Capítