo a variedade paulista como sendo a sua). A tabela abaixo demonstra como os estímulos são organizados:

	Estímulo	
Resposta	Sim	Não
Sim	<i>Hit</i>	<i>False Alarm</i>
Não	<i>Miss</i>	<i>Correct Rejection</i>

Tabela 2: Exemplo de organização dos dados em TD.

A TD trabalha com a medição da sensibilidade de discriminação do participante, que é dada pela taxa de acertos e alarmes falsos. O resultado da diferença entre as inversas da distribuição gaussiana de acertos e alarmes falsos é chamado de d' (*dee-prime* ou *d linha*):

$$d' = z(H) - z(FA)$$

O valor d' indica o quanto os informantes foram capazes de diferenciar o sinal do ruído. Assim, quando os informantes não diferenciam os estímulos significa que $H = FA$ e $d' = 0$. Por exemplo, uma situação hipotética em que o participante respondeu positivamente a todos os estímulos em um experimento com 20 estímulos e 10 possibilidades de identificação correta. Nesse caso, o número de identificações acertadas ($H = 10$) se iguala ao número de identificações equivocadas ($FA = 10$). Em outra hipótese, quando os informantes apresentam uma alta taxa de acertos e uma baixa taxa de alarmes falsos, então $H > FA$ e $d' > 0$. Nesse caso, tendo como exemplo o mesmo experimento hipotético com 20 estímulos e 10 possibilidades de identificação correta, o participante também identifica corretamente os 10 estímulos ($H = 10$), mas não identifica equivocadamente os demais estímulos. Assim, o participante acaba por rejeitar corretamente aquilo que deveria não ser identificado ($FA = 0$). As duas situações supracitadas são tidas como raras, representando situações ideais de reconhecimento e não reconhecimento completos.

O comportamento mais comum dos informantes em experimentos como esse, tendo êxito ou não na discriminação, é situar seus desempenhos em nenhum dos extremos. Assim, no mesmo experimento hipotético com 20 estímulos e 10 possibilidades de identificação, os informantes obteriam, por exemplo, um bom resultado como $H = 7$ e $FA = 1$. Nesse caso, ter-se-ia o índice $d' = 1,37$. Num exemplo de participação com desempenho não tão satisfatório, com $H = 5$ e $FA = 4$, ter-se-ia $d' = 0,25$.

A vantagem da Teoria da Detecção é a de não só dar conta dos acertos, das respostas positivas, mas de conseguir isolar as respostas dadas a esmo. Essa clareza dos dados é muito útil para assumir posições acerca do desempenho de cada participante.

5. ANÁLISE AUTOMÁTICA DA FALA - *EXPROSODIA*

Em linhas gerais, o projeto *ExProsodia* objetiva estabelecer um modelo para a análise automática da entoação na língua portuguesa. A entoação é entendida, como uma sequência de tons, iguais ou diferentes, produzidos pela voz durante a fala. A análise baseia-se na hipótese inicial de Ferreira Netto (2006) de que a entoação pode ser entendida como uma série temporal, ou seja, um conjunto de observações sequenciadas e interdependentes, sendo o resultado da observação feita no momento $t+1$ condicionada ao momento t . Em outras palavras, o valor observado em cada momento de medição depende do valor do momento imediatamente anterior, não permitindo inversão dos valores observados. Assim, a ordem de ocorrência de cada momento deve ser mantida como uma informação inerente ao valor obtido.

A análise está baseada na proposta de Xu e Wang (1997) de que alguns fatos prosódicos são decorrências de restrições *mecânico-fisiológicas* e outros configuram as *intenções expressivas* dos falantes.

As intenções expressivas seriam as variações intencionais de F0 com fins linguísticos, em outras palavras, alterações na onda sonora que configuram a intenção de imprimir significados diversos à fala. As restrições mecânico-fisiológicas seriam condicionadas pelas propriedades fisiológicas do falante.

Para dar conta dessa proposta, Ferreira Netto (2008) propôs a decomposição automática de F0 por meio do aplicativo *ExProsodia*, dividindo F0 em *componentes estruturadoras* e *semântico-funcionais*. As componentes ainda seriam subdivididas em finalização (F) e sustentação (S) — formando o

ritmo tonal — foco/ênfase (E) e acento lexical (A), formando a face significativa da entoação frasal.

Dessa forma, cada mensuração feita ao longo da série temporal, no caso presente, as variações do parâmetro F0, é resultado da coocorrência das componentes que a formam. Em esquema, tem-se:

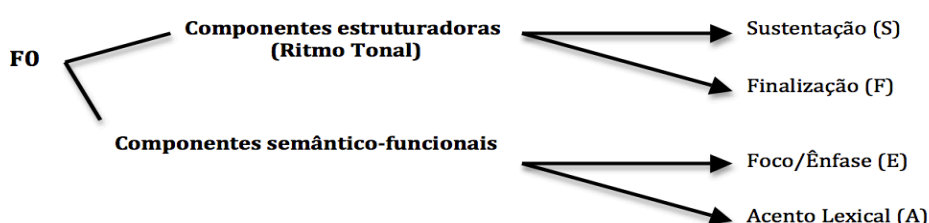


Figura 14: Esquema das componentes de F0.

A composição do valor Z obtido no momento t — notado como Z_t — será a soma das componentes que concorreram para o estabelecimento desse valor. Os núcleos silábicos, por portarem a massa de sonoridade mais significativa da fala, serão os momentos de medição (Z_t) escolhidos pela rotina de análise *ExProsodia* como candidatos a pico silábico (cps), pois as variações de frequência ocorrem especialmente nesses núcleos (FERREIRA NETTO, 2006). Assim, tem-se: $Z_t = S + F + E + (A)$, ou seja, um candidato a pico silábico (cps) é igual à soma das componentes *sustentação*, *finalização*, *foco/ênfase* e *acento lexical*.

A componente acento lexical (A) parece não ser tão significativa para a constituição da série temporal em português, pois a frequência fundamental dificilmente é, em português, correlato de acento lexical.²⁶ Em resumo, as

²⁶ (A) seria estabelecido pela seleção lexical do falante em línguas *pitch accent* ou em línguas tonais (FERREIRA NETTO, 2008).

componentes propostas por Ferreira Netto (2008) para a composição de F0 são:

Componentes estruturadoras (Ritmo Tonal):

- *Sustentação* (S) – Reação à tendência natural do falante em abandonar o esforço de manutenção da energia de um candidato a pico silábico (cps).
- *Finalização* (F) – Componente consequente da atividade de sustentação da fala é o valor mínimo a ser atingido nos candidatos a pico silábico.

Assim, a produção da fala necessita de um esforço para a sua manutenção, deixando a voz com uma frequência estável, definida aqui como tom médio (TM) ideal de F0. A supressão desse esforço desencadeia uma declinação pontual que exige a retomada da tensão inicial.

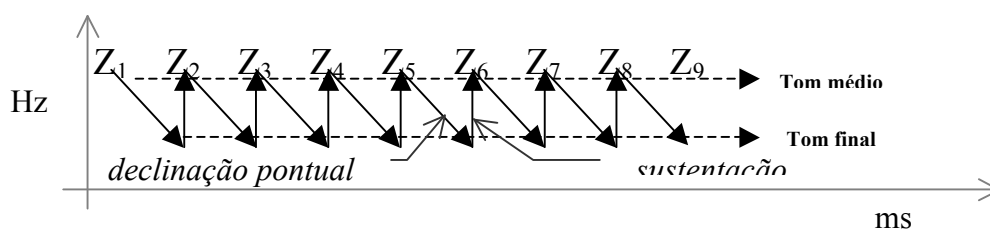


Figura 15: As setas diagonais representam a declinação desencadeada pelo desejo de supressão do esforço e as setas verticais representam a sustentação desencadeada pela retomada da tensão inicial, ambas formando o ritmo tonal.

Componentes semântico-funcionais:

- *Foco/Ênfase* (E) – A sua ocorrência estabelece variações de sentido, decorrendo da necessidade expressiva do falante.

- Acento Lexical (A) – A sua ocorrência depende da seleção lexical feita pelo falante, em línguas *pitch accent* ou tonais.

Por permitir a análise isolada de cada uma das componentes de F0, a rotina *ExProsodia* difere de outras experiências realizadas como as que vão descritas em Hirst e Espesser (1993), relativa ao projeto MOMEL, e em Dutoit et al. (1996), relativa ao projeto MBROLA.

A decomposição de F0 parte da definição de seu tom médio como forma prototípica, para a qual todas as frequências próximas convergem. Essa proposta parte do modelo *perceptual magnet effect* (PME), desenvolvido por Kuhl e seus colegas (IVERSON; KUHL, 1995; KUHL, 2000; KUHL; TSAO; LIU; ZAHNG e DE BOER, 2001)

O modelo PME pressupõe a existência de uma forma prototípica fixa, não abstrata, que sirva de parâmetro de comparação inicial para todas as demais formas posteriormente percebidas. Da mesma forma, os tons prosódicos estão sujeitos à formação de categorias que sirvam de parâmetros para as demais, fundadas em um modelo específico, localizado nos extremos máximos das possibilidades de variação do fenômeno (FERREIRA NETTO, 2008). Assim, pode-se pensar em dois tons específicos, um grave e outro agudo, que serão as variações extremas de um tom médio (TM).

Na proposta do projeto *ExProsodia*, o TM é o ponto de partida da decomposição da série temporal de F0, definido pelo próprio falante e decorrente de suas intenções prosódicas, cujo “poder magnético” parece estar em torno de 3 st ascendentes (CONSONI; FERREIRA NETTO, 2008) — o

mesmo encontrado por t'Hart (1981) no holandês — e 4 st descendentes²⁷ (CONSONI, 2010; FERREIRA NETTO; CONSONI; PERES, 2011).

A rotina *ExProsodia*, elaborada em Visual Basic para Excel, faz de maneira automática a análise de frequência, intensidade e duração de segmentos de frases convertidos na forma de texto e extraídos pelo programa *Speech Filing System v. 4.7* (SFS Win v1.7; HUCKVALE et al., 1987). Os arquivos *.txt*, um com as medidas de frequência tomadas a cada 5 ms pelo SFS, e outro, com medidas de intensidade tomadas no mesmo intervalo de tempo da frequência, são lidos pela rotina.

A partir desses dados, a rotina *ExProsodia* toma alguns parâmetros para a escolha dos candidatos a pico silábico(cps). São eles:

- Limiar inferior de frequência: 50 Hz. Opção do usuário.
- Limiar superior de frequência: 350 Hz. Opção do usuário.
- Limiar inferior de duração: 4 frames ou 20 ms. Opção do usuário
- Limiar superior de duração: 60 frames ou 300 ms. Opção do usuário.
- Limiar de intensidade: 2000 RMS. Opção do usuário.

A partir do tom médio (TM), são estabelecidas as bandas de frequência, nas quais as variações de F0 serão plotadas. As bandas de frequência são necessárias, pois, diferentemente das variações melódicas musicais, as variações da fala são mais largas do ponto de vista perceptual (3 e 4 st), tal como demonstrado nos trabalhos supracitados. Como resultado final, a rotina produz um gráfico de distribuição temporal de frequência e de intensidade,

²⁷ Esses parâmetros perceptivos foram encontrados por trabalhos desenvolvidos pelo grupo *ExProsodia* e estão sujeitos a revisões à medida que novos problemas possam aparecer em trabalhos futuros.

categorizada com a transcrição²⁸ dos dados em cada ponto plotado no gráfico.

A rotina também fornece as tabelas detalhadas referentes a esse gráfico.

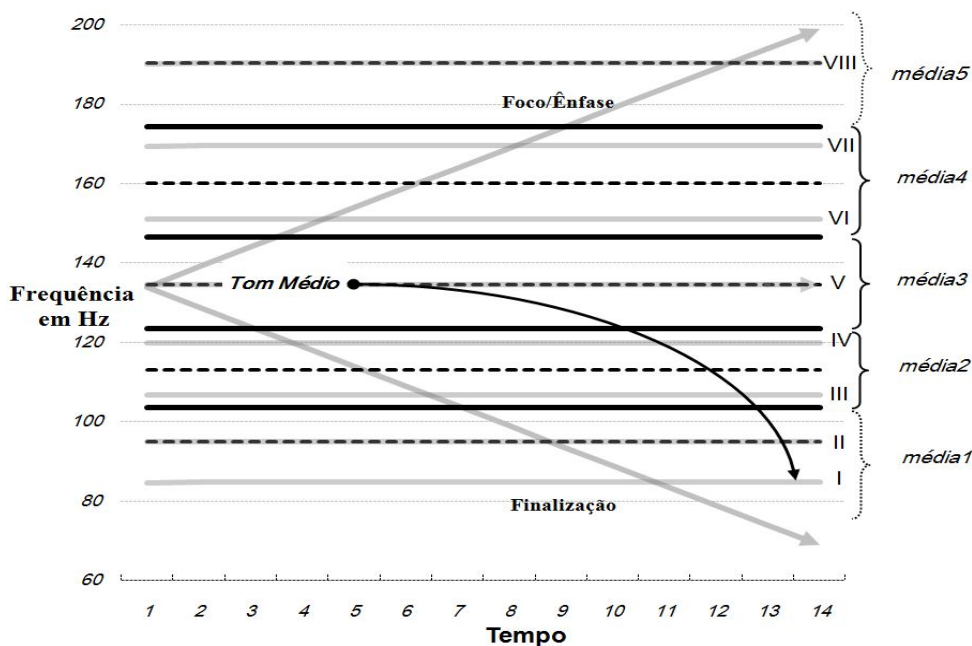


Figura 16: Gráfico da escala de tons. As linhas pontilhadas mostram os valores médios de cada banda de frequência. As linhas pretas contínuas mostram os limites de cada banda. As linhas cinzas mostram os pontos exatos de cada grau da escala diatônica ocidental (algarismos romanos à direita do gráfico), tal como em Pierce, (1983). A linha curva central mostra a tendência à finalização da frase, e as linhas diagonais apontam para foco e ênfase e para a finalização (FERREIRA NETTO, 2008).

Um exemplo de uso da rotina *ExProsodia* pode ser visto na figura 17:

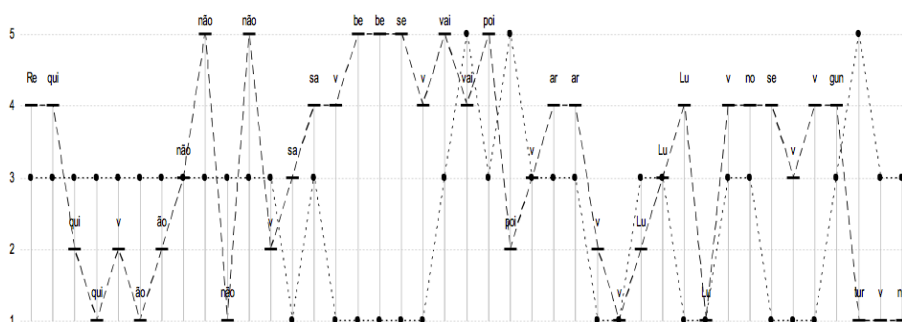


Figura 17: Gráfico produzido pela rotina *ExProsodia*. Os traços marcadores pretos mostram a entoação na escala de tons e os círculos marcadores cinza mostram a intensidade.

²⁸ Caso a transcrição seja feita manualmente no SFS e fornecida à rotina *ExProsodia*.

Para a demonstração dos resultados da decomposição de F0 em componentes estruturais e semântico-funcionais, e sabendo-se que a série temporal proposta é $Z_t = S_t + F_t + E_t (+A_t)$, pode-se fazer a decomposição dos momentos Z_t a partir de TM_{t-1} e definindo-se F_t como $TM_{t-1} - 7st$. Para a definição de E_t com Z_t maior do que TM_{t-1} extrai-se $TM_{t-1} + 1,5st$ de Z_{t-1} ; para a definição de E_t com Z_t menor do que TM_{t-1} extrai-se $F_{t-1} + 1,5st$ de Z_{t-1} . Para a definição de S_t , apenas extraem-se F_t e E_t de Z_t (FERREIRA NETTO, 2008). O resultado obtido pela rotina para uma frase qualquer é o seguinte:

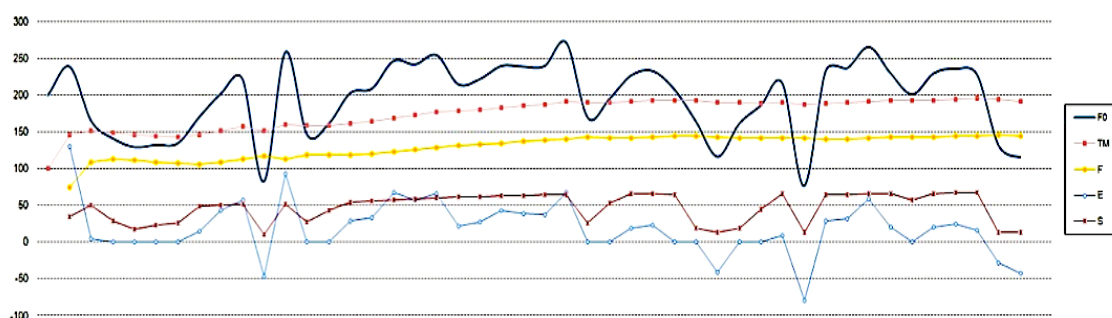


Figura 18: Decomposição da série temporal feita pela rotina *ExProsodia*. A linha azul escuro é a soma das componentes: $Z_t = F_t + S_t + E_t$, obtida nos pontos significativos: os candidatos a pico silábico (cps). As linhas amarela, azul claro e vinho são as componentes “finalização”, “foco/ênfase” e “sustentação” respectivamente. A linha rosa representa o tom médio (TM) das frequências.

Neste trabalho, a rotina *ExProsodia* será utilizada para a extração de tom médio e finalização de cada estímulo utilizado no experimento de percepção. Esses valores também serão usados na análise de produção. A aplicação e resultado da decomposição de F0 podem ser encontrados em Vergés et al. (2008) sobre uma análise de distinção de gênero em português.

6. ANÁLISE DOS DADOS

6.1 Análise com a Teoria da Detecção

Após a explanação sobre a Teoria da Detecção dada no subcapítulo 4.2.3, pode-se iniciar a análise dos dados. As tabelas abaixo apresentam os resultados de cada participante em cada variedade no experimento 1 com técnica PURR, envolvendo trechos curtos e longos:

Tabelas dos dados (CE)		
Informante	TC	TL
1	1,72	1,02
2	0,76	0,57
3	1,66	1,53
4	1,61	0,71
5	0,57	0
6	1,2	0,9
7	1,35	0,57
8	1,05	1,42
9	0,9	0,42
10	1,65	0,57
11	0	-0,43
12	1,05	0,42
13	0,57	1,29
Média	1,08	0,69
Desvio Padrão	0,52	0,55

Tabela 3: Resultados d' obtidos no experimento 1 (PURR) para cada participante de Senador Pompeu (CE).

Tabelas dos dados (SP)		
Informante	TC	TL
1	1,02	1,36
2	1,11	0,87
3	0,71	0,9
4	1,36	0,45
5	0,57	0,72
6	1,2	1,12
7	1,53	1,2
8	0,9	0,86
9	1,05	0,28
10	1,4	1,5
11	1,36	1,02
12	0,57	0,71
13	0,83	1,03
Média	1,05	0,92
Desvio Padrão	0,32	0,34

Tabela 4: Resultados d' obtidos no experimento 1 (PURR) para cada participante de São Paulo (SP).

Tabelas dos dados (RS)		
Informante	TC	TL
1	0,71	0,57
2	1,11	1,4
3	0,86	0,42
4	1,2	1,08
5	0,51	0,14
6	0,86	1,05
7	0,57	1,2
8	1,4	0,86
9	0,58	0,28
10	1,2	1,5
11	1,36	1,02
12	1,73	0,71
13	0,71	0,73
Média	0,98	0,84
Desvio Padrão	0,38	0,42

Tabela 5: Resultados d' obtidos no experimento 1 (PURR) para cada participante de Pelotas (RS).

Os resultados demonstram que há uma certa variabilidade nos valores d' obtidos. Alguns informantes exibiram maior sensibilidade para a discriminação da própria variedade, bem como a negação das variedades alheias. Quanto

maior o valor d' , melhor foi a tarefa de reconhecimento. Os informantes senadorenses 1, 3, 4, 7, 8, 10 e 12, em trechos curtos, e 1, 3, 6, 8, e 13, em trechos longos, tiveram um desempenho satisfatório (tabela 3). De maneira análoga, os informantes paulistanos 1, 2, 4, 6, 7, 9 e 11, em trechos curtos, e 1, 6, 11 e 13, em trechos longos, tiveram um bom desempenho no reconhecimento de sua variedade (tabela 4). O mesmo se deu para os informantes pelotenses 2, 4, 10, 11 e 12, em trechos curtos, e 2, 4, 6, 10 e 11, em trechos longos (tabela 5). Por outro lado, pode-se observar que alguns informantes não tiveram uma *performance* satisfatória quanto ao reconhecimento, seja pelo não reconhecimento da própria variedade ou pelos alarmes falsos na aceitação de variedades alheias. Nesses casos, o valor d' ficou próximo de 0 ou negativo quando a taxa de alarmes falsos superou a taxa de acertos. Exemplos de baixo desempenho podem ser vistos nos informantes senadorenses 5 e 11, nos informantes paulistanos 5 e 12 e nos informantes pelotenses 5 e 9, em trechos curtos e longos.

Para visualizar de maneira mais clara essa relação, foi utilizada a curva ROC.²⁹ A curva ROC é um gráfico de sensibilidade que relaciona a taxa de acertos (verdadeiros positivos) com a taxa de alarmes falsos (falsos positivos). A utilização primeira desse tipo de gráfico ocorreu no campo das comunicações como uma forma de demonstrar as relações entre sinal e ruído. Novamente, sinal e ruído serão interpretados metaforicamente como variedade do participante (sinal) e variedade alheia (ruído). A seguir serão colocados alguns exemplos de curva ROC para alguns informantes do experimento 1 (PURR). O informante 1 (TC [CE]), na tabela 2, apresentou o índice $d' = 1,72$, resultando na seguinte curva ROC:³⁰

²⁹ Receiver Operating Characteristic.

³⁰ As demais curvas ROC para cada um dos informantes estão em anexo.

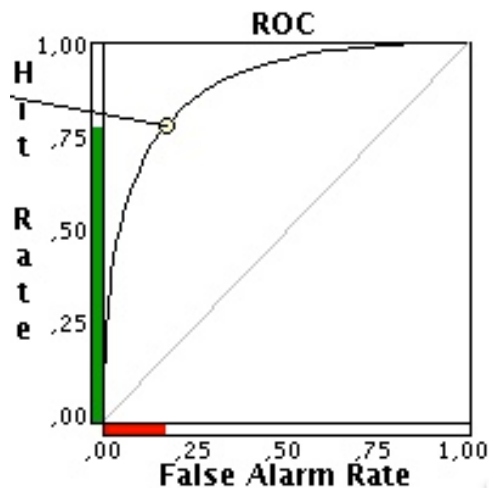


Gráfico 2:³¹ Resultado do informante 1 TC (CE).

O resultado acima demonstrou que o informante 1 teve êxito na tarefa de discriminação das variedades manipuladas com técnica PURR. O eixo y indica qual a taxa de acertos, na proporção de 0,78, o eixo x indica a taxa de alarmes falsos, na proporção de 0,17. No exemplo abaixo, a curva ROC demonstra um desempenho menor do informante 9 (TL [RS]):

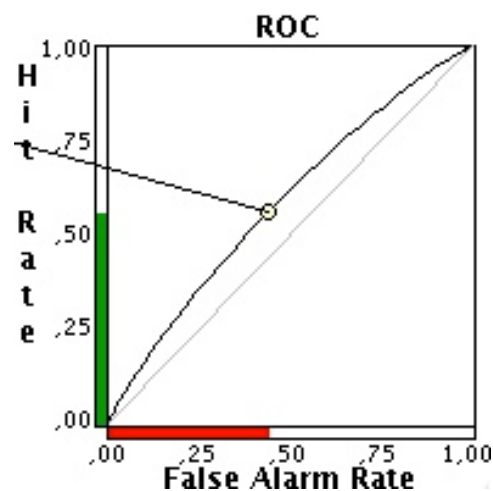


Gráfico 3: Resultado do informante 9 TL (RS).

O eixo y indica a taxa de acertos, na proporção de 0,56, o eixo x indica a taxa de alarmes falsos, na proporção de 0,44. O valor d' para essas taxas de

³¹ Os gráficos ROC foram feitos pelo site <http://wise.cgu.edu/sdtmod/measures5.asp>.

acertos e alarmes falsos foi de 0,28. Poucos informantes tiveram $d' = 0$ ou $d' < 0$, obtendo as seguintes representações na curva ROC, respectivamente:

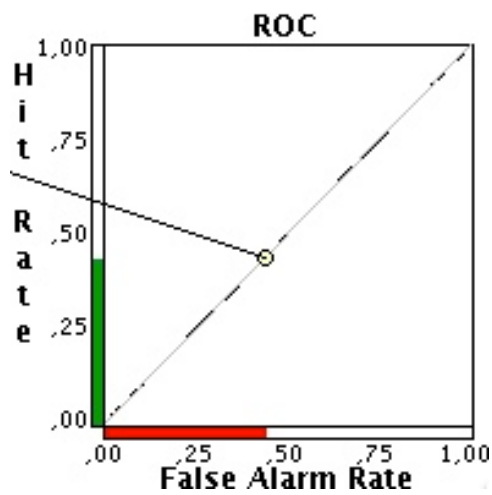


Gráfico 4: Resultado do informante 11 TC (CE) para $d' = 0$.

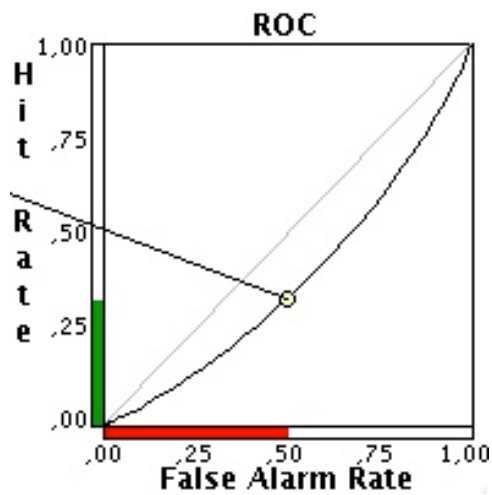


Gráfico 5: Resultado do informante 11 TL (CE) para $d' < 0$.

A tabela com os resultados de cada participante para o experimento 2 com curva melódica monotônica (CMM) está a seguir:

Inf.	(CE)	Inf.	(SP)	Inf.	(RS)
1	1,64	1	2,35	1	2,35
2	1,27	2	2,35	2	2,35
3	2,54	3	1,93	3	2,35
4	3,29	4	2,76	4	2,76
5	1,18	5	0,97	5	0,97
6	1,4	6	1,93	6	1,93
7	3,29	7	2,76	7	1,93
8	1,4	8	2,35	8	2,76
9	0,43	9	1,81	9	1,81
10	0,86	10	0,43	10	0,43
11	1,71	11	1,93	11	1,93
12	1,93	12	1,93	12	1,93
13	1,93	13	1,81	13	1,81
Média	1,76	Média	1,95	Média	1,95
Desvio Padrão	0,86	Desvio Padrão	0,65	Desvio Padrão	0,65

Tabela 6: Resultados d' para curva melódica monotônica (CMM) obtidos para cada participante em cada região.

Os resultados demonstram que não houve a mesma variabilidade encontrada no experimento 1 (PURR). Os valores d' obtidos mantiveram-se altos com relação aos índices do experimento 1. Parece que a tarefa de reconhecimento das variedades torna-se mais árdua quando os segmentos estão ausentes. Para ilustrar, duas curvas ROC foram feitas com o melhor e o pior desempenho no experimento 2 (CMM):

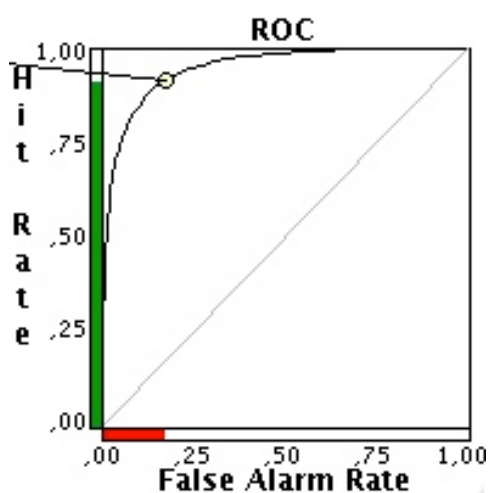


Gráfico 6: Resultados do informante 4 (CE).

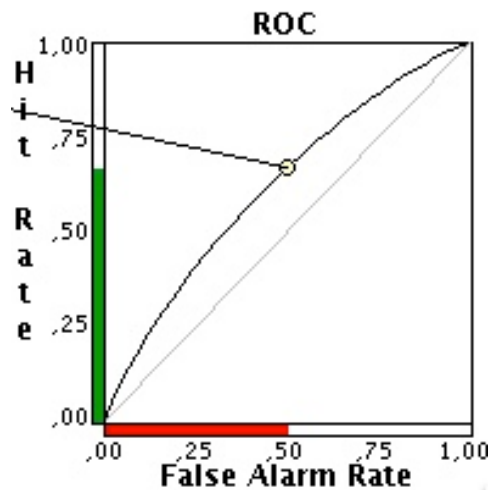


Gráfico 7: Resultado do informante 5 (CE).

O resultado do gráfico 6 demonstrou que o informante 4 (CE) obteve êxito na tarefa de discriminação das variedades manipuladas com curva melódica monotônica.

O resultado do gráfico 6, por outro lado, demonstrou um menor desempenho do informante 5 (CE) na tarefa de reconhecimento proposta para o experimento 2.

Após a análise intra-sujeito baseada na TD, pode-se ter um panorama de como foi o desempenho dos informantes nos dois experimentos desenvolvidos. O próximo passo da análise abordará o desempenho dos informantes pensando em categorias regionais.

6.2 Análise geral dos dados do experimento 1

Nesta seção, será feita uma análise geral dos dados, os quais serão abordados pelo reconhecimento das variedades regionais e pela duração dos estímulos, este último critério avaliado somente no experimento 1 (PURR). O primeiro desempenho avaliado foi o dos informantes (ouvintes) de Pelotas

(RS). A tabela abaixo representa os acertos e os erros dos informantes de Pelotas no experimento 1:

PURR (RS)		
	acertos	erros
Valores brutos	468	234
Porcentagem	67%	33%

Tabela 7: Desempenho – RS.

É importante esclarecer que se entende por acerto todos os reconhecimentos da variedade em questão e todas as negações das variedades alheias, nos termos da TD, todos os acertos e rejeições corretas somados. Os acertos dos 13 informantes pelotenses estão distribuídos entre trechos curtos (TC) e longos (TL) no gráfico abaixo:

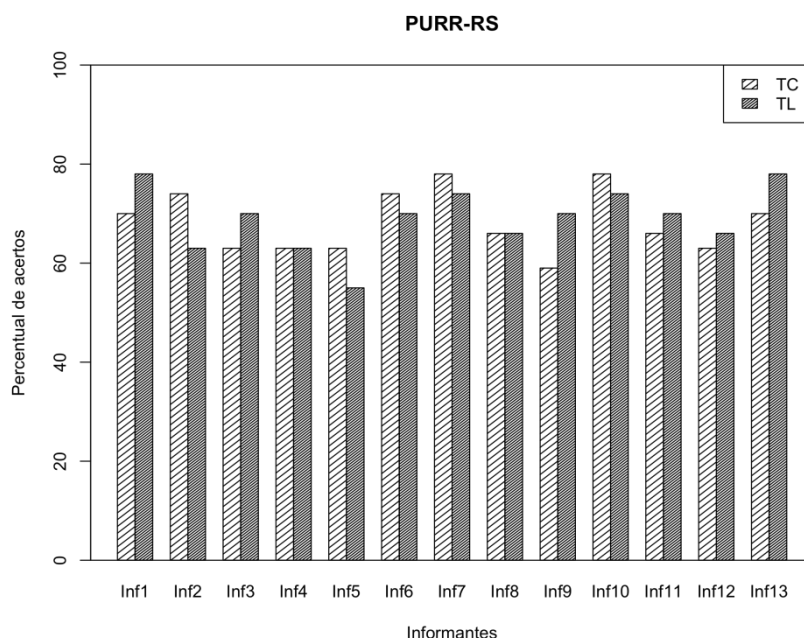


Gráfico 8: Distribuição de acertos do experimento 1 (PURR) por duração de estímulo (RS).

A média de acertos dos informantes pelotenses foi de 67%, tanto em trechos curtos (TC), como em trechos longos (TL), para as duas sequências de

27 estímulos cada. Passando ao desempenho dos informantes paulistanos, temos os seguintes valores para acertos e erros:

PURR (SP)		
	Acertos	erros
Valores brutos	483	219
Porcentagem	69%	31%

Tabela 8: Desempenho – SP.

O desempenho dos paulistanos foi similar ao dos pelotenses, obtendo uma média de acertos para trechos curtos de 66%, diferenciando-se sutilmente nos trechos longos, os quais obtiveram uma média de acerto de 70%. O gráfico abaixo ilustra o desempenho dos 13 informantes paulistanos no experimento 1:

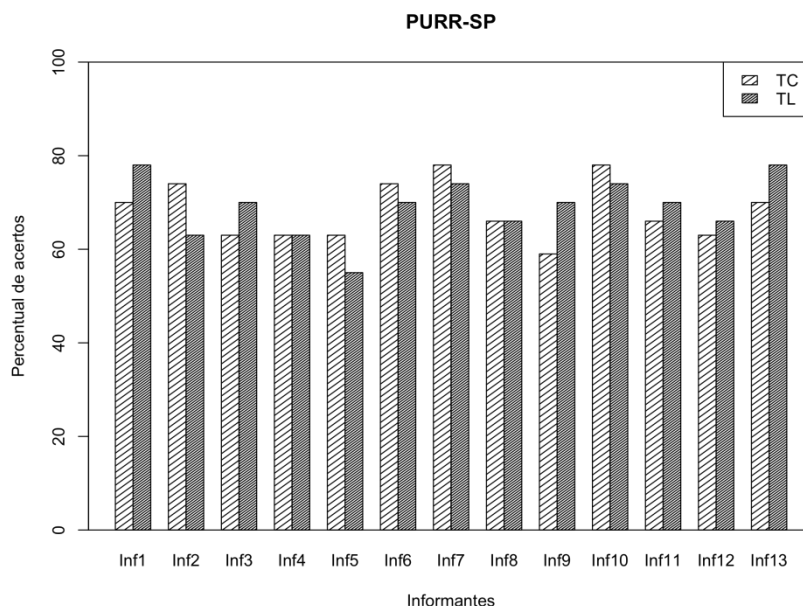


Gráfico 9: Distribuição de acertos do experimento 1 (PURR) por duração de estímulo (SP).

Passando ao desempenho dos informantes senadorense, houve os seguintes valores de acertos e erros:

PURR (CE)		
	acertos	Erros
Valores brutos	443	259
Porcentagem	63%	37%

Tabela 9: Desempenho – CE.

Os acertos dos informantes cearenses foram ligeiramente inferiores aos demais, apresentando uma média de 63% de acertos para trechos curtos e longos. O gráfico a seguir ilustra o desempenho dos 13 informantes senadorenses no experimento 1:

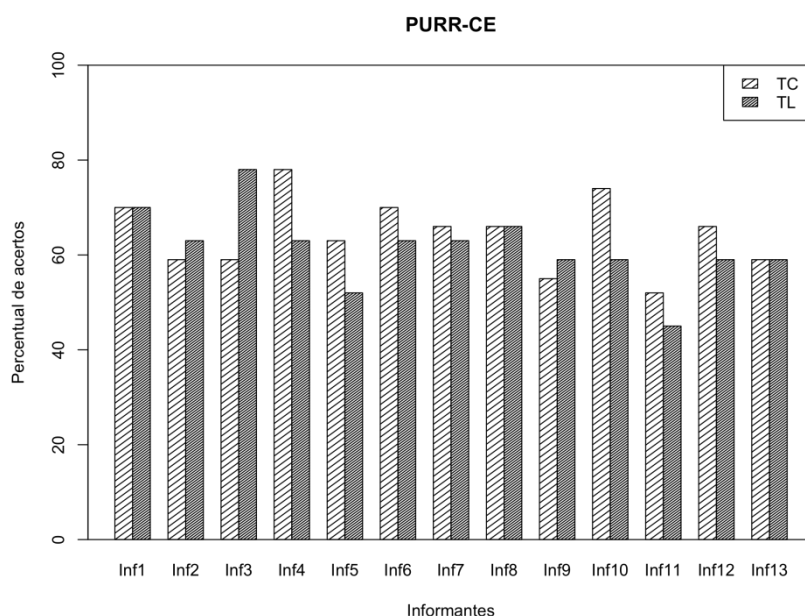


Gráfico 10: Distribuição de acertos do experimento 1 (PURR) por duração de estímulo (CE).

No desempenho dos senadorenses, nota-se, visualmente, que há algumas diferenças pontuais mais salientes entre trechos longos e curtos. O papel da duração dos estímulos no reconhecimento das variedades regionais será testado adiante. Para se ter um panorama geral de desempenho entre as variedades, o gráfico abaixo ilustra as três variedades e seus acertos e erros no experimento 1 (PURR):

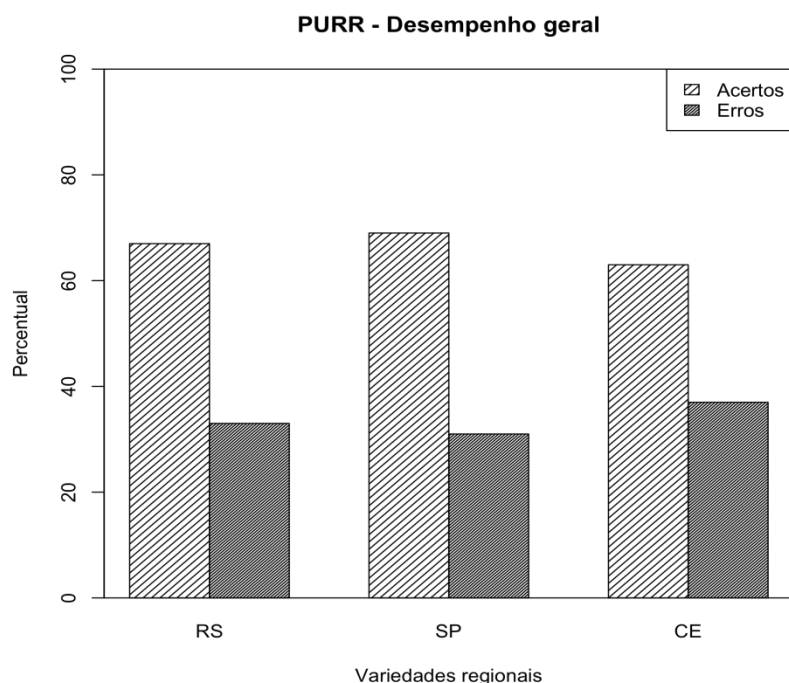


Gráfico 11: Acertos e erros para cada variedade no experimento 1 (PURR).

A exposição dos dados acima é apenas uma visão global do desempenho dos participantes, mas não é possível afirmar alguma tendência, excetuando-se o fato de que os acertos tiveram vantagem com relação aos erros, nas três variedades.

Partindo para a análise probabilística dos dados, foi feito um teste ANOVA de um fator, envolvendo os dados brutos de acertos e a duração dos estímulos dados aos informantes (ouvintes) no experimento 1 (PURR). O resultado não foi significativo: $F(1, 72) = 0,63$, $p > 0,05$. Com isso, o fator duração de estímulo parece não exercer influência no desempenho dos informantes.

Um outro teste ANOVA de 1 fator foi feito, agora tendo como fator a variedade dos informantes. O resultado foi significativo: $F(2, 75) = 4,04$, $p < 0,05$. Para constatar qual era a diferença significativa entre as variedades, foi feito um teste de Tukey. O resultado apontou uma diferença significativa entre as variedades de SP e CE, e entre SP e RS ($p < 0,05$). A seguir, os gráficos 12

e 13 mostram os resultados dos fatores duração de estímulo e variedades regionais, respectivamente:

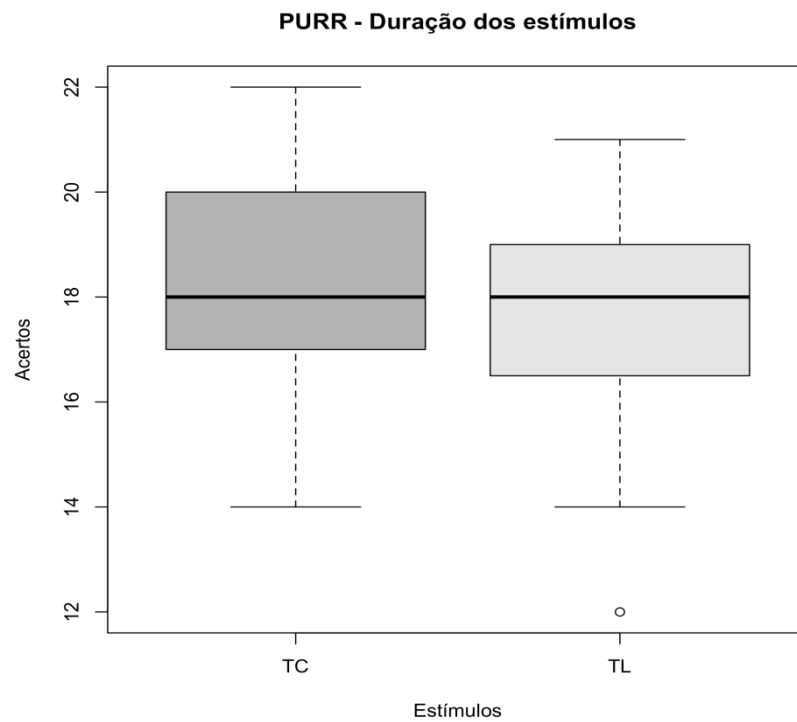


Gráfico 12: Boxplot de acertos e duração de estímulo no experimento 1 (PURR).

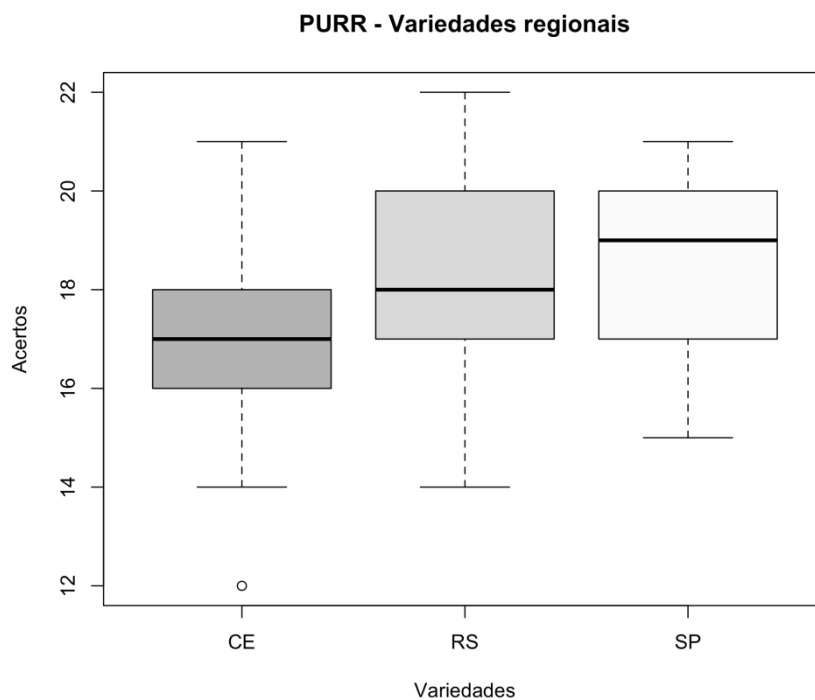


Gráfico 13: Boxplot de acertos e variedades regionais no experimento 1 (PURR).

O gráfico abaixo ilustra as diferenças entre as médias das variedades regionais obtidas no teste de Tukey:

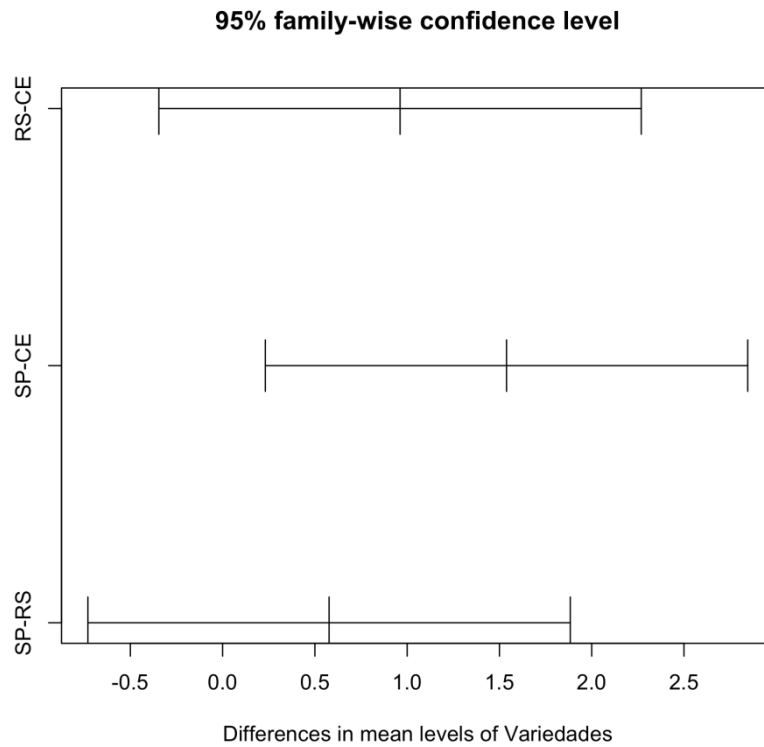


Gráfico 14: Diferenças entre as médias das variedades regionais – Tukey.

Num segundo momento da análise do experimento 1 (PURR), foi feito um teste ANOVA de 1 fator, mas com os índices *d-prime* obtidos no subcapítulo 6.1. O fator foi a duração de estímulos e o resultado obtido foi $F(1,76) = 5,14$, $p < 0,05$. Os índices *d-prime* demonstraram que o desempenho dos informantes pode ter sido influenciado pela duração do estímulos, o qual não ocorreu com os acertos brutos.

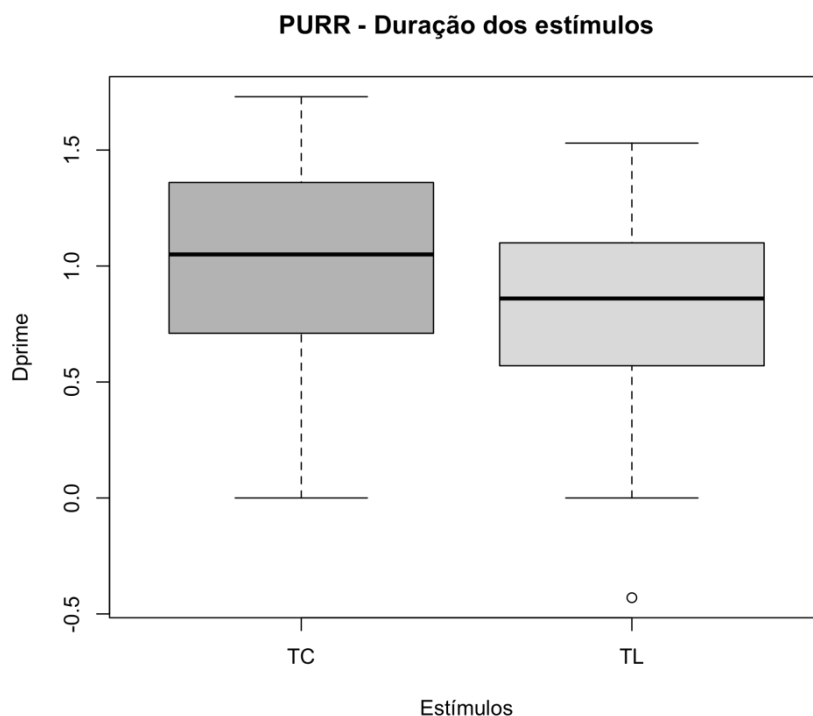


Gráfico 15: Boxplot de índices *d-prime* e duração de estímulos no experimento 1 (PURR).

O resultado diverso do teste ANOVA envolvendo duração de estímulo e *d-prime* pode ser atribuído ao índice d' , pois este pode fornecer uma informação mais detalhada do desempenho de cada informante, por levar em consideração também os erros cometidos. Nas análises feitas no experimento 1, pode-se considerar que, com os dados brutos, o reconhecimento da variedade de São Paulo teve um comportamento diverso em relação ao desempenho no reconhecimento da variedade do Ceará. Na discussão deste trabalho, tratar-se-á da relevância da utilização da TD.

A distribuição das médias de acertos por variedade regional e os índices *d-prime* por duração de estímulo estão nos gráficos 16 e 17, respectivamente:

PURR - Média de acertos

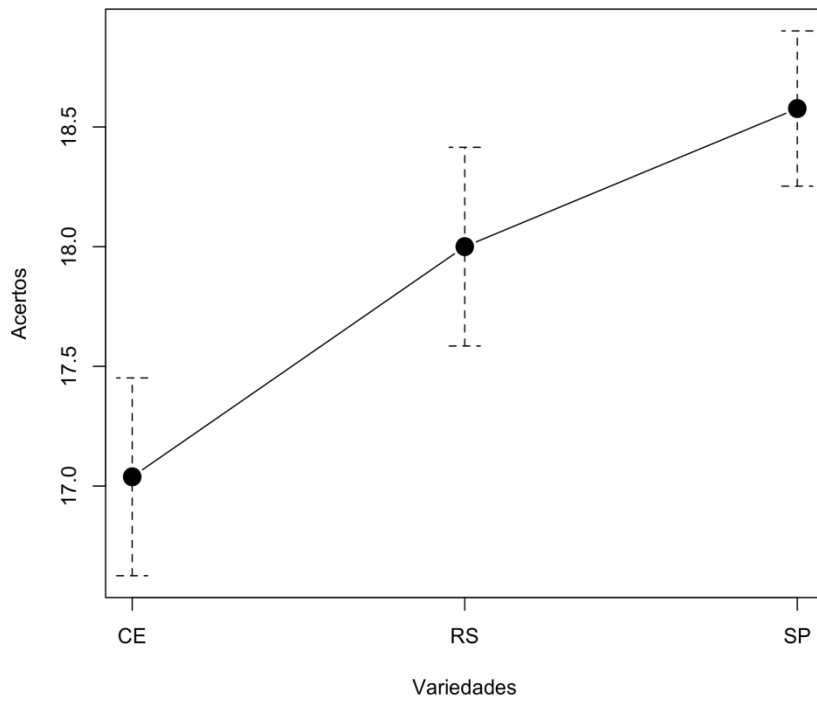


Gráfico 16: Médias de acertos por variedade regional.

Médias d-prime

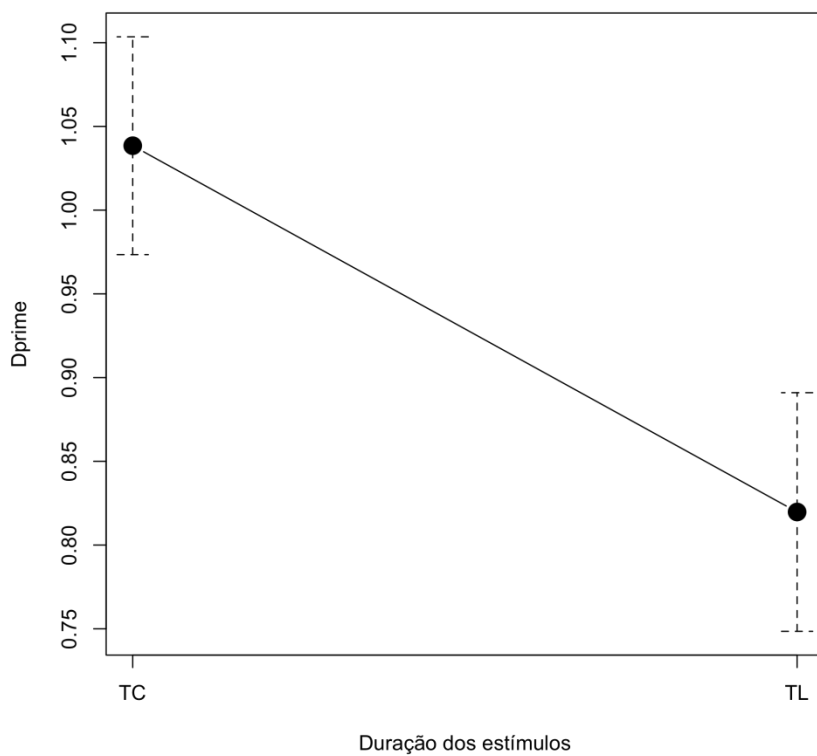


Gráfico 17: Médias de índices *d-prime* por duração de estímulo.

6.3 Análise geral dos dados do experimento 2

O experimento 2, formado por trechos de fala com curva melódica monotônica (CMM), foi apresentado ao mesmo grupo que participou do experimento 1 (PURR).

O desempenho dos 13 informantes de cada variedade regional está disposto no gráfico abaixo:

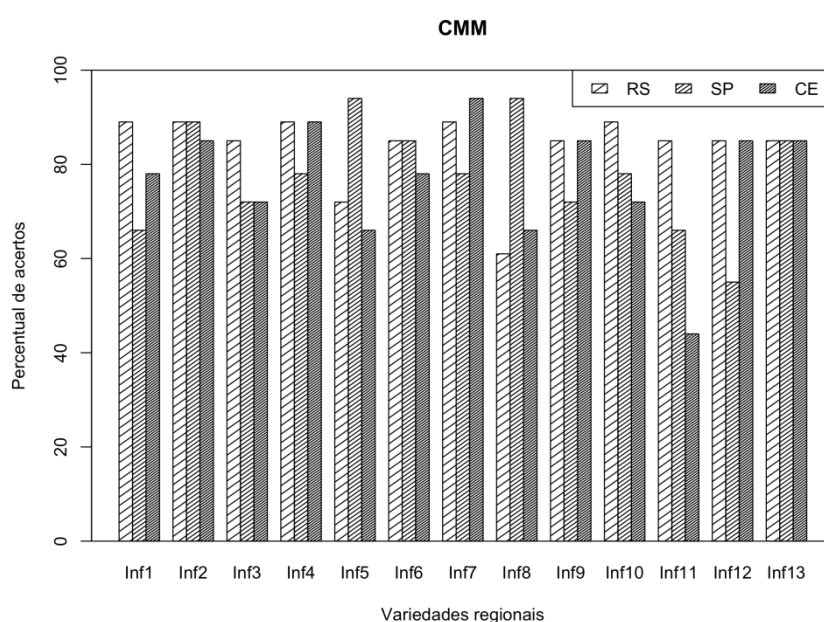


Gráfico 18: Desempenho dos 13 informantes de cada variedade no experimento 2 (CMM).

Parece haver uma regularidade entre os informantes quanto ao desempenho, excetuando-se algumas disparidades mais visíveis numa primeira análise. Abaixo, há o gráfico de desempenho geral dos informantes na tarefa de reconhecimento do experimento 2. À primeira vista, a diferença entre acertos e erros parece ser proeminente.

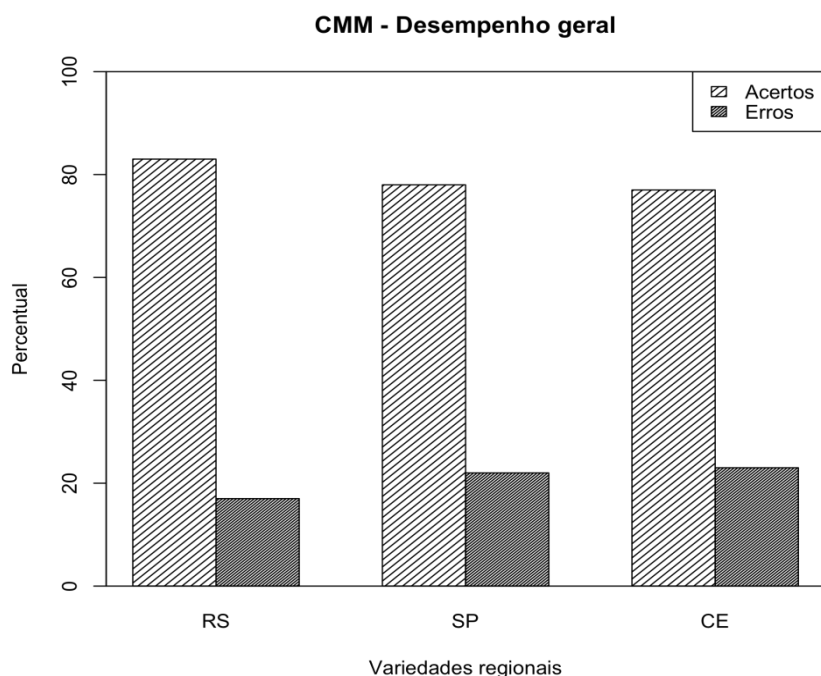


Gráfico 19: Total de erros e acertos por variedade no experimento 2 (CMM).

Seguindo a mesma metodologia de análise do experimento 1, tentou-se verificar o papel da variável “variedade regional” no nível de acertos e nos índices *d-prime*. A análise prévia de resíduos, via teste de normalidade de Shapiro-Wilk, demonstrou que a distribuição dos acertos não era normal ($p < 0,05$). A análise de homocedasticidade, via teste de Fligner-Killeen, apontou para uma distribuição não homogênea da variâncias ($p < 0,05$). Não cumpridos os pressupostos para a feitura da ANOVA, foi feito um teste de Kruskal-Wallis. O resultado não foi significativo. Os acertos parecem não ter um desempenho diferenciado por variedades.

A análise do experimento 2 (CMM) com índice *d-prime* também não passou no testes de pressupostos da ANOVA. Assim, foi feito um teste de Kruskal-Wallis, apontando para um resultado não significativo ($p > 0,05$). Portanto, o fator variedade regional não apresentou diferenças significativas.

Os gráficos 20 e 21 apresentam os acertos por variedade e os índices *d-prime* por variedade regional, respectivamente:

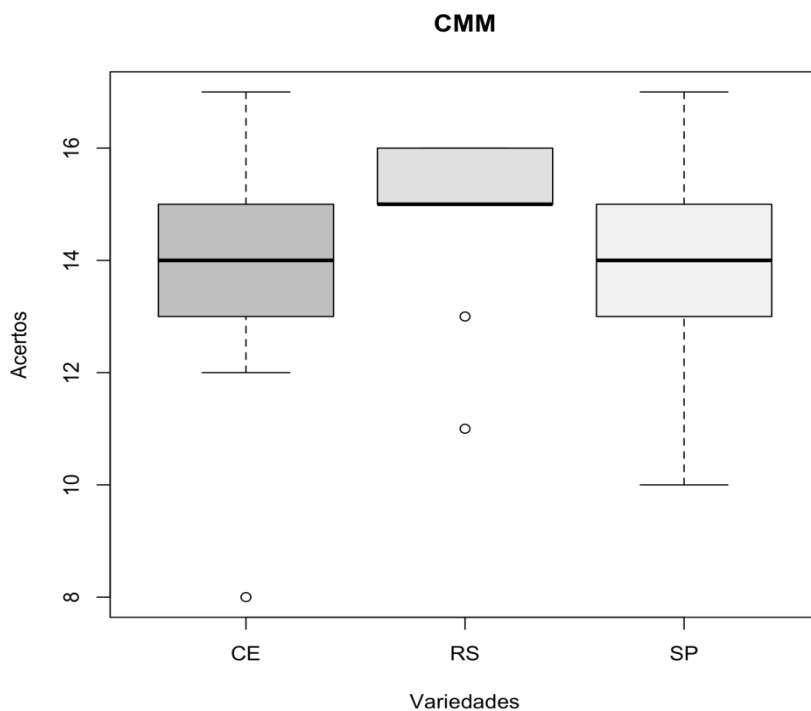


Gráfico 20: Boxplot de acertos e variedades regionais no experimento 2 (CMM).

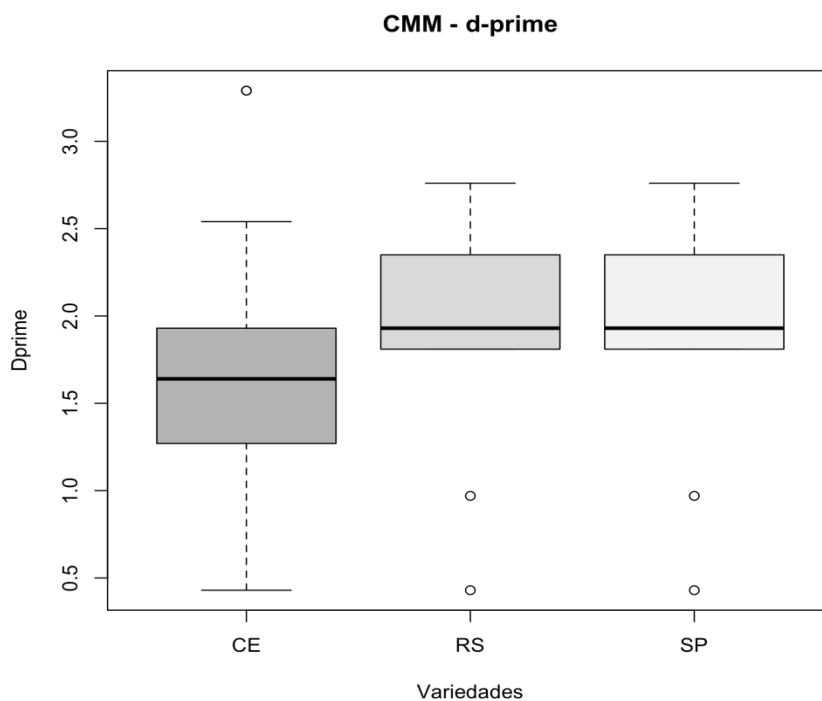


Gráfico 21: Boxplot de índices *d-prime* e variedades regionais no experimento 2 (CMM).

A análise dos dados dos experimentos 1 e 2 demonstrou que a tarefa de reconhecimento proposta no experimento 1, com PURR, foi mais árdua para os informantes do que a proposta no experimento 2, com CMM. Era esperado, de maneira análoga ao ocorrido no trabalho de Bezooijen e Gooskens (1999), que a tarefa se tornaria mais difícil quando a parte “verbal”, nos termos dos autores, fosse retirada. De fato, o índice de acertos no experimento 2 (CMM) foi de 75%, em média, ao passo que o experimento 1 (PURR) teve, em média, 62% de acerto. A diferença entre os resultados dos experimentos 1 e 2 foi verificada por meio do teste de Kruskal-Wallis ($p < 0,05$). A maior dificuldade de reconhecimento de estímulos delexicalizados apontada por Bezooijen e Gooskens (1999) é sustentada neste trabalho. Os gráficos 22 e 23 ilustram as diferenças de desempenho nos dois tipos de experimento:

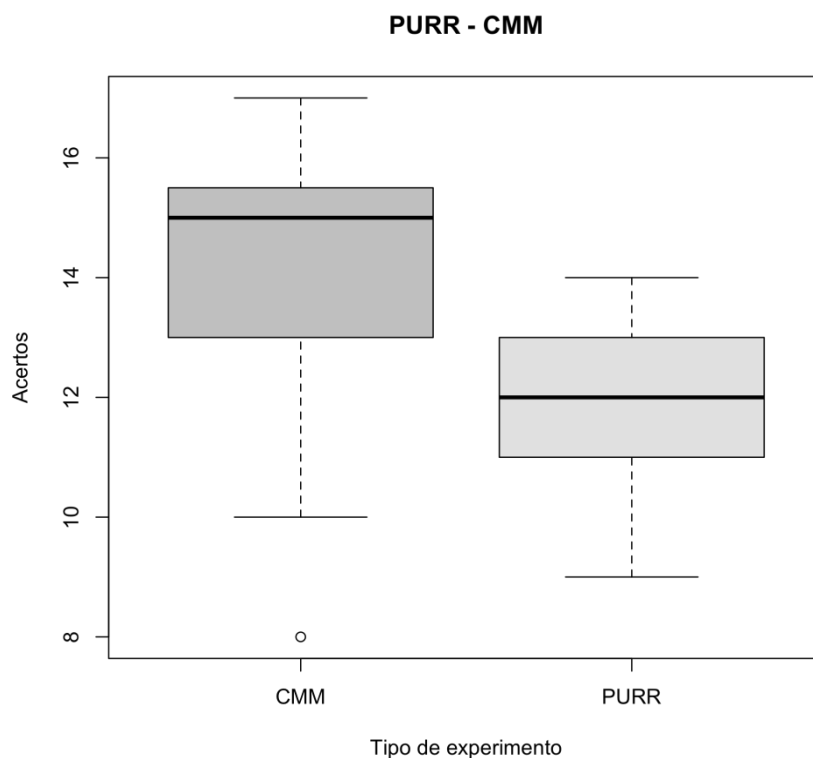


Gráfico 22: Boxplot de acertos e tipos de experimento (PURR vs CMM).

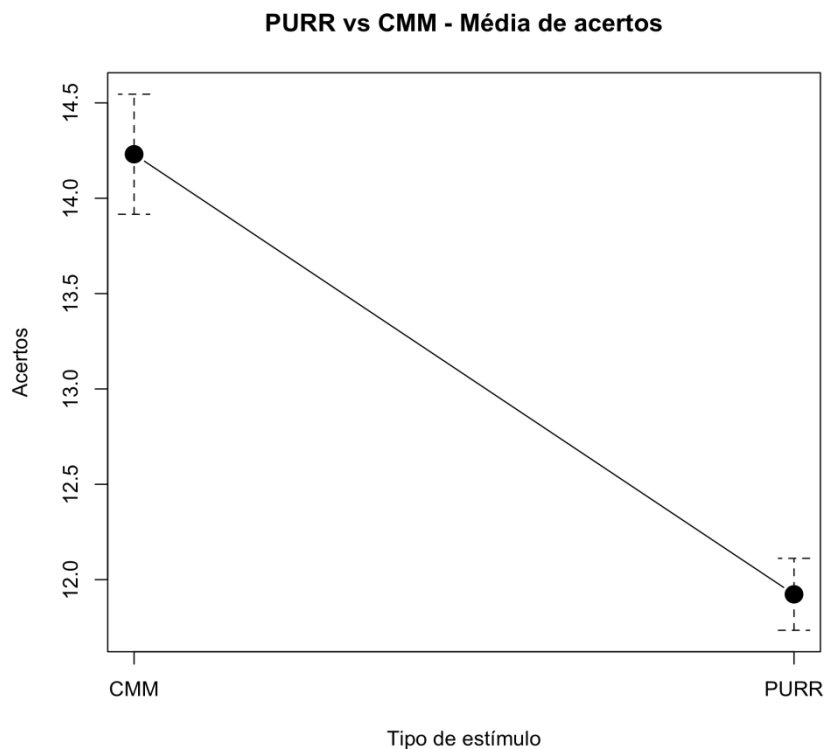


Gráfico 23: Médias de acertos nos experimentos 1 e 2.

6.4 Análise de produção

A análise de produção feita neste trabalho partiu dos dados fornecidos via *ExProsodia*. É importante ressaltar que a análise feita abordou tão somente as variações do parâmetro F0, pelo entendimento de que uma análise com os demais parâmetros prosódicos acarretaria um trabalho que não se encerraria nesta pesquisa. Para a análise de F0, tomou-se os *cps* (candidatos a picos silábicos) gerados pelo *ExProsodia*, obedecendo aos limiares fornecidos no capítulo 5. Três características foram abordadas: i) finalização dos estímulos; ii) tom médio dos estímulos; iii) dispersão das variações de F0 (em Hz). No primeiro passo da análise, finalização entre as variedades, o teste de Kruskal-Wallis forneceu um resultado não significativo ($p > 0,05$). O gráfico 24

demonstra os valores de finalização por variedade regional. É possível notar alguns valores espúrios nas variedades do Ceará e de São Paulo.

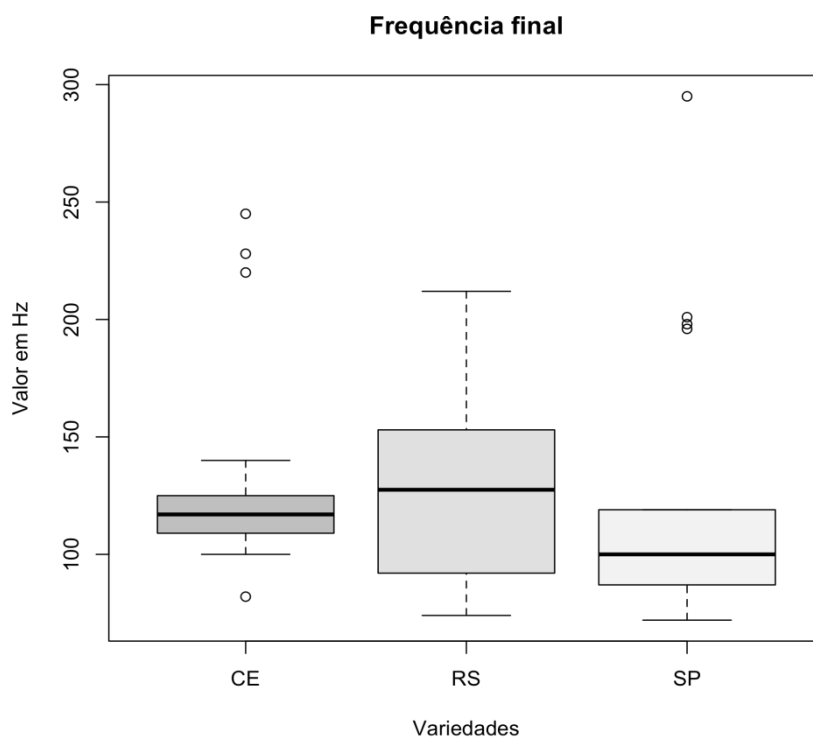


Gráfico 24: Boxplot de finalização dos estímulos e variedades regionais.

Com esse resultado, a finalização dos estímulos não apresentou diferença quanto à variedade regional dos falantes, fazendo com que essa característica seja descartada como fator de distinção entre os estímulos.

No segundo passo da análise, tom médio entre as variedades, um teste ANOVA de 1 fator apresentou resultado significativo: $F(2,51) = 17,09$, $p < 0,05$. O resultado apontou uma diferença significativa entre os tons médios dos estímulos de cada variedade regional. Para certificar quais variedades diferiram entre si, foi feito um teste de Tukey. O resultado apontou para uma diferença significativa entre as variedades paulistana e senadoreense, e entre as variedades pelotense e paulistana ($p < 0,05$). Os gráficos 25 e 26 demonstram as diferenças de tom médio por variedade e as diferenças obtidas pelo teste de

Tukey, respectivamente. O gráfico 27 mostra os valores dos tons médios dos estímulos por variedade.

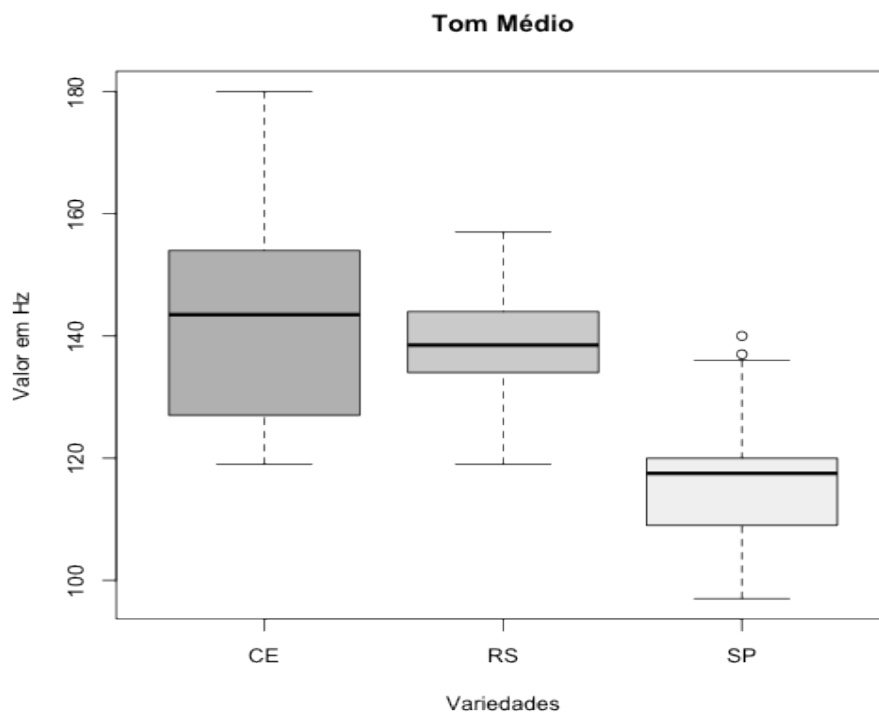


Gráfico 25: Boxplot de tom médio dos estímulos e variedades regionais.

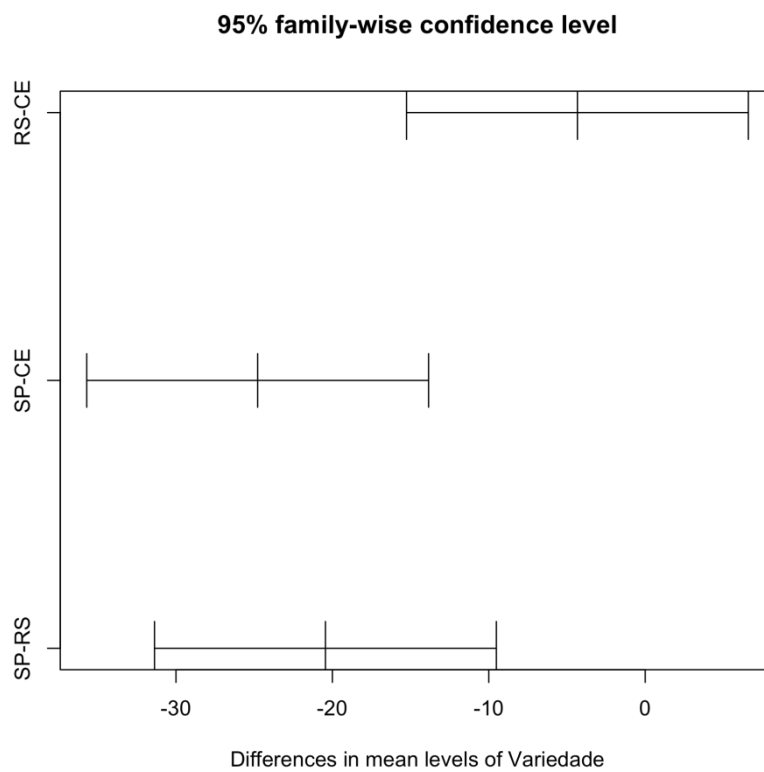


Gráfico 26: Diferenças entre as médias de tom médio e variedades regionais – Tukey.

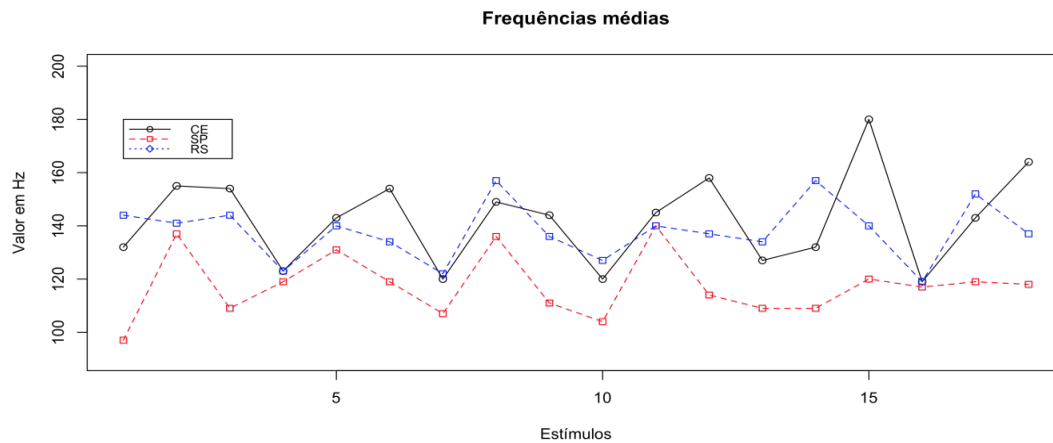


Gráfico 27: Tons médios dos estímulos por variedades regionais.

No terceiro passo da análise, dispersão de F0, um teste ANOVA de um fator apresentou resultado significativo: $F(2, 51) = 18,04, p < 0,05$. No teste de Tukey, as variedades que apresentaram diferença significativa entre si foram a de São Paulo e Ceará, de um lado, e a do Rio Grande do Sul e São Paulo, de outro ($p < 0, 05$). O gráfico abaixo ilustra as três variedades regionais e suas respectivas dispersões de F0.

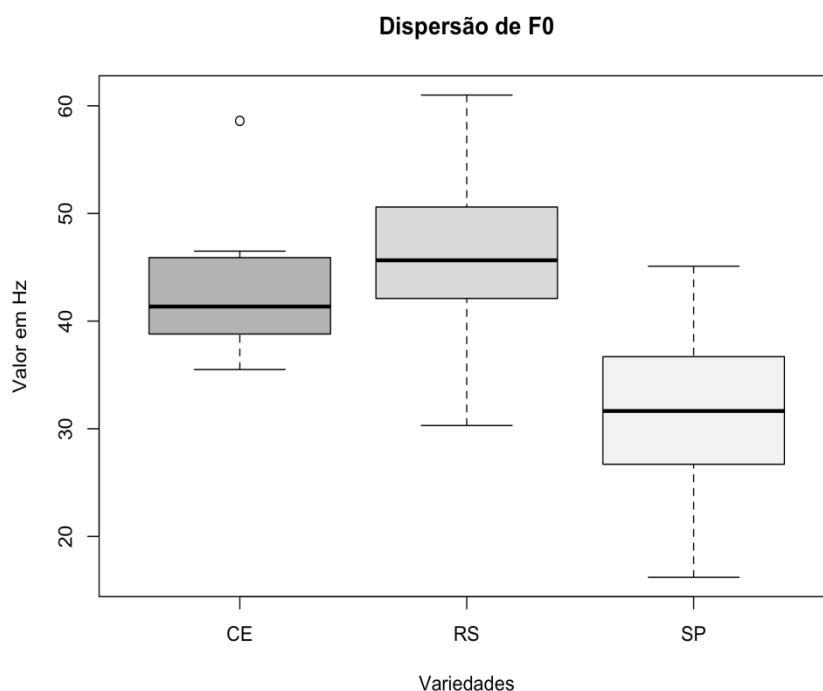


Gráfico 28: Boxplot de dispersão de F0 e variedades regionais.

As diferenças obtidas pelo teste de Tukey estão expostas no gráfico abaixo:

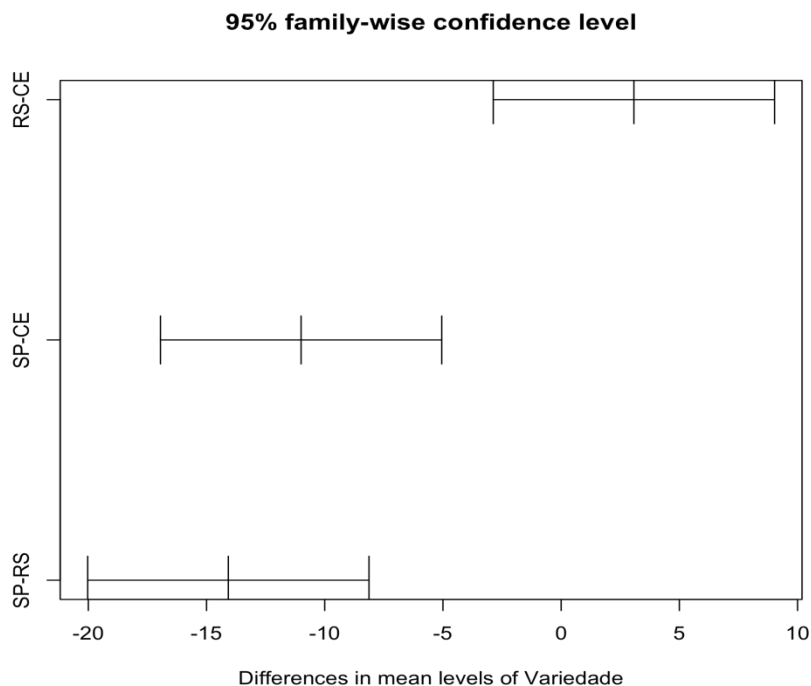


Gráfico 29: Diferenças entre médias de dispersão de F0 entre variedades regionais – Tukey.

Das três possíveis características que poderiam estar influenciando a escolha, o tom médio dos estímulos e a dispersão de F0 apresentaram resultado significativo, ao passo que a finalização não teve resultado significativo. A implicação dessas características na escolha das variedades regionais pelos ouvintes será tratada no próximo capítulo.

7. DISCUSSÃO

Antes do início da discussão dos resultados, cabe aqui uma síntese das análises feitas no capítulo anterior:

	Testes	Resultados
Experimento 1 (PURR)	Anova – 1 fator – variável dependente – acertos; variável independente – duração de estímulo (TL vs TC).	$F(1, 72) = 0,63, p > 0,05$
	Anova – 1 fator – variável dependente – acertos; variável independente – variedades regionais (SP, CE, RS).	$F(2, 75) = 4,04, p < 0,05$
Experimento 2 (CMM)	Teste de Tukey - acertos e variedades regionais (SP, CE, RS).	Diferença significativa entre as variedades de SP e CE, e SP e RS $< 0,05$.
	Anova – 1 fator – variável dependente – índices <i>d-prime</i> ; variável independente – duração de estímulo (TL vs TC).	$F(1, 76) = 5,14, p < 0,05$
	Kruskal-Wallis – variável dependente – acertos; variável independente – variedades regionais (SP, CE, RS).	$p > 0, 05$
	Kruskal-Wallis – variável dependente – índices <i>d-prime</i> ; variável independente – variedades regionais (SP, CE, RS).	$p > 0, 05$
Experimentos 1 e 2	Kruskal-Wallis – variável dependente – acertos; variável independente – tipo de experimento (PURR vs CMM).	$p < 0,05$
	Kruskal-Wallis – variável dependente – F0 final; variável independente – variedades regionais (SP, CE, RS).	$p > 0, 05$
Análise de produção	Anova – 1 fator – variável dependente – F0 médio; variável independente – variedades regionais (SP, CE, RS).	$F(2, 51) = 17,09, p < 0,05$
	Teste de Tukey – F0 médio e variedades regionais (SP, CE, RS).	Diferença significativa entre as variedades de SP e CE, e SP e RS, $p < 0, 05$.
	Anova – 1 fator – variável dependente – dispersão de F0; variável independente – variedades regionais (SP, CE, RS).	$F(2, 51) = 18,04, p < 0,05$
	Teste de Tukey – dispersão de F0 e variedades regionais (SP, CE, RS).	Diferença significativa entre as variedades de SP e CE, e entre a SP e RS, $p < 0, 05$.

Quadro 2: Resumo das análises.

Retomando os trabalhos de Atkinsons (1968), Bonte (1975), Maidment (1976), Ohala e Gilbert (1978) e Bezooijen e Gooskens (1999), os ouvintes são capazes de reconhecer uma língua ou variedade linguística somente tendo acesso à informação prosódica. De fato, levando-se em conta os resultados deste trabalho, pode-se afirmar que os informantes (ouvintes) notaram diferenças entre os estímulos a eles apresentados.

A primeira questão deste trabalho, sobre a capacidade dos ouvintes em encontrar suas próprias variedades em meio às demais e de forma atípica (ora com prosódia, ora com curva melódica monotônica), foi respondida de maneira positiva. Assim, os resultados deste estudo sobre o português brasileiro corroboram os estudos anteriores. Mas responder a essa questão sobre o reconhecimento das variedades não traz à luz o que ocasionaria o reconhecimento das variedades regionais. Como exposto por Clopper e Pisoni:

By studying and identifying the acoustic cues used in dialect categorization, we can better determine what kinds of information about dialect variation are encoded, stored, and represented by the naïve listener based on his or her everyday experiences with linguistic variation in the environment. (2004, p. 320)

Para poder buscar quais pistas os informantes têm à sua disposição, a análise de produção necessita buscar diferenças entre os estímulos dados aos informantes. Mas isso por si só não responde à questão sobre em quais características os ouvintes ancoram sua percepção. Os parâmetros estudados na análise de produção devem se comportar semelhantemente às respostas dadas pelos ouvintes. Assim, a manipulação desses parâmetros deveria implicar uma mudança de direção nos julgamentos dos informantes.

Essa manipulação *post hoc* não foi feita, mas pode-se tentar organizar os resultados de percepção e produção para que seja possível visualizar alguma sistematicidade nas escolhas dos ouvintes e nos estímulos produzidos.

O foco desta discussão será o experimento 1, com técnica PURR, por ser comprovadamente a tarefa mais difícil enfrentada pelos informantes, não só neste estudo, como demonstrado pelo teste de Kruskal-Wallis e nos gráficos 21 e 22, mas em Bezooijen e Gooskens (1999). É nessa tarefa que o reconhecimento das variedades regionais foi guiado pela prosódia, mesmo que a análise de produção esteja levando em consideração somente 1 dos parâmetros constituintes da prosódia, a saber, a frequência fundamental (F0).

O primeiro passo é observar como foi o desempenho dos informantes no experimento 1. A ANOVA (acertos e variedades regionais) apresentou diferenças significativas de reconhecimento, especificadas pelo teste de Tukey entre as variedades de São Paulo e do Ceará, e entre as do Rio Grande do Sul e de São Paulo. Essa primeira divisão no desempenho já é um indício de que os informantes de São Paulo comportaram-se de maneira diferente dos demais.

No tocante à análise de produção, a finalização foi descartada, mas o tom médio apresentou um resultado que merece uma análise mais detida. A análise das variedades regionais pautada pelos valores de tom médio parece se organizar de maneira análoga à distribuição do teste perceptivo com PURR. Outra vez, a variedade de São Paulo ficou isolada, diferenciando-se da do Ceará e do Rio Grande do Sul.

De fato, esse comportamento pode indicar alguma pista, algo em que ouvintes leigos poderiam estar se baseando para fazer suas escolhas. Por outro lado, há o problema de o tom médio ser uma tendência central das

variações de F0. Por ser uma medida de tendência central, toda e qualquer variação distante da média é parcialmente ignorada, pois mesmo que não tomadas como medida *per se*, elas determinam o valor médio de F0.

Para analisar as variações de F0 de maneira mais abrangente, não estando preso somente ao tom médio, uma análise de dispersão de F0 pode dar mais uma pista sobre qual o motivo dos acertos significativos em direção ao reconhecimento das variedades.

Os resultados apontaram um comportamento semelhante à análise de tom médio, diferenciando a dispersão de F0 da variedade de São Paulo das demais.

O que parece haver de sistemático, quando os resultados são cruzados, é uma distribuição espelhada entre percepção e produção, isolando a variedade de senadoreense num extremo, a variedade pelotense numa posição central, mas ainda sim um pouco mais próxima da senadoreense, e a variedade paulistana noutro extremo (ver gráficos 25 e 28). A dispersão média de F0 também corrobora essa divisão, tendo a variedade do Rio Grande do Sul com 45Hz de dispersão, em média, seguida das variedades do Ceará e de São Paulo com 43Hz e 32Hz, respectivamente.

Pode-se atribuir para os informantes (falantes) do experimento 1, um comportamento diferente tanto nas medidas de tom médio, como nas de dispersão de F0. Essa característica também se encontra nas respostas dadas pelos informantes (ouvintes) do experimento 1.

Ainda sim, há uma questão a ser respondida. Como dito acima, apesar de as variedades do Ceará e de São Paulo estarem em posições extremas, a variedade do Rio Grande do Sul não figura exatamente numa posição central

entre delas. Há, pois, uma aproximação entre as variedades do Rio Grande do Sul e do Ceará, tanto na percepção dos dados quanto na produção.

O problema da análise com dados brutos é de somente contar acertos e erros dos informantes, mas não qualifica qual tipo de erro seria. Neste caso, a TD consegue analisar mais detalhadamente o desempenho dos participantes. Se os senadorenses são distintos dos paulistanos, mas nem tanto dos pelotenses, como seriam os alarmes falsos deles? Da mesma forma, qual seria a tendência dos alarmes falsos dos demais?

Verificar os alarmes falsos de cada variedade é uma maneira de corroborar a análise acima ou de colocar um ponto de interrogação na influência das características tidas, até então, como prováveis pistas utilizadas na escolha das variedades. A tendência seria de aumento no número de alarmes falsos quando o informante se deparasse com uma variedade mais próxima da sua, como, por exemplo, pelotenses cometendo alarmes falsos com a variedade senadorenses. Abaixo, há o esquema das análises de alarmes falsos:

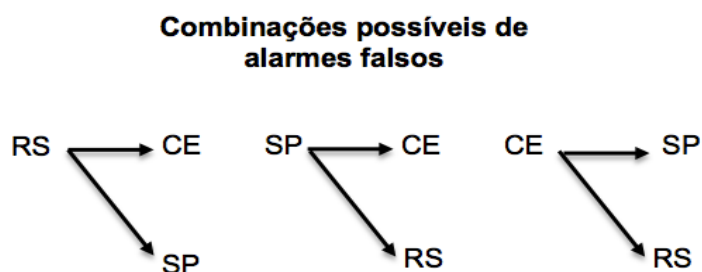


Figura 19: Esquema dos alarmes falsos.

Alarmes falsos – RS			Alarmes falsos – SP			Alarmes falsos – CE		
Estímulos	SP	CE	Estímulos	CE	RS	Estímulos	SP	RS
1	2	5	1	2	5	1	7	5
2	2	7	2	2	6	2	4	5
3	4	4	3	2	3	3	4	6
4	3	6	4	4	8	4	3	7
5	1	4	5	2	4	5	3	3
6	3	5	6	4	8	6	6	7
7	4	6	7	3	2	7	4	3
8	2	5	8	5	6	8	6	9
9	3	7	9	2	5	9	4	3
10	5	5	10	2	8	10	4	5
11	2	4	11	2	4	11	5	8
12	3	7	12	2	6	12	3	4
13	6	6	13	1	5	13	5	4
14	4	9	14	5	6	14	6	3
15	4	3	15	2	8	15	4	6
16	1	4	16	2	3	16	7	4
17	8	2	17	5	4	17	8	4
18	0	2	18	2	8	18	5	10
TOTAL	57	91	TOTAL	49	99	TOTAL	88	96
Porcent.	31%	69%	Porcent.	34%	66%	Porcent.	47,9%	52,1%

Tabela 10: Resultado da contagem de alarmes falsos.

A distribuição de alarmes falsos obedece, em partes, à divisão encontrada nos resultados de percepção e produção, tendo um aumento de ocorrência em direção à variedade que se mostrou mais semelhante na produção e na percepção. Um teste ANOVA para os alarmes falsos dos informantes pelotenses teve como resultado $F(1,34) = 12,98$, $p < 0,05$.

Um teste de Kruskal-Wallis para os alarmes falsos cometidos pelos informantes paulistanos teve como resultado $p < 0,05$. Nesse caso, houve um enviesamento nas respostas, quando incorretas, para uma variedade regional mais próxima. Somente os alarmes falsos cometidos pelos senadorenses não condizem com essa tendência ($p > 0,05$).

Em outras palavras, a hipótese da tendência dos alarmes falsos foi mantida no desempenho das variedades paulistana e pelotense. No caso da variedade senadoreense, parece não haver direção para os equívocos, apesar da ligeira vantagem em direção à variedade mais semelhante, a pelotense, segundo os resultados de produção e percepção.

A relação que se pôde estabelecer entre tom médio, dispersão de F0 e a tendência dos alarmes falsos foi significativa, mas este trabalho tem a ciência de que os resultados obtidos aqui necessitam de um estudo mais detalhado, sobretudo na produção dos estímulos, abarcando os demais parâmetros acústicos da prosódia.

Um outro refinamento possível da análise seria a manipulação das características do parâmetro F0 estudadas neste trabalho e a replicação do experimento 1, ou seja, a manipulação dos tons médios e da dispersão de F0 e compará-los com os resultados deste trabalho. Tal procedimento reafirmaria o papel do tom médio e da dispersão de F0 na percepção das variedades regionais em questão, uma vez que a mudança dessas características deveria, em tese, ocasionar uma mudança na percepção. Com relação ao ritmo, poder-se-ia aplicar um teste com CMM e PURR somados de maneira que somente as durações pudessem ser percebidas.

Para os fins deste trabalho, a prosódia pode ser encarada como parte determinante no reconhecimento das variedades regionais aqui estudadas. A análise de produção conseguiu evidenciar pistas relevantes nas características de F0, as quais podem ser consideradas, nos limites experimentais deste trabalho, como responsáveis pelo reconhecimento das variedades pelos informantes.

8. REFERÊNCIAS

AMARAL, L.; BORGES, P. *Banco de dados sociolinguísticos variáveis por classe social - VarX*. Pelotas: UFPel, 2001.

ABDI, H. Signal Detection Theory. In: SALKIND, Neil (Ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks (CA): Sage, 2007.

ATKINSON, K. Language identification from non-segmental cues. *Working Papers in Phonetics* (UCLA), 10, p. 85-89, 1968.

BARBOSA, P. A. Estrutura rítmica da frase revelada por aspectos de produção e percepção de fala. Manuscrito referente à comunicação do XLIII Seminário do GEL. São Paulo, 25 a 27 de maio, 1995

_____. Prominence and boundary-related acoustic correlations in Brazilian Portuguese read and spontaneous speech. In: SPEECH PROSODY 2008. Campinas, 2008. Disponível em: <http://aune.lpl.univ-aix.fr/~sprosig/sp2008/papers/id060>. Acesso em: 15 abr. 2011.

van BEZOOIJEN, R.; GOOSKENS, C. Identification of language varieties: the contribution of different linguistic levels. *Journal of Language and Social Psychology*, v. 18, p. 31-48, 1999.

BLOOMFIELD, L. *Language*. New York: Holt, 1933.

BOERSMA, P.; WEENINK, D. Praat, a system for doing phonetics by computer, version 5.1.10. *Institute of Phonetic sciences of the University of Amsterdam*, Report, p. 132-182, 1996.

BONTE, R. *Can you identify a language by its prosody?* (Dissertação de Mestrado) – University of California: Berkeley, 1975.

BUSH, C. Some acoustic parameters of speech and their relationships to the perception of dialect difference. *TESOL Quarterly*, v. 1, n. 3, p. 20-30, 1967.

CASTILHO, A. T. Para um programa de pesquisas sobre a história social do português de São Paulo. In: MATTOS E SILVA, R. V. (Org.). *Para a História do Português Brasileiro*, v. 2, Primeiros Estudos. São Paulo: Humanitas / Fapesp, pp. 337-369, 2000.

_____. Nova Gramática do Português Brasileiro. São Paulo: Contexto, 2010.

CCITT. *A method for subjective performance assessment of the quality of speech version output devices*, Draft ITU-T Recommendation P. 85, study group 12 – report R 6, 1989.

CHAMBERS, J. K.; TRUDGILL, P. *Dialectology*. Cambridge University Press, 1980.

CHERRY, E. C. Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, n. 25, p. 975-979, 1953.

CLOPPER, C. G.; PISONI, D. B. Perception of Dialect Variation. In: PISONI, D. B.; Remez, R. E. (Eds.) *The Handbook of perception*. Oxford: Blackwell, p. 314-337, 2004.

CONSONI, F.; FERREIRA NETTO, W. *A sensibilidade de percepção da variação em semitons ascendentes em palavras isoladas*. Texto apresentado no X CONGRESSO NACIONAL DE FONÉTICA E FONOLOGIA E IV CONGRESSO INTERNACIONAL DE FONÉTICA E FONOLOGIA, Niterói RJ, 2008.

CONSONI, F. Aspectos da Percepção da Proeminência Tonal em Português Brasileiro. Tese (Doutorado em Filologia e Língua Portuguesa) – Universidade de São Paulo, São Paulo, 2010.

DESHAIES-LAFONTAINE, D. *A Socio-Phonetic Study of a Québec French*

Community: Trois-Rivières. (Dissertação de Mestrado) - University College London, 1974.

DIEGUES J. M. *Etnias e culturas no Brasil*. Rio de Janeiro: Biblioteca do Exército, 1960.

DUTOIT, T; PAGEL, V; PIERRET, N; BATAILLE, F.; VRECKEN; O. van der. The mbrola project: towards a set of high quality speech synthesizers free of use for non-commercial purposes. In: INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 4th, Philadelphia, 1996. Disponível em: <http://www.asel.udel.edu/icslp/cdrom/vol3/920/a920.pdf>. Acesso em: 24 abr. 2008.

FANT, G.; LILJENCRANTS, J.; LIN, Q. A four-parameter model of glottal flow. *STL-QPSR*, n. 4, v. 26, p.1-13, 1985.

FAURE, G. *Recherches sur les caractères et le rôle des éléments musicaux dans la prononciation anglaise*. Paris: Didier, 1962.

FEAGIN, C. Peaks and Glides in Southern States Short-a. In: GUY, G. R.; Feagin, C.; Schiffrin, D.; Baugh, J. (Eds.) *Towards a Social Science of Language: Papers in Honor of William Labov*, v. 1. *Variation and Change in Language and Society*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, 1996.

FERREIRA NETTO, W. *Variação de frequência e constituição da prosódia da língua portuguesa*. 2006. (Tese de Livre-Docência na Área de Filologia e Língua Portuguesa) - USP, São Paulo, 2006.

_____. *Decomposição da entoação frasal em componentes estruturadoras e semântico-funcionais*. Trabalho apresentado no X CONGRESSO NACIONAL DE FONÉTICA E FONOLOGIA / IV CONGRESSO INTERNACIONAL DE FONÉTICA E FONOLOGIA. Universidade Federal Fluminense, Niterói, RJ,

2008.

_____. ExProsodia. Proc. 8992-2 *Revista da Propriedade Industrial - RPI*, n. 2038, de 26/01/2010. p. 167.

_____; PERES, D. O. A variação tonal em português brasileiro a partir de frases espontâneas e textos lidos. *Estudos Linguísticos*, Campinas, n. 38, v. 1, p. 9-20, 2009.

_____; CONSONI, F.; PERES, D. O. A influência da cadeia segmental na percepção das variações tonais. *LL Journal*, New York, v. 6, n. 1, 2011. Disponível em: <http://ojs.gc.cuny.edu/index.php/lljournal/article/view/652/895>. Acesso em: 20 jun. 2011.

FÓNAGY, I. Des fonctions de l'intonation: Essai de synthèse. *Flambeau*, v. 29, p.1-20, 2003.

FOULKES, P. Sociophonetics. In: BROWN, K. (Ed.) *Encyclopedia of Language and Linguistics*. 2. ed. Amsterdam: Elsevier Press, p. 495-500, 2005.

GREEN, D. M.; SWETS, J. A. *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.

GROSJEAN, F.; FRAUENFELDER, U. H. *A Guide to Spoken Word Recognition Paradigms*. Hove, UK: Psychology Press, 1997.

GUMPERZ, J. The speech community. In: *International Encyclopedia of the Social Sciences*. London: Macmillan, p. 381-386, 1968.

't HART, J.; COLLIER, R.; COHEN, A. *A Perceptual Study of Intonation*. Cambridge: Cambridge University Press, 1990.

HAWKINS, S. Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, v. 31, p. 373-405, 2003.

_____. Puzzles and patterns in 50 years of research on speech perception. Conference Proceedings of From Sound to Sense: 50+ years of Discoveries in Speech Communication. June 11-13, 2004. MIT, p. 223-246, 2004.

HIRST, D.; DI CRISTO, A. (Orgs.) *Intonation Systems*. Cambridge University Press, 1998.

_____; ESPESSER, R. *Automatic Modelling of Fundamental Frequency using a quadratic spline function*. Travaux de l'Institut de Phonétique d'Aix, v. 15, p. 75-85, 1993.

HUCKVALE, M.A.; BROOKES, D.M.; DWORKIN, L.T.; JOHNSON, M.E.; PEARCE, D.J.; WHITAKER, L. *The SPAR Speech Filing System*. In: EUROPEAN CONFERENCE ON SPEECH TECHNOLOGY, 1987, Edinburgh. Disponível em: <http://www.phon.ucl.ac.uk/home/mark/papers/sparsfs87.pdf>. Acesso em: 08 abr. 2010.

HYMES, D. H.; FOUGHT, J. *American Structuralism*. The Hague: Mouton, 1981.

KERSWILL, P.; WRIGHT, S. On the limits of auditory transcription: a sociophonetic perspective. *Language Variation and Change*, v. 2, p. 255-75, 1990.

KLASMEYER, G. The perceptual importance of selected voice quality parameters. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'97). Munich, v. 3, p. 1615-1618, 1997.

KREIMANN, J. Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics*, v. 10, p. 163-175, 1982.

KUHL, P. K.; IVERSON, P. Linguistic experience and the perceptual magnet effect. In: STRANGE, W. (Ed.). *Speech perception and linguistic experience: issues in cross-language research*. Baltimore: York Press, p. 121-154, 1995.

KUHL, P. K. A new view of language acquisition. *Proceedings of the National Academy of Sciences*, n. 97, v. 22, p. 1850-1857, 2000.

KUHL, P.K.; TSAO, F. M.; LIU, H. L.; ZAHNG, Y. e DE BOER, B. *Language/Culture/Mind/Brain. Progress at the margins between disciplines. Annals of New York Academy of Sciences*, n. 935, p. 136-74, 2001. [KUHL, P.K.; TSAO, F. M.; LIU, H. L.; ZAHNG, Y. e DE BOER, B. *Língua, cultura, mente, cérebro: progresso nas fronteiras entre disciplinas. Tradução de Waldemar Ferreira Netto e Fernanda Consoni. São Paulo: Paulistana, 2006.*]

JAKOBSON, R. *Fonema e Fonologia*. Rio de Janeiro: Livraria Acadêmica, 1972.

LABOV, W. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.

_____. *Principles of Linguistic Change*. Philadelphia: University of Pennsylvania Press, 1989.

LADEFOGED, P.; BROADBENT, D. E. Information conveyed by vowels, *Journal of the Acoustical Society of America*, v. 29, p. 98-104, 1957.

LAMBERT, W. E.; HODGSEN, R. C.; GARDNER, R. D.; FILLENBAUM, S. Evaluational Reaction to Spoken Language. *Journal of Abnormal and Social Psychology*, v. 60, p. 44-51, 1960.

LIBERMAN, A. M.; MATTINGLY, I. G. The motor theory of speech perception revised. *Cognition*, v. 21, p. 1-36, 1985.

MACMILLAN, N. A.; CREELMAN, C. D. *Detection Theory: a user's guide*. New Jersey: LEA, 2005.

MAIDMENT, J. Voice fundamental frequency characteristics as language differentiators. *Speech and Hearing Working in Progress*, London, University College London, p. 74-93, 1976.

MARTINET, A. *Elementos de linguística geral*. São Paulo: Martins Fontes, 1970.

MATTOS e SILVA, R. V. Ideias para a história do português brasileiro: fragmentos para uma composição posterior. In: CASTILHO, A. T. (Org.). *Para uma história do português brasileiro. Primeiras idéias*. v. 1. São Paulo: Humanitas. p. 21-52, 1998.

_____. *Para a história do português brasileiro*. São Paulo: Humanitas/Fapesp, v. 2, Primeiros Estudos, 2001.

MCGURK, H.; MACDONALD, J. Hearing lips and seeing voices. *Nature*, v. 264, n. 5588, p. 746-748, 1976.

MOFTAH, A.; ROACH, P. Language Recognition from Distorted Speech: Comparison of Techniques. *Journal of the International Phonetic Association*, v.18, n. 1, p. 50-52, 1988.

MOUGEON, R.; BENIAK, E. Social class and language variation in bilingual speech communities. In: GUY, G. R.; Feagin, C.; Schiffrin, D.; Baugh, J. (Eds.) *Towards a Social Science of Language: Papers in Honor of William Labov*, v. 1. *Variation and Change in Language and Society*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, 1996.

MUFWENE, S. Population Movements and Contacts in Language Evolution. *Journal of Language Contact (THEMA)*, v. 1, p.63-91, 2007.

NATTIEZ, J. J. Modelos linguísticos e análise das estruturas musicais. PER MUSI *Revista Acadêmica de Música*. Belo Horizonte, v. 9, p. 5-46, 2004.

NOLL, V. *O português brasileiro: formação e contrastes*. Tradução de Mário Eduardo Viaro. São Paulo: Globo, 2008.

OHALA, J. J.; GILBERT, J. B. Listener's ability to identify languages from their prosody. *Report of the Phonology Laboratory*, Berkeley, v. 2, p. 126-132, 1978.

OHALA, J. J. There is no interface between phonetics and phonology: a personal view. *Journal of Phonetics*, v. 18, p.153-171, 1990.

PATRICK, P. L. The Speech Community. In: Chambers, J. K.; Trudgill, P.; Schilling-Estes, N. (Eds.) *Language Variation and Change*. Oxford: Blackwell Publishing. p. 573-596, 2004.

PICKETT, J. M.; POLLACK, I. Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech*, v. 6, p. 151-164, 1963.

PIERCE, J R. *The science of musical sounds*. New York: Scientific American Library, 1983.

RIBEIRO, J. *História do Brasil*. Curso Superior. Rio de Janeiro: Editora Francisco Alves, 1914.

RICHARDSON, J. A. C. The Identification by Voice of Speakers Belonging to Two Ethnic Groups. 1973. (Dissertação em Mestrado). Columbus: Ohio State University, 1973.

ROMANO, A.; MOUTINHO, L. C. Alguns esquemas entoacionais característicos de leitura de um texto por portugueses e brasileiros. *Revista de Estudos da Linguagem* (Revista da Faculdade de Letras da Universidade Federal de Minas Gerais), v. 12, n. 2, p. 93-111, 2004.

SAPIR, E. *Linguagem: Introdução ao estudo da fala*. Rio de Janeiro: Acadêmica, 1971.

SONNTAG, G. P.; PORTELE, T. Comparative evaluation of synthetic prosody with the PURR method. In: ICSLP, 1998, Sydney, v. 5, p. 1715-1718, 1998a.

_____. PURR – a method for prosody evaluation and investigation. *Computer Speech and Language, Special Issue on Evaluation*, v. 12, n. 3, p. 437-451, 1998b.

SILVA NETO, S. *Introdução ao estudo da língua portuguesa no Brasil*. Rio de Janeiro: Presença, 1986.

SAUSSURE, F. *Curso de Linguística Geral*. São Paulo: Cultrix, 2003.

STEELE, J. *An essay towards establishing the melody and measure of speech*. Menston, UK: The Scolar Press Limited, 1969 [1775].

STEVENS, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In: DAVID, E.E.; DENES, P.B. (Ed.) *Human Communication: A Unified View*. New York: McGraw-Hill, p. 51-66, 1972.

STRAND, E. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, v. 18, p. 86-99, 1999. Disponível em: <http://www.stanford.edu/~eckert/PDF/strand99.pdf>. Acesso em: 10 de nov. 2010.

SUMBY, W.H.; POLLACK, I. Visual contribution to speech intelligibility in noise, *Journal of the Acoustical Society of America*, v. 26, p. 212-215, 1954.

THOMAS, E. R. Sociophonetics applications of speech perception experiments. *American Speech*, v. 77, n. 2, p. 115-146, 2002.

TRUBETZKOY, N. S. *Principios de fonologia*. Madrid: Cincel, 1970.

VAINIO, M.; SUNI, A.; RAITIO, T.; NURMINEN, J.; JÄRVIKIVI, J.; ALKU, P. New method for delexicalization and its application to prosodic tagging for text-to-speech synthesis. In: 10th ANNUAL CONFERENCE OF THE INTERNATIONAL SPEECH COMMUNICATION ASSOCIATION – INTERSPEECH, p. 1703-1706, 2009.

VAISSIÈRE, J. Sur les universaux de substance prosodiques, In: WAUQUIER, Sophie (Ed.) *Les universaux sonores*. Nantes: Presses Universitaires de Rennes, 2002. Disponível em:

www.personnels.univparis3.fr/users/vaissier/pub/ARTICLES/index_fichiers/2005.pdf. Acesso em: 9 mar. 2007.

_____. Perception of intonation. In: PISONI, D. B.; REMEZ, R. E. (Eds.) *Handbook of Speech Perception*. Oxford: Blackwell, p. 236-263, 2004.

VAISSIÈRE, J.; BOULA de MAREÜIL, P. Identifying a language or an accent: from segments to prosody. In: MODELLING FOR THE IDENTIFICATION OF LANGUAGES (MIDL) WORKSHOP. Paris, p. 1-6, 2004.

VERGÉS, G. M.; CONSONI, F.; CONCEIÇÃO, G. M.; CUNHA, A. F. R. ROSA, R. C. M.; PERES, D. O.; FERREIRA NETTO, W. *Relações entre variação de gênero e variação tonal na fala de língua portuguesa do Brasil*. Trabalho apresentado no X CONGRESSO NACIONAL DE FONÉTICA E FONOLOGIA / IV CONGRESSO INTERNACIONAL DE FONÉTICA E FONOLOGIA. Universidade Federal Fluminense, Niterói, RJ, 2008.

WERTHEIMER, M. Laws of organization in perceptual forms. In: ELLIS, W. (Ed.). *A Source Book of Gestalt Psychology*. London: Routledge & Kegan Paul, p. 71-88, 1938. [1923]. Disponível em <http://psychclassics.yorku.ca/>. Acesso em: 20 abr. 2010.

WILLIS, C. Perception of Vowel Phonemes in Fort Erie, Ontario, Canada, and Buffalo, New York: An Application of Synthetic Vowel Categorization Tests to Dialectology. *Journal of Speech and Hearing Research*, v. 15, p. 246-55, 1972.

XU, Y.; WANG, Q. E. *Components of intonation: what are linguistic, what are mechanical/physiological?* Presented at International Conference on Voice Physiology and Biomechanics. Evaston, 1997.

9. ANEXOS

- Versão escrita dos experimentos:

EXPERIMENTO 1 - PURR

Este teste tem como objetivo saber se as pessoas conseguem reconhecer uma variedade linguística (o seu sotaque) baseado somente na entoação. Se tudo for removido da fala, exceto a ‘música da fala’, é possível identificar a variedade linguística apresentada?

Como exemplo, ouça as seguintes frases:

- *Você sabe do que eu estou falando.*
- *Você sabe do que eu estou falando?*

Nessas frases idênticas somente o padrão entoacional foi mudado.

Este teste é formado por trechos de fala gravados de conversa em português, tendo como variedades o português paulistano, nordestino e gaúcho. Estes trechos foram manipulados eletronicamente para que restasse somente a entoação (música da fala). A representação da entoação (música da fala) soa como um zumbido. Será pedido a você que escute este zumbido e responda se o trecho ouvido corresponde à sua variedade.

As frases serão divididas em curtas e longas. Cada uma delas será seguida por um aviso sonoro como este [bip], então você terá 5 segundos de pausa para marcar com um **X** em **SIM**, caso você reconheça o sotaque da sua região, ou **NÃO**, caso você não reconheça o seu “jeito” de falar.

Para ajudá-lo(a) a ouvir os zumbidos melhor, será feito um pequeno treinamento:

- Ouça um trecho retirado da variedade paulistana (Exemplo tocado).
- Ouça o zumbido que representa essa frase (Exemplo tocado).
- Ouça a frase seguida pelo seu zumbido (Exemplo tocado)

- Ouça um trecho retirado da variedade gaúcha (Exemplo tocado).
- Ouça o zumbido que representa essa frase (Exemplo tocado).
- Ouça a frase seguida pelo seu zumbido (Exemplo tocado)

- Ouça um trecho retirado da variedade cearense (Exemplo tocado).
- Ouça o zumbido que representa essa frase (Exemplo tocado).
- Ouça a frase seguida pelo seu zumbido (Exemplo tocado)

O teste terá início após o sinal [bip].

*** Obrigado ***

EXPERIMENTO 2 – CURVA MELÓDICA MONOTÔNICA

Este teste será bem mais breve que o anterior, pois não requer uma fase de treinamento. Trata-se de um experimento inverso ao que foi feito anteriormente, ou seja, serão dadas as frases sem a entoação.

Escute uma frase sem alteração:

(Exemplo tocado)

Agora, escute a mesma frase alterada:

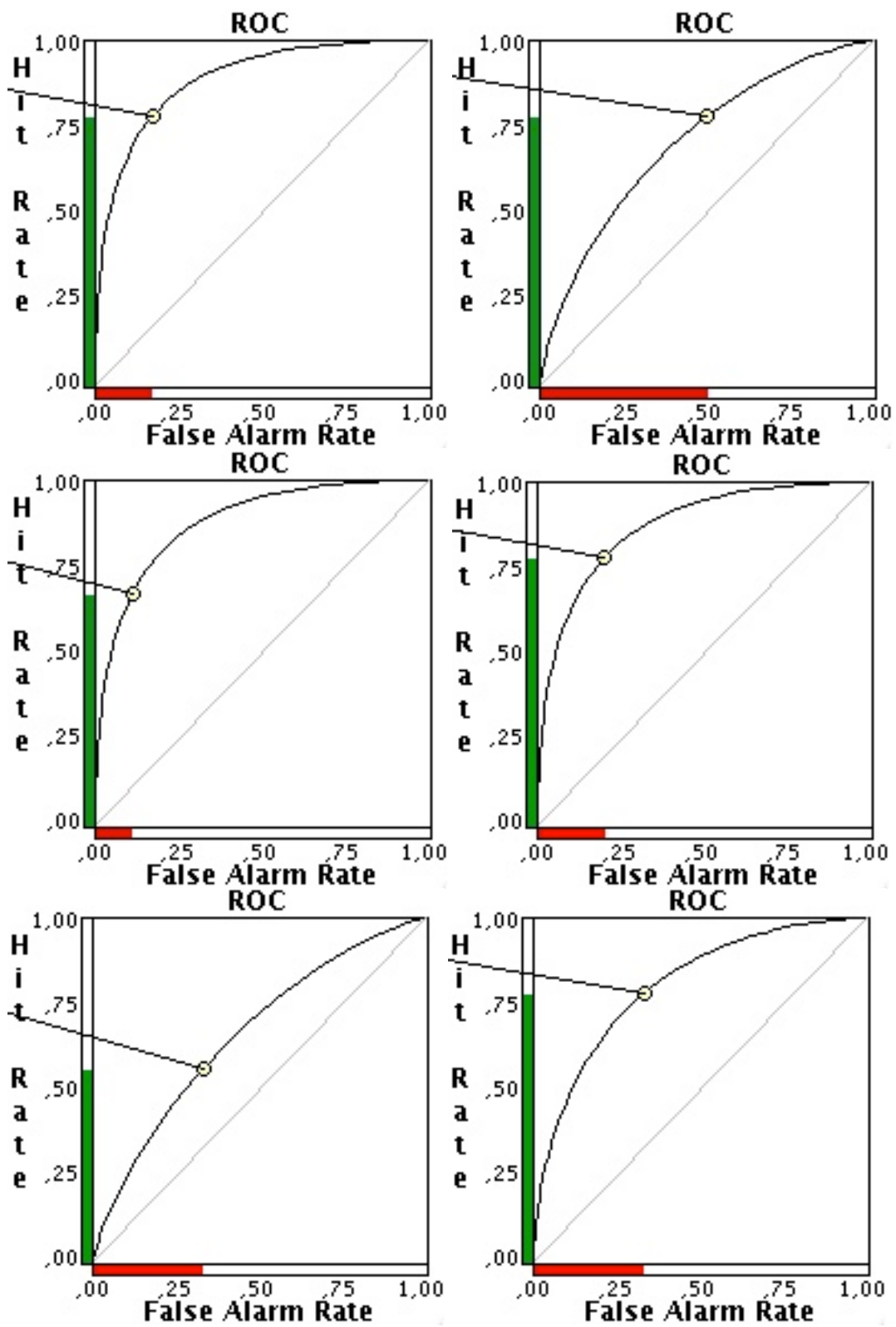
(Exemplo tocado)

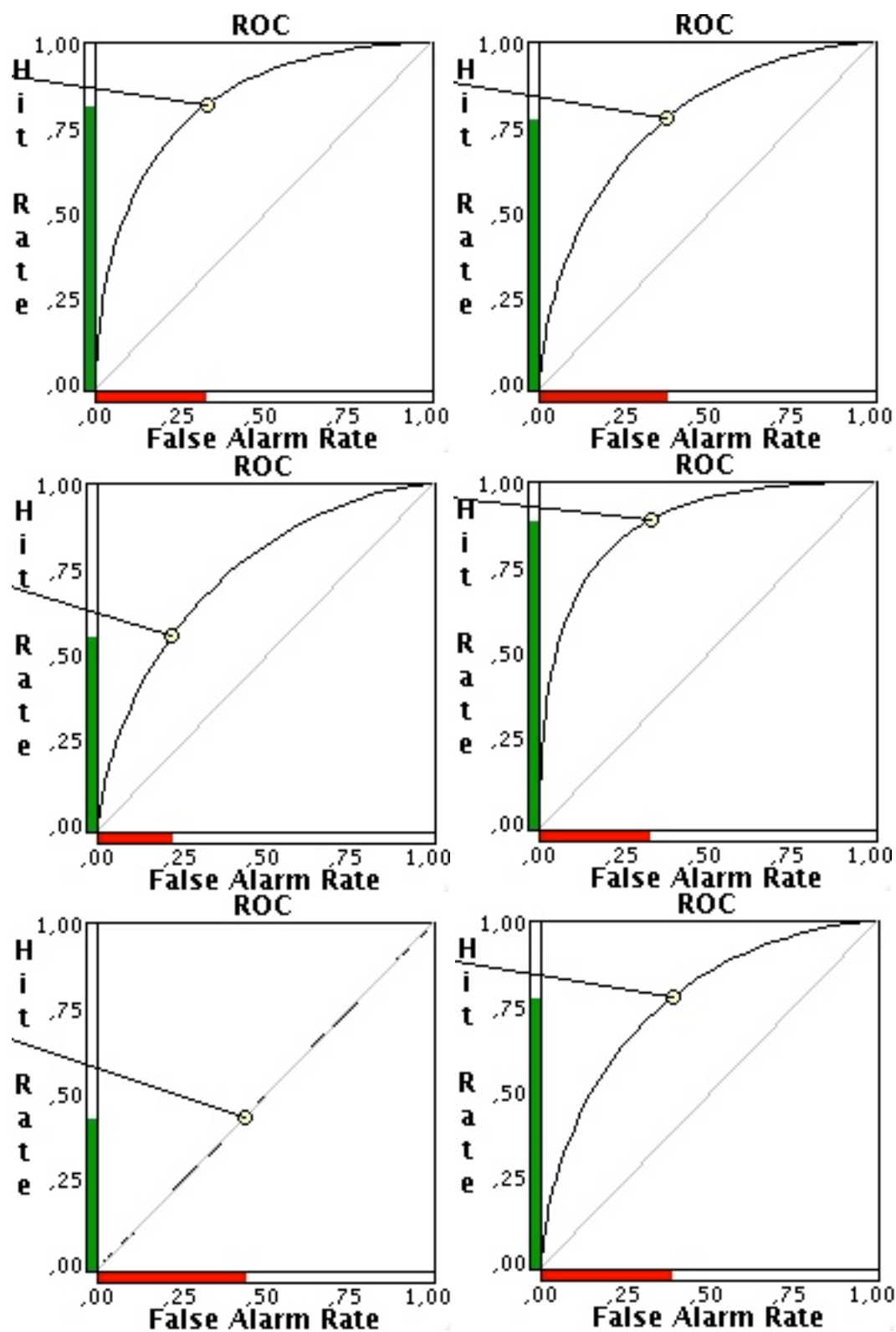
Cada frase será separada das demais por este aviso sonoro: [bip]. Após esse sinal, você terá 6 segundos para marcar se faz parte da sua variedade (marcando SIM) ou se o trecho não é da sua variedade (marcando não).

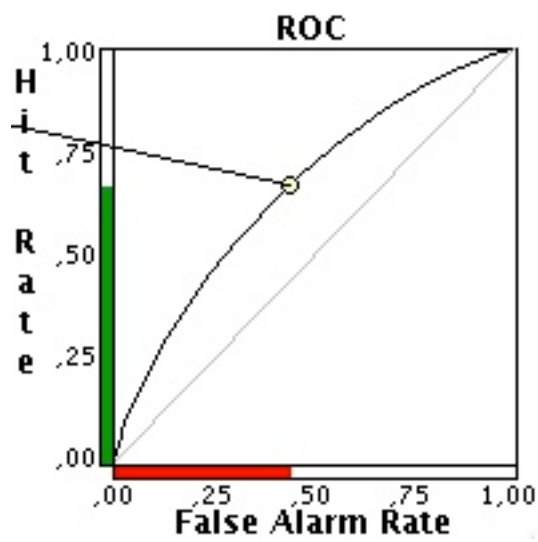
O teste terá início após o sinal [bip].

*** Obrigado ***

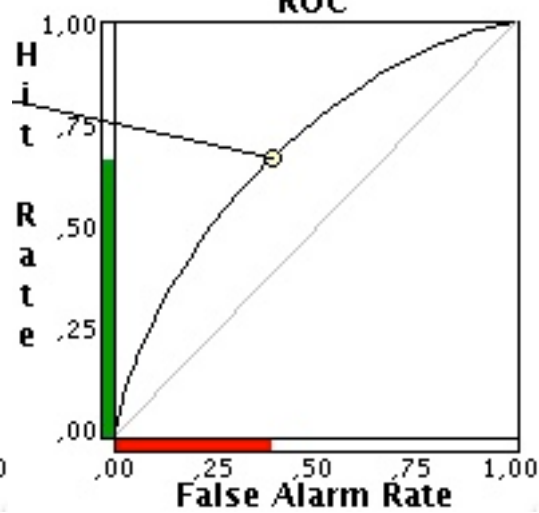
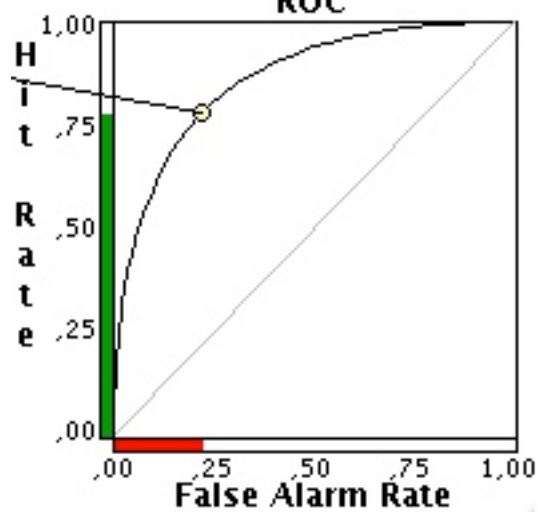
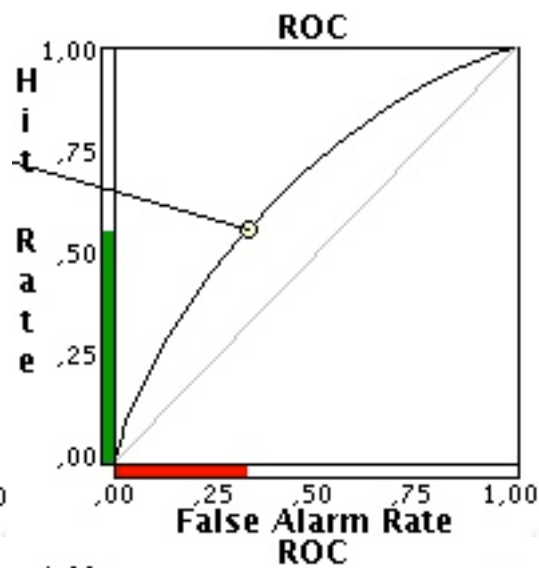
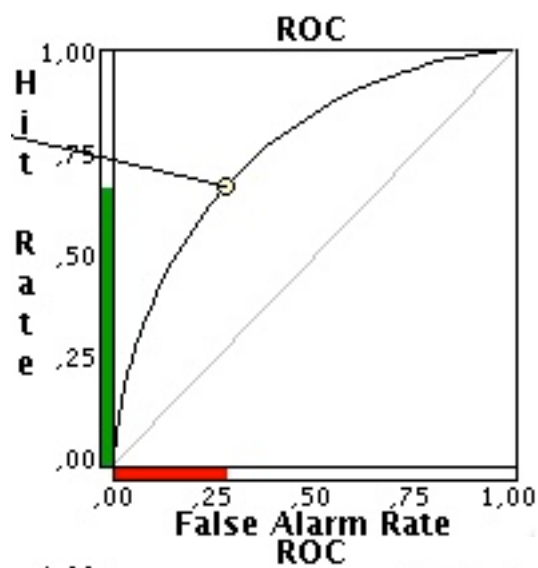
Gráficos ROC – CE – PURR TC

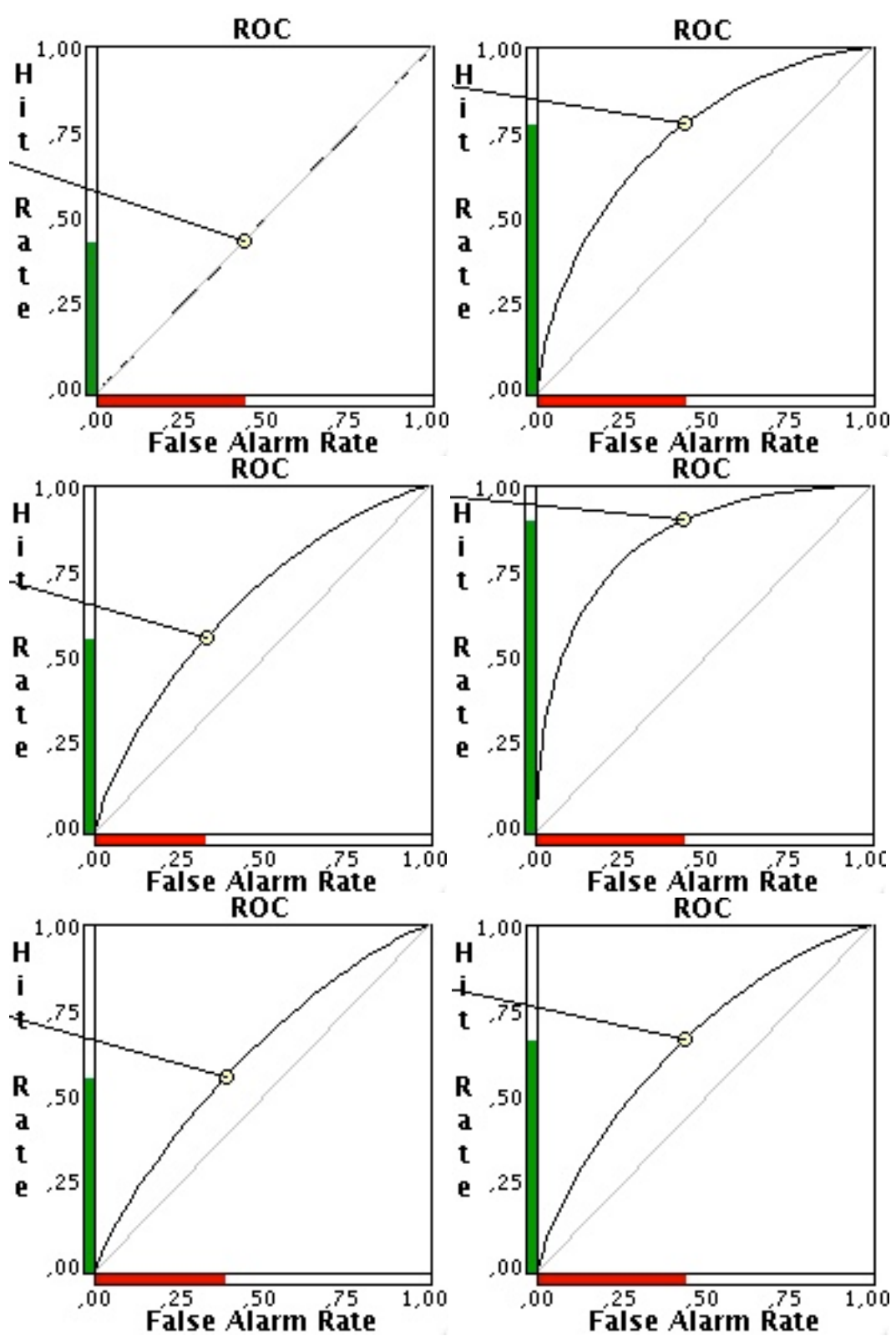


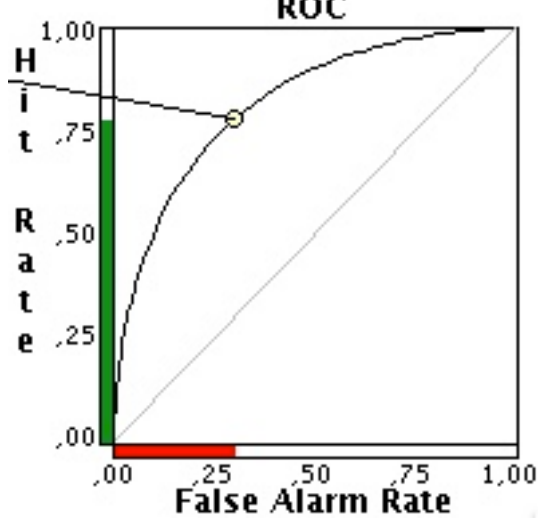
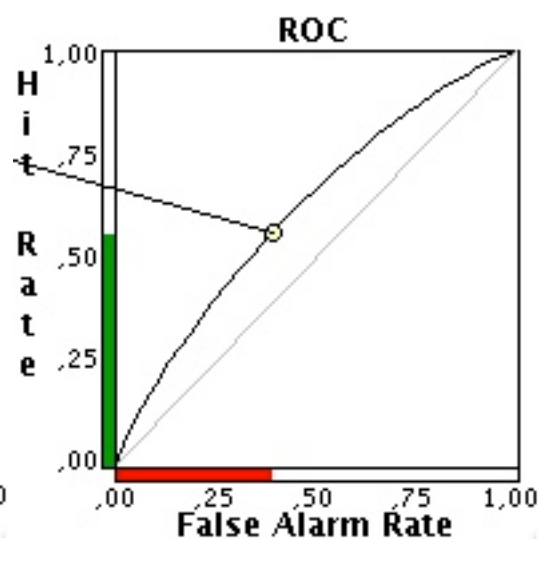
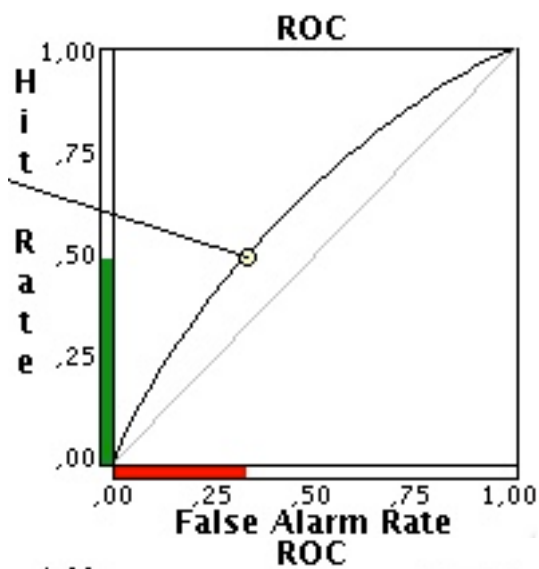




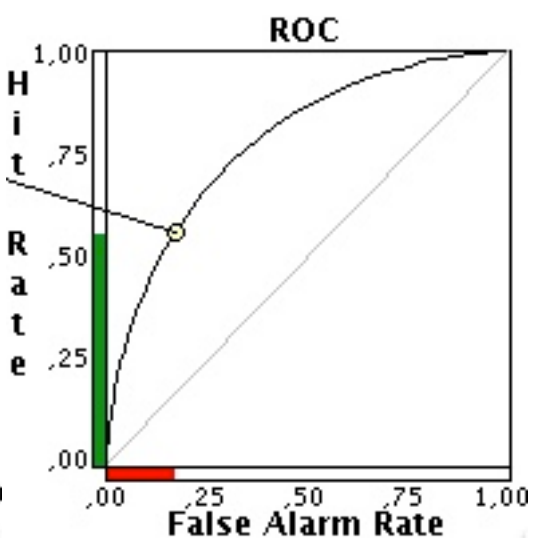
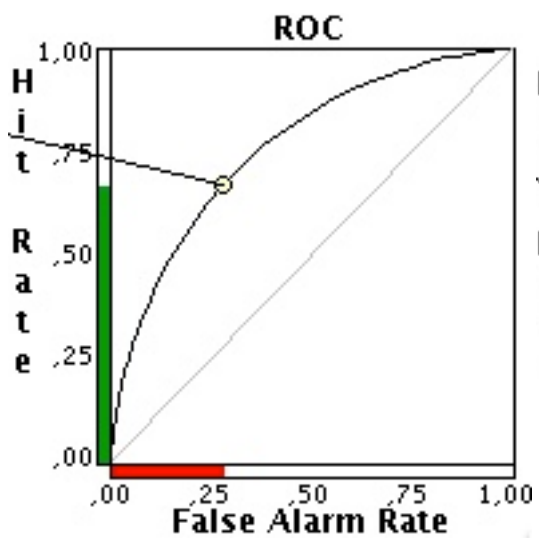
Gráficos ROC – CE – PURR TL

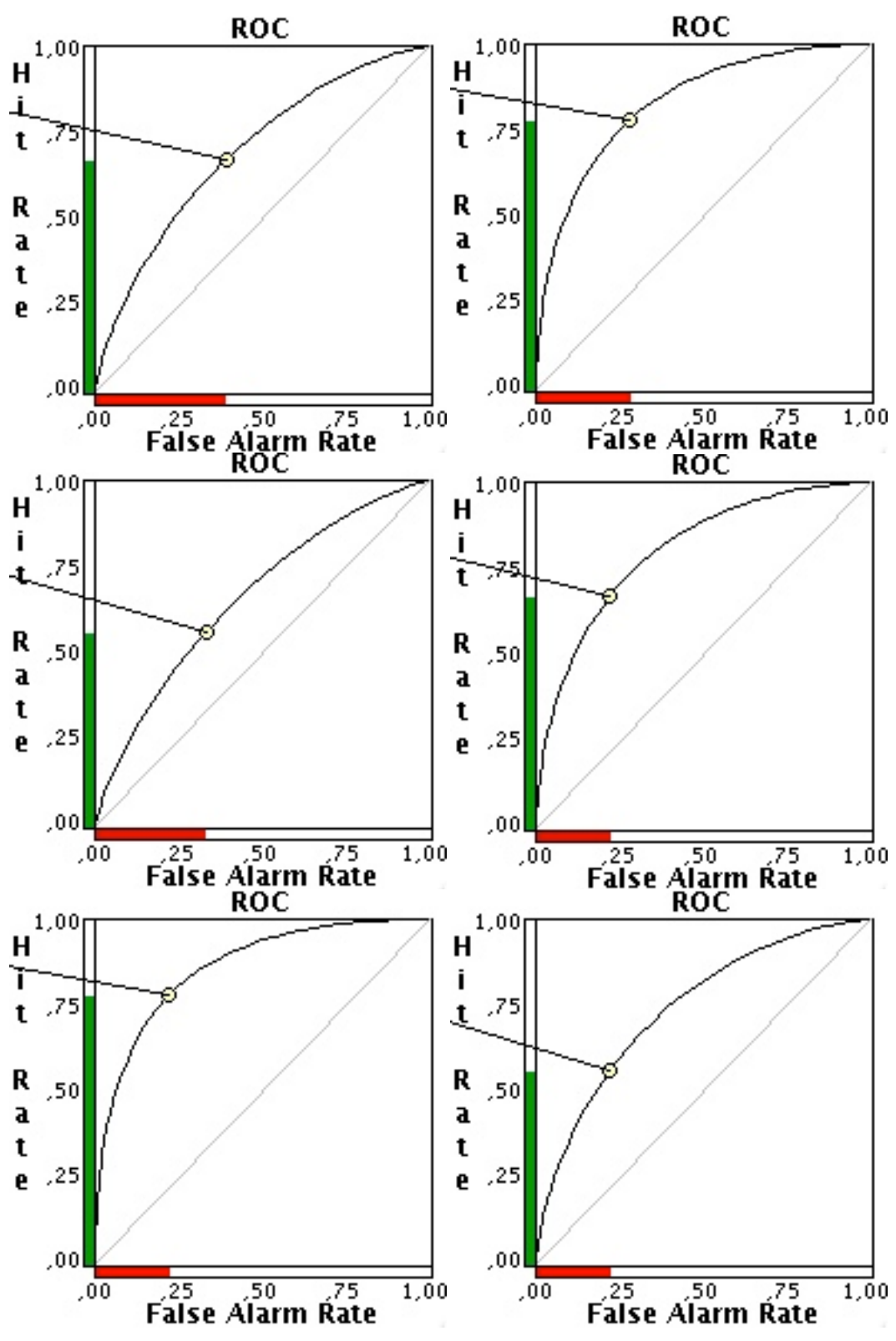


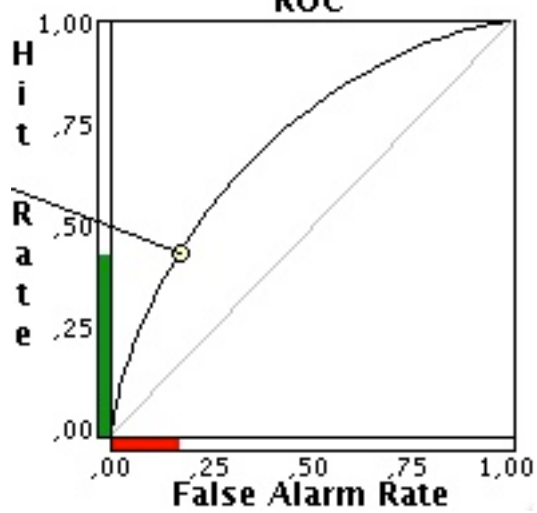
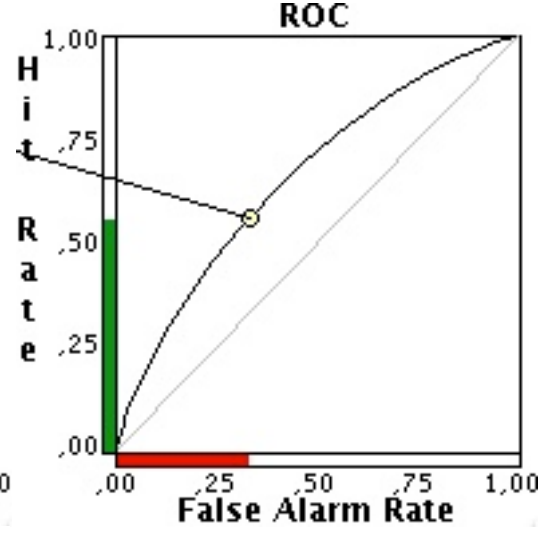
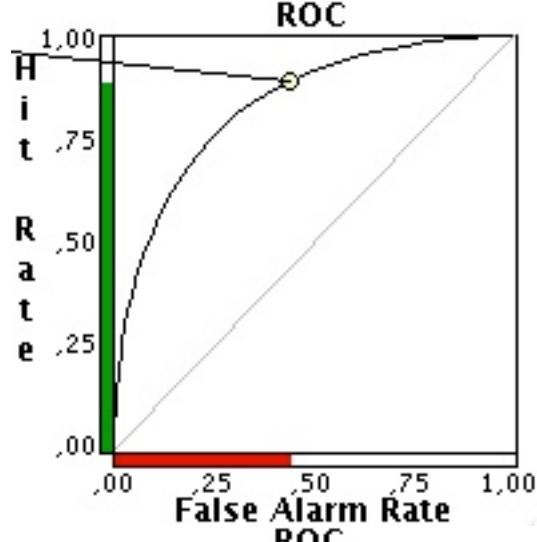
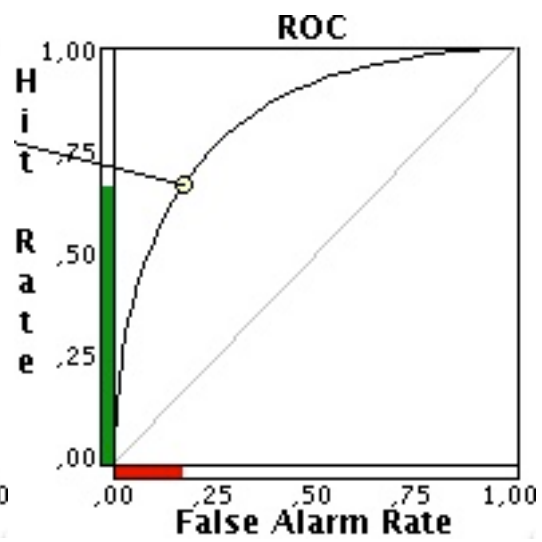
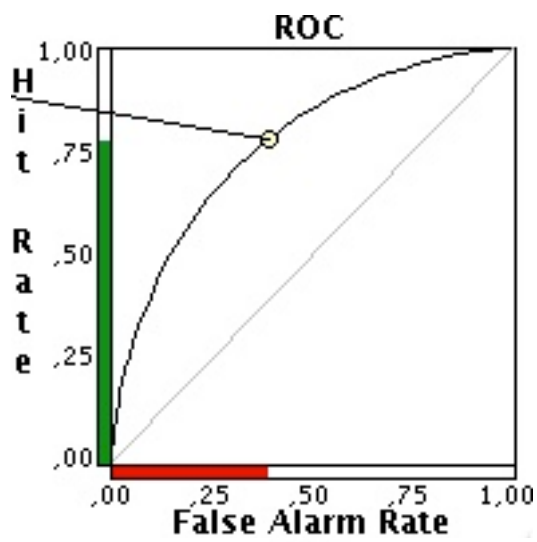




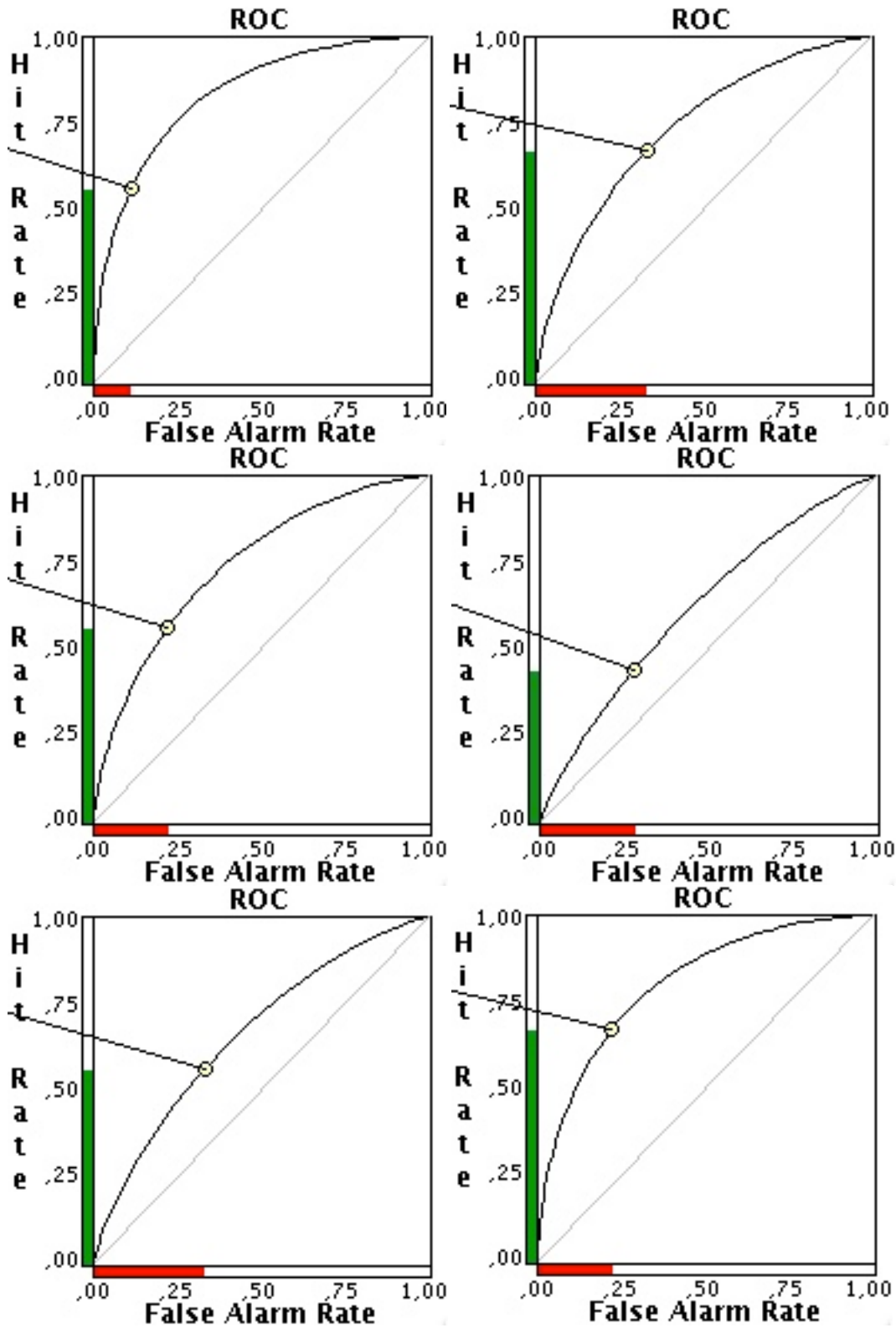
Gráficos ROC – SP – PURR TC

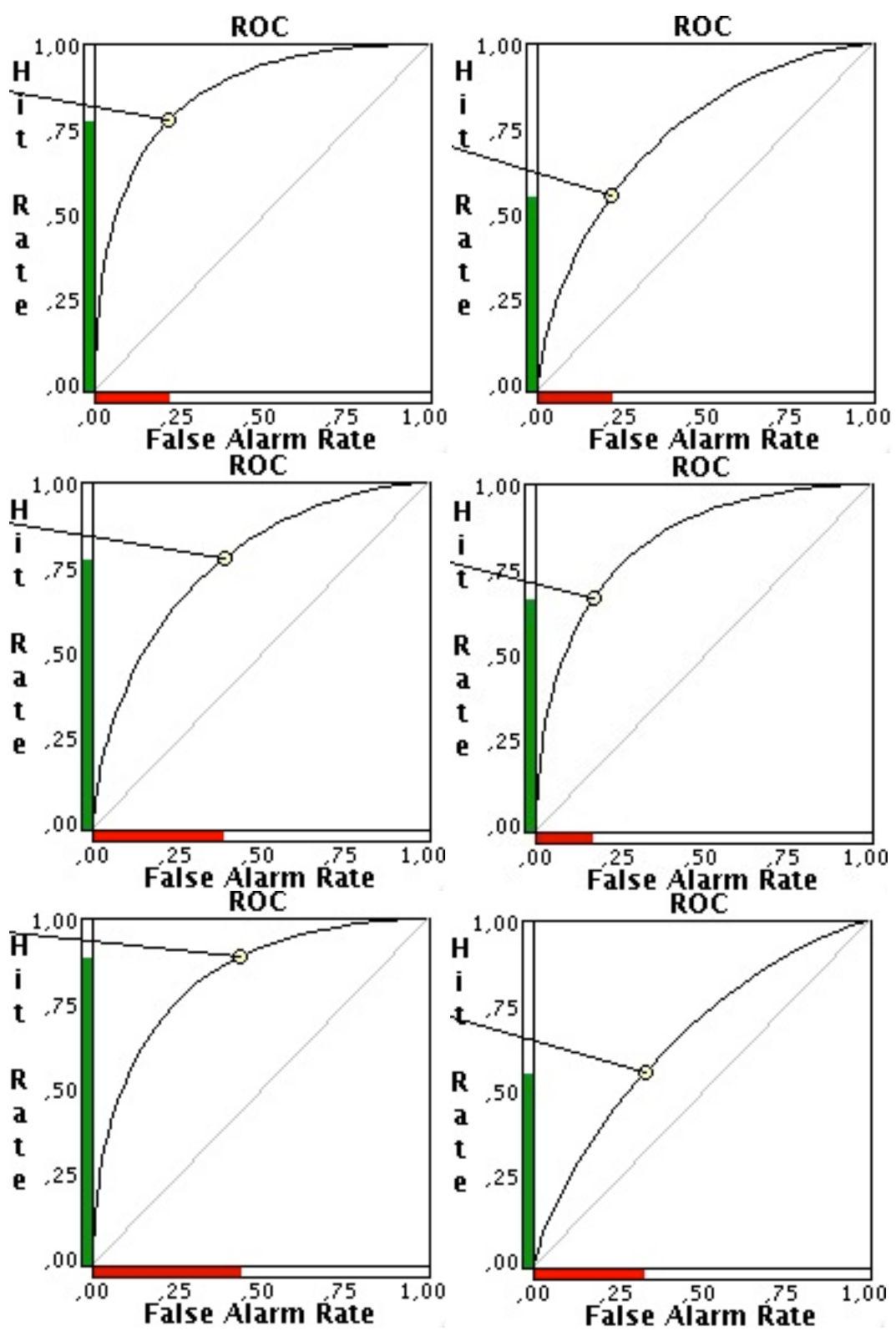


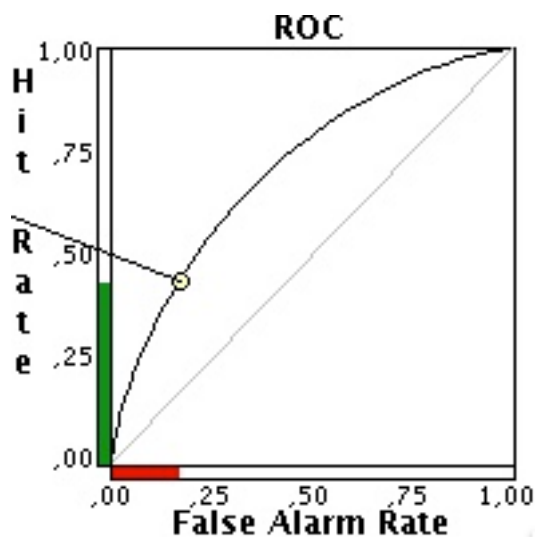




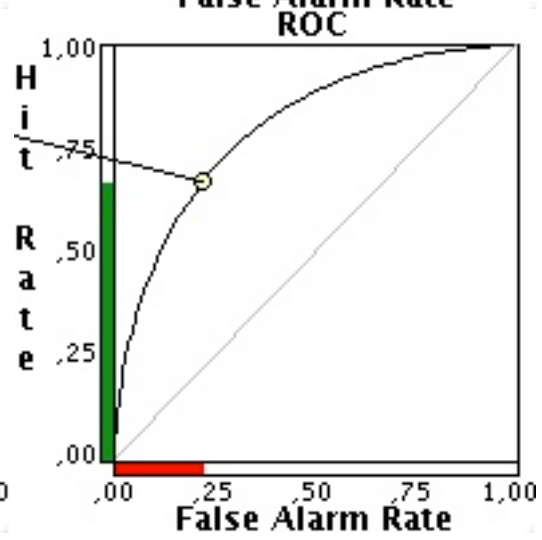
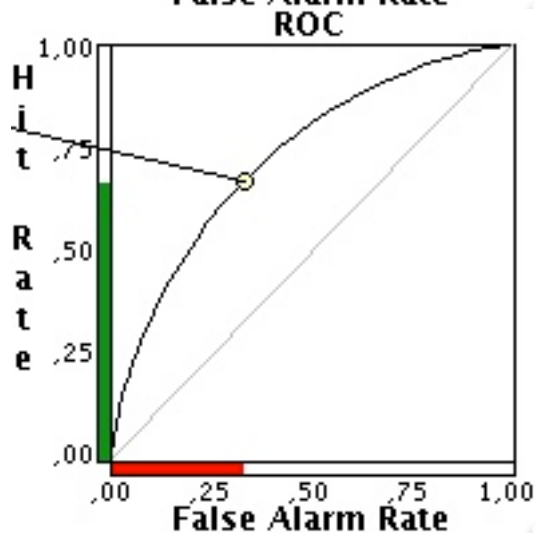
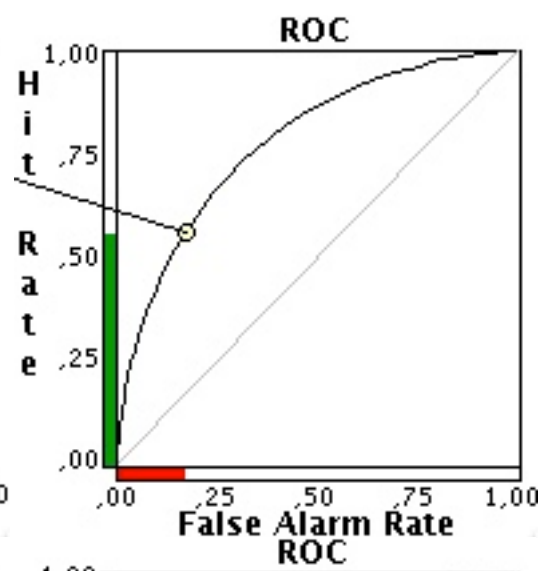
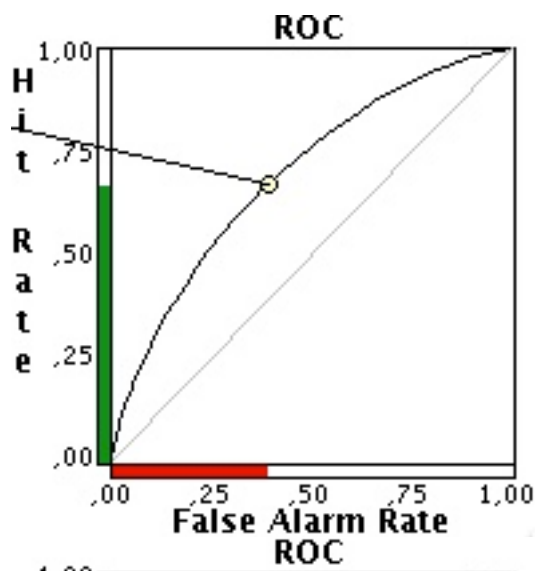
Gráficos ROC – SP – PURR TL

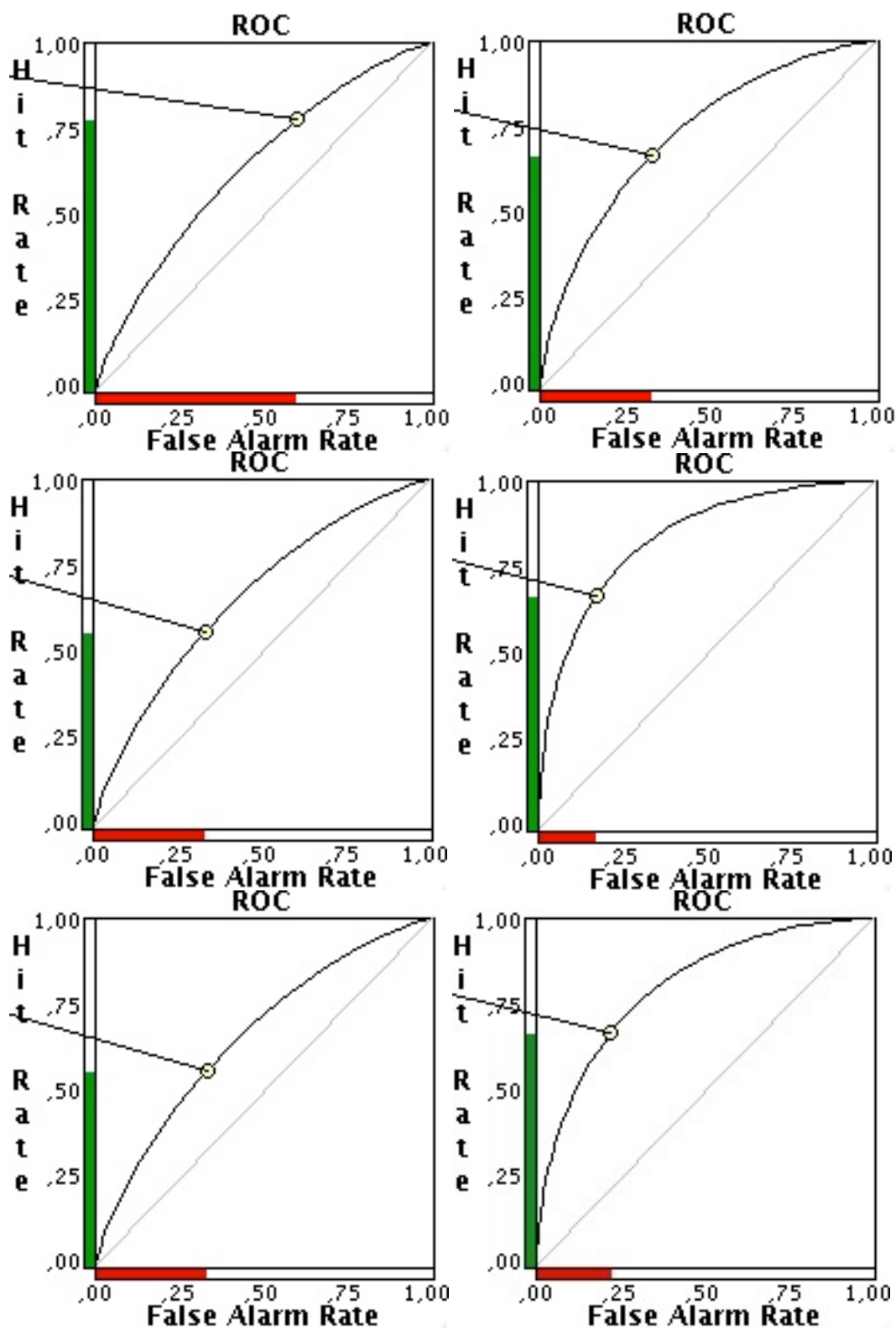


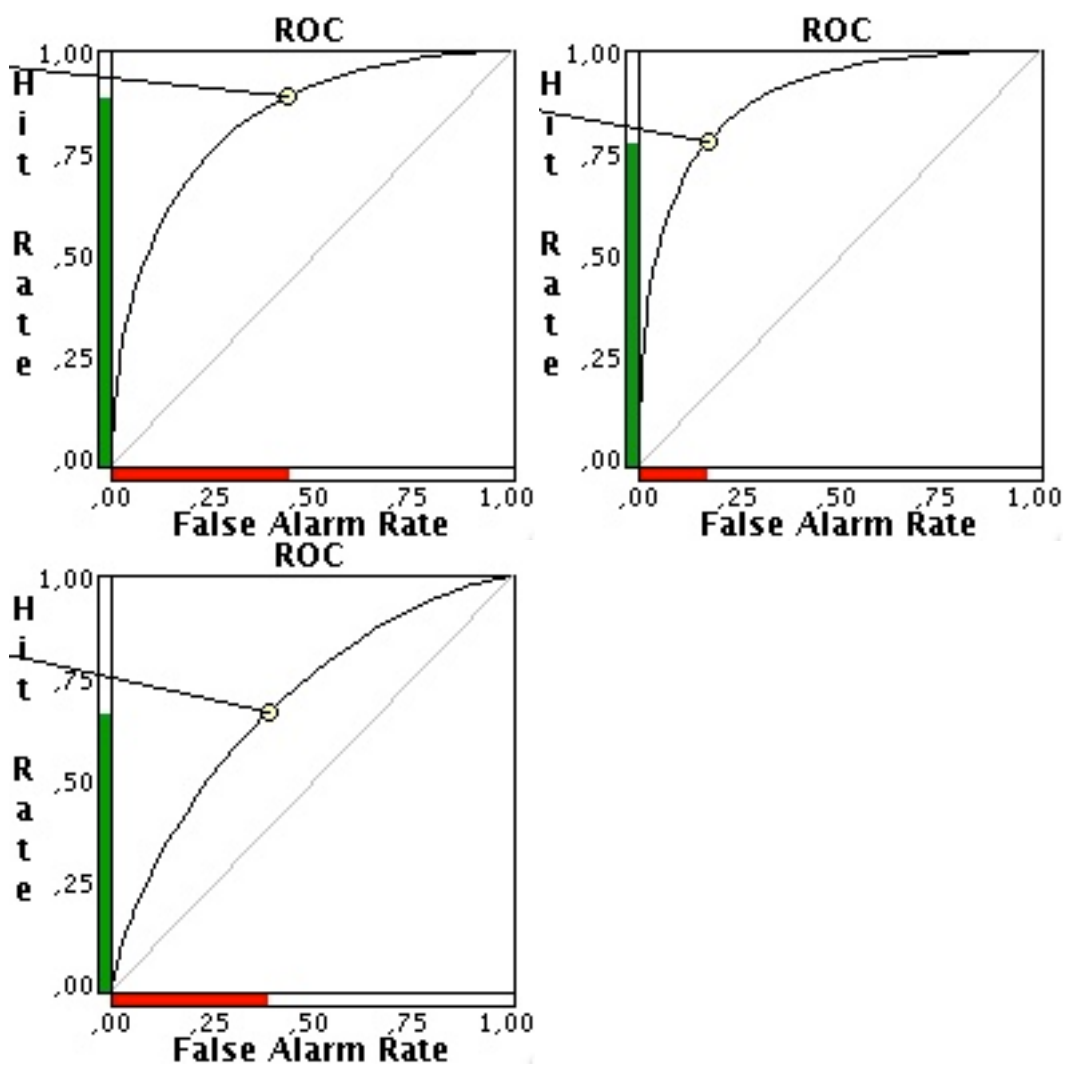




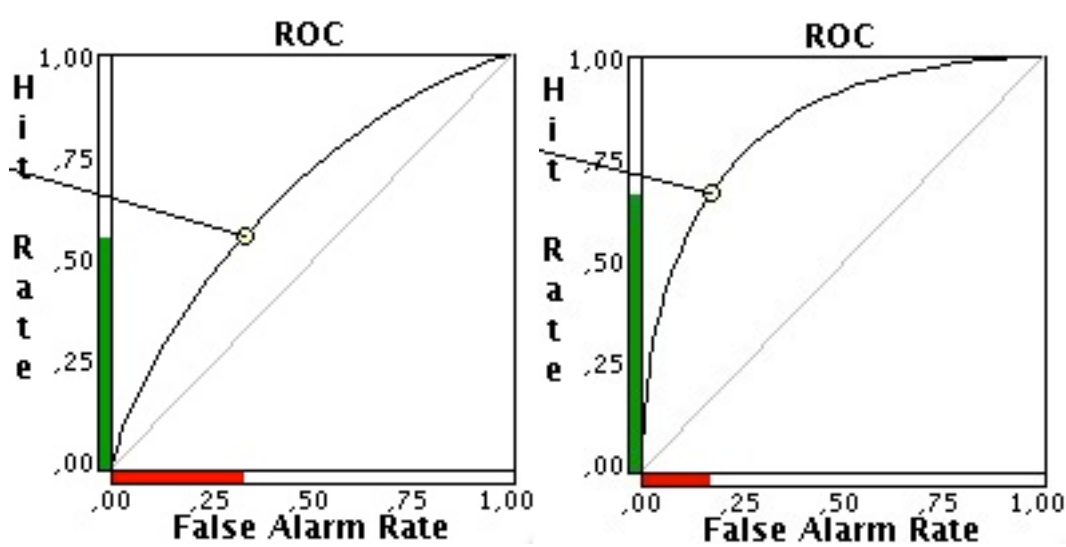
Gráficos ROC – RS – PURR TC

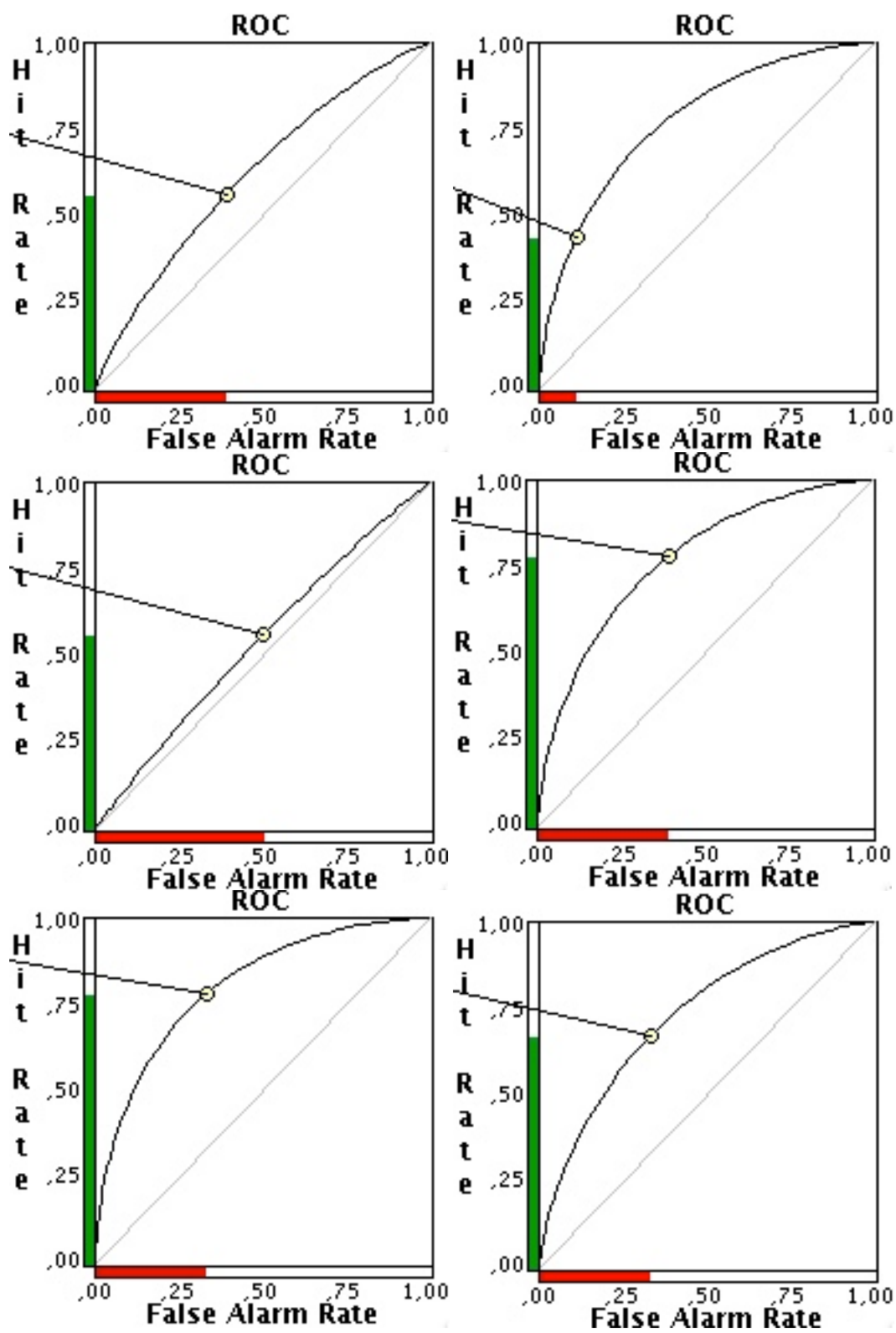


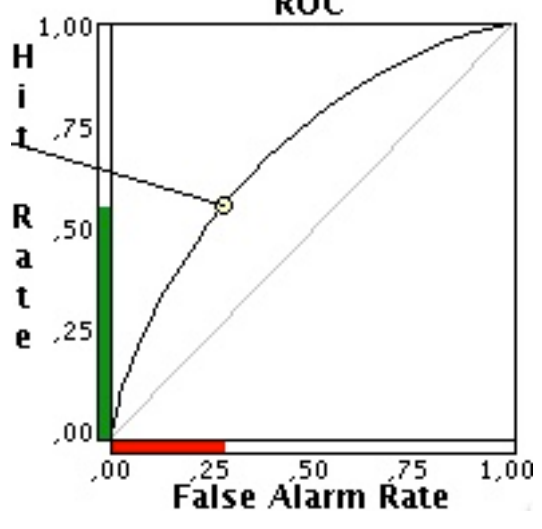
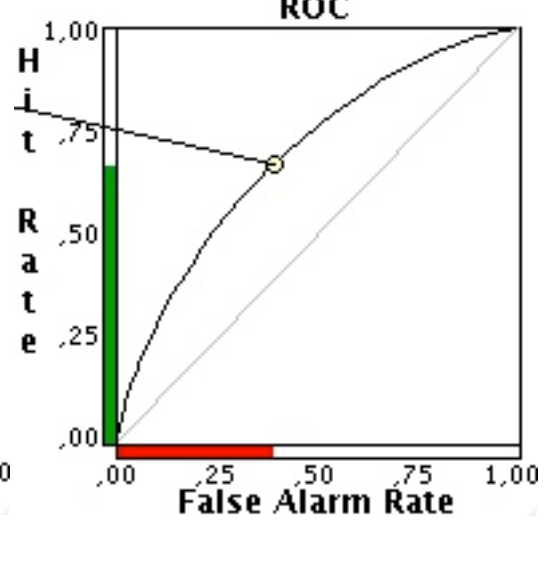
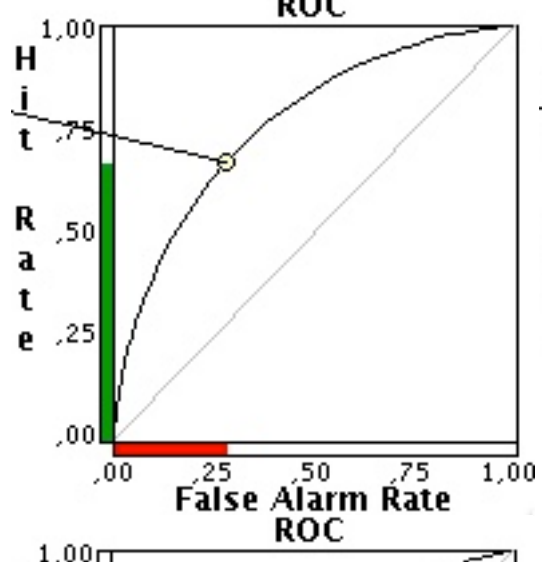
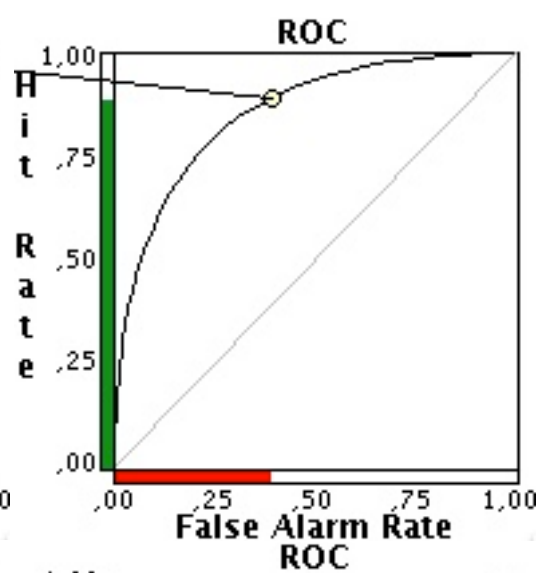
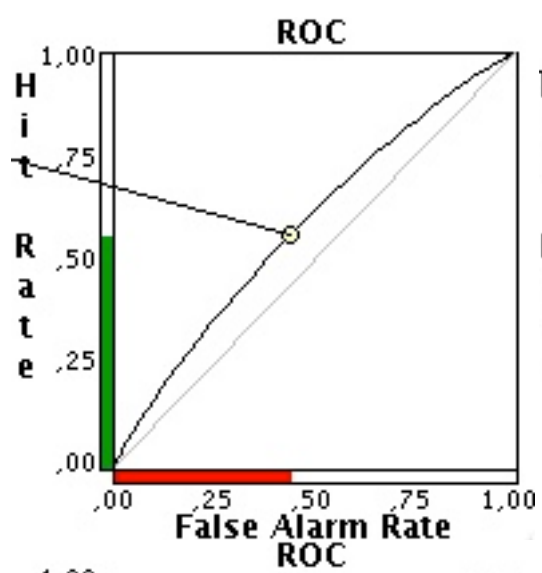




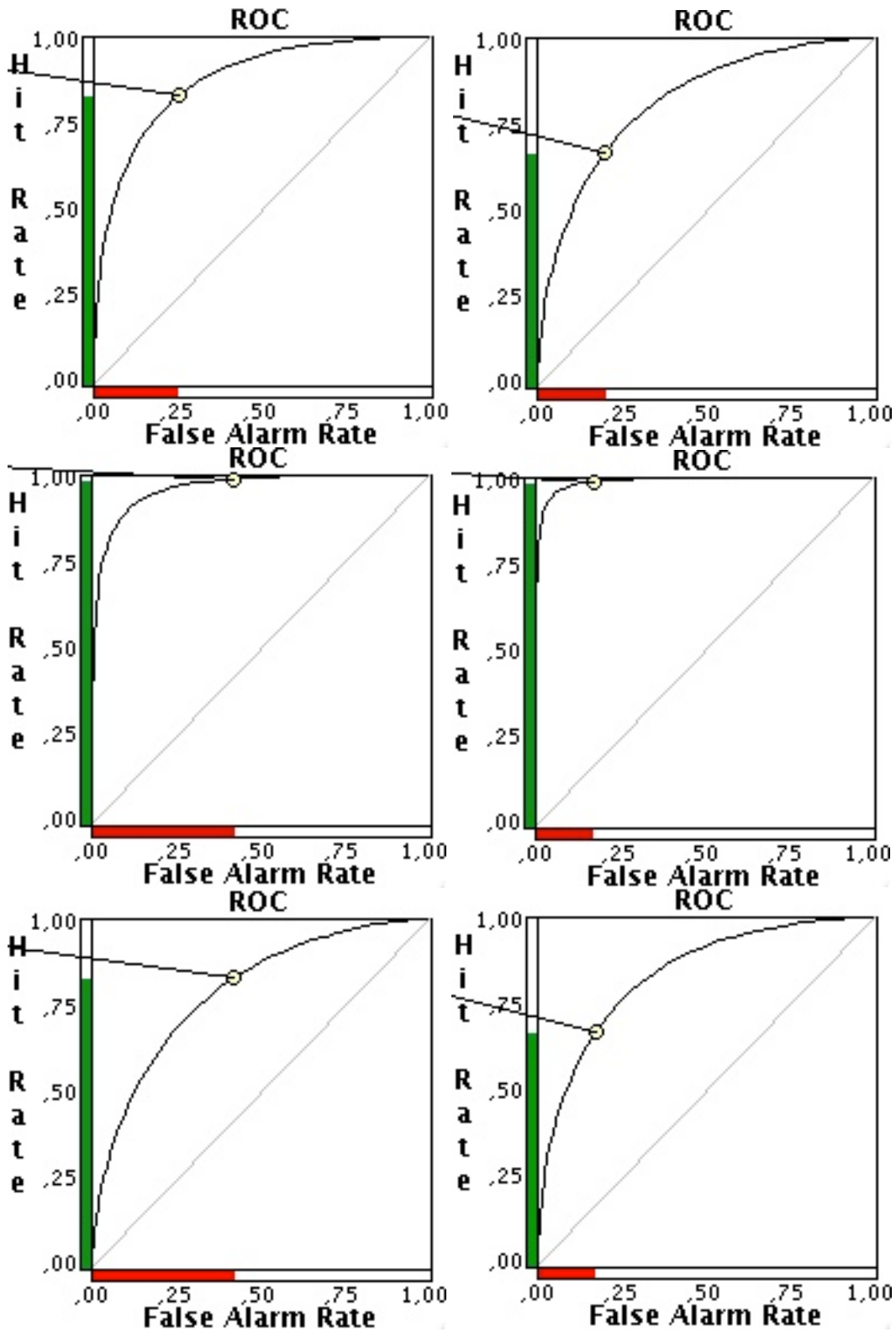
Gráficos ROC – RS – PURR TL

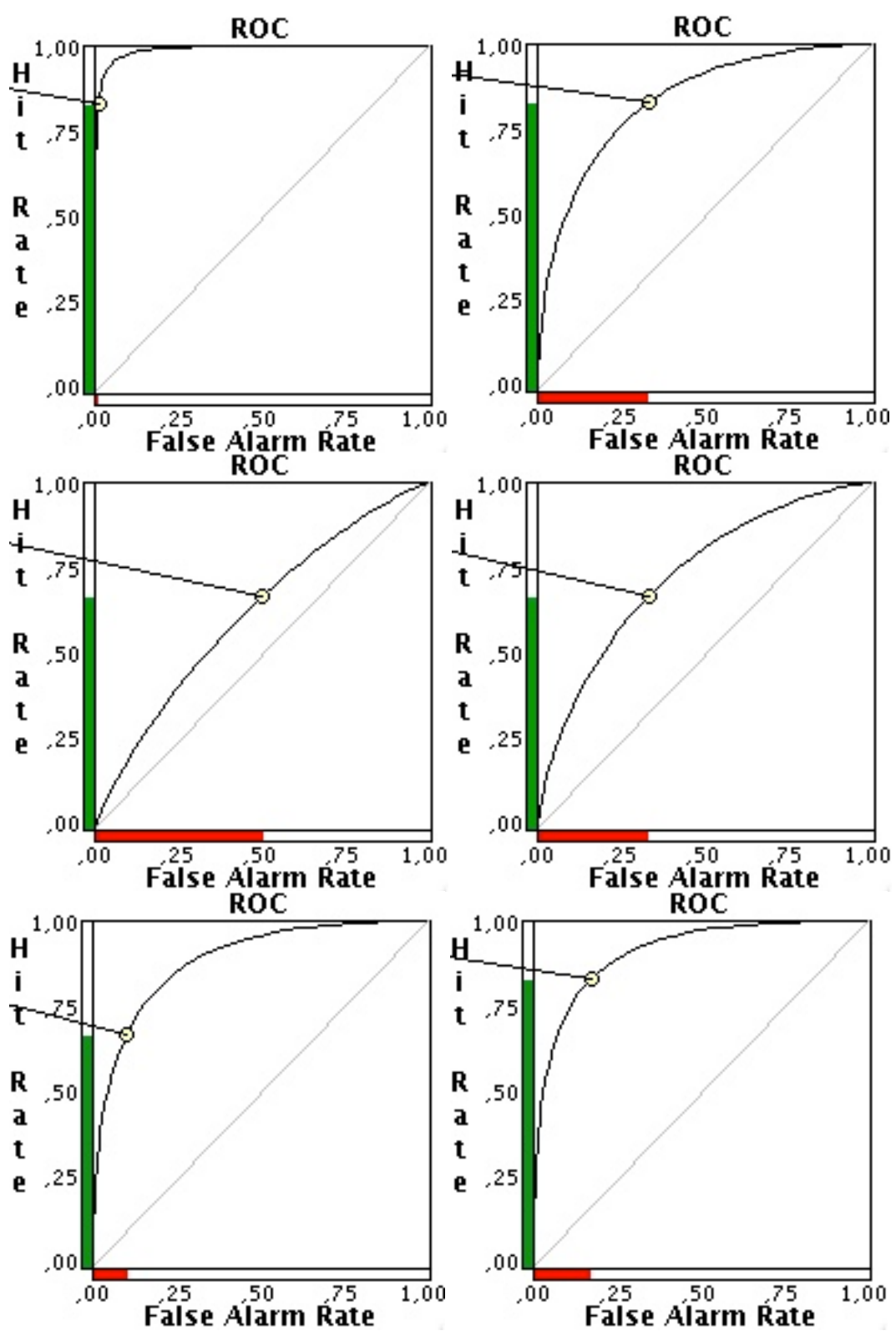


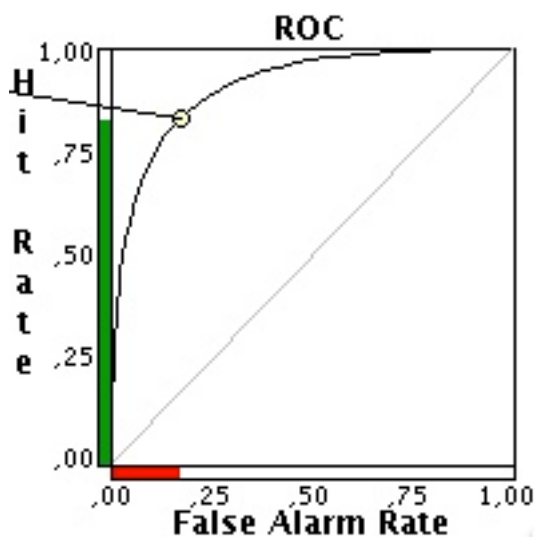




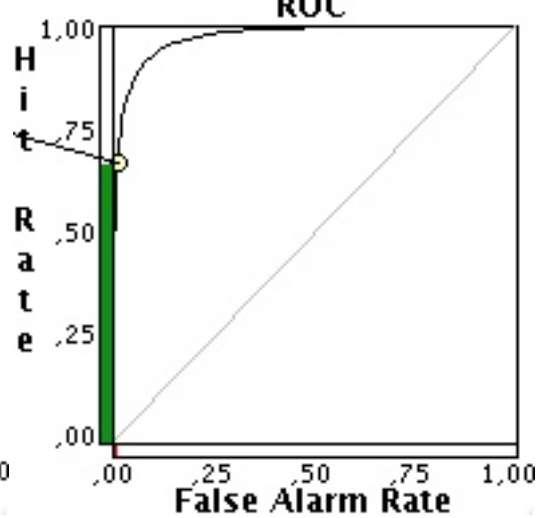
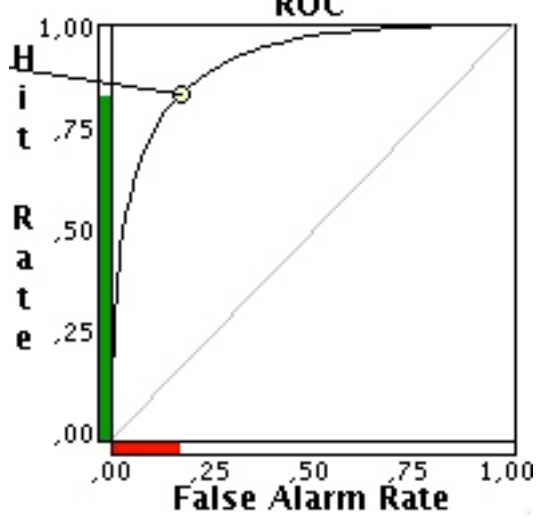
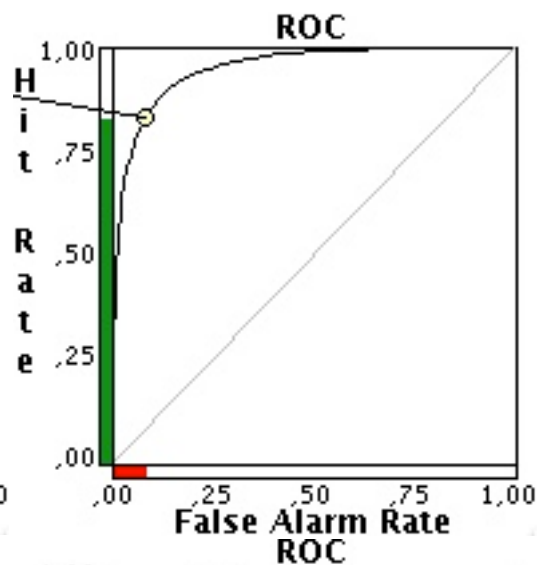
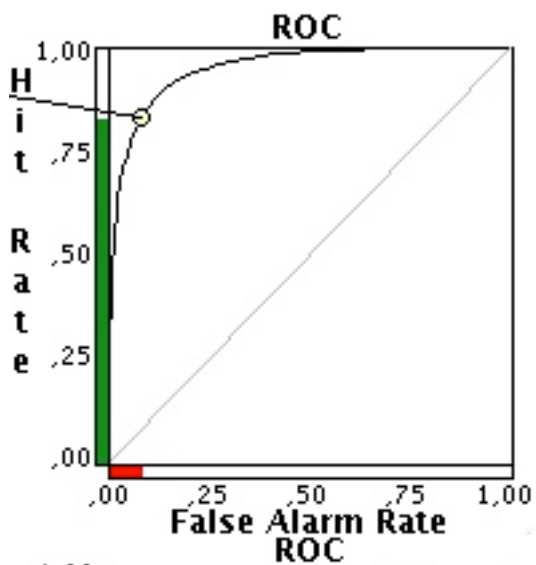
Gráficos ROC – CE – CMM

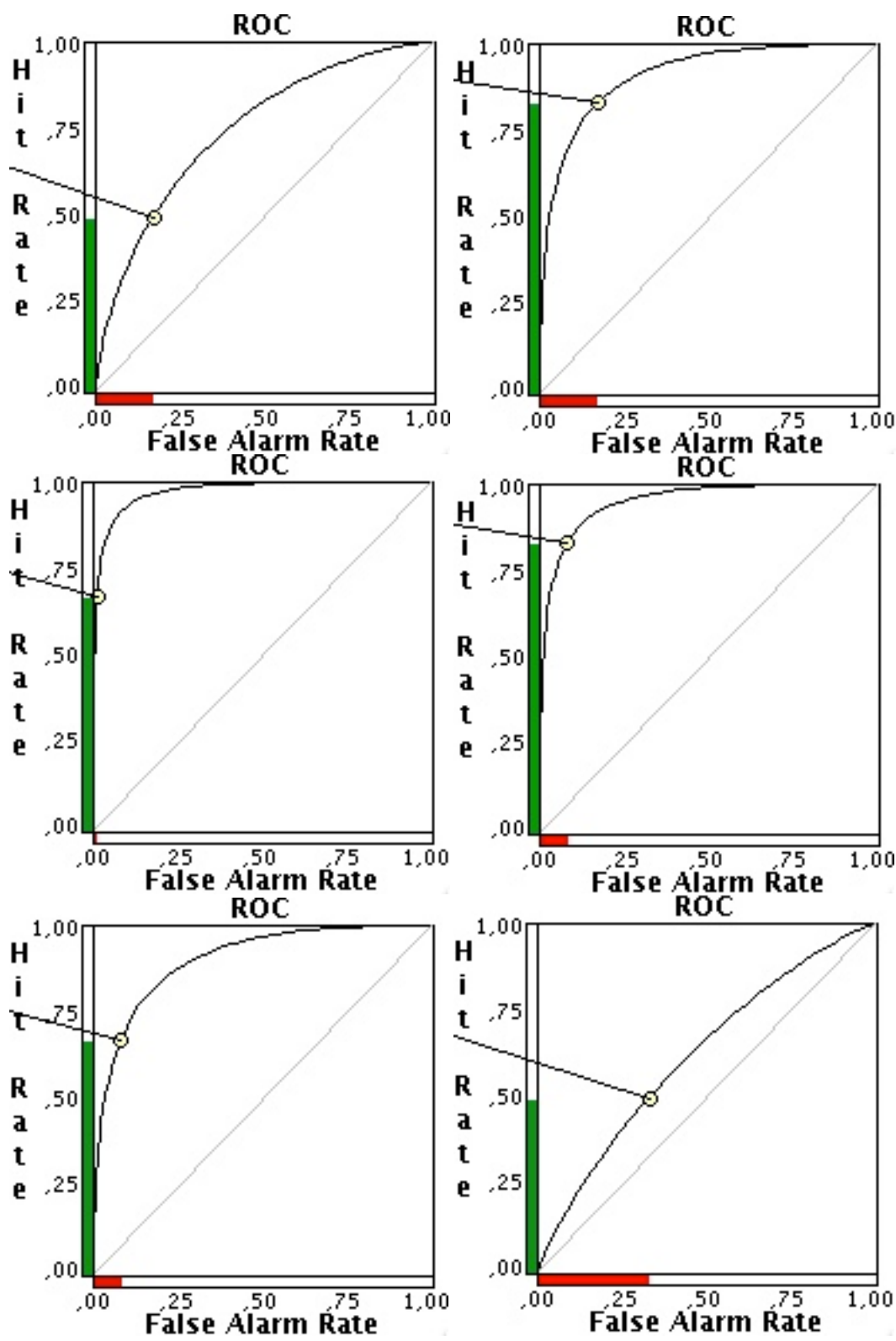


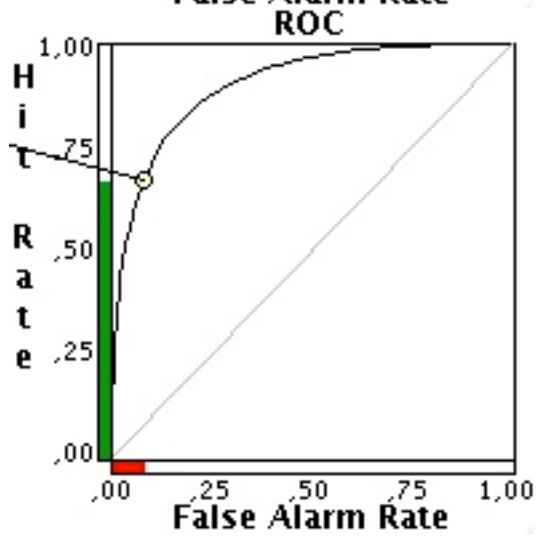
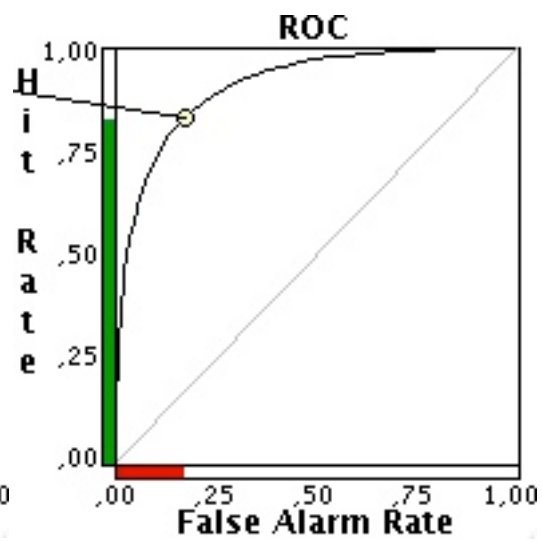
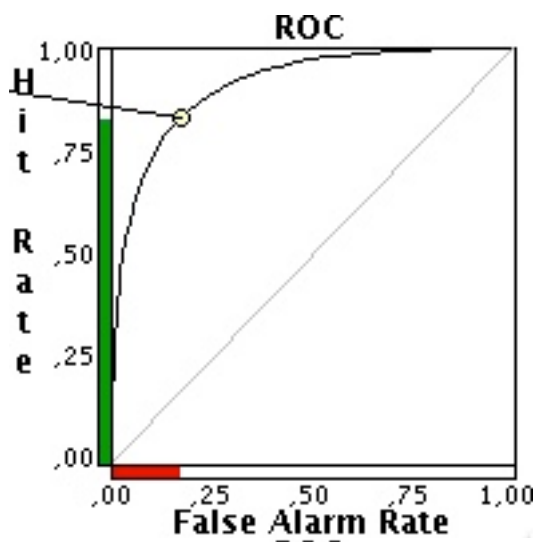




Gráficos ROC – SP – CMM







Gráficos ROC – RS – CMM

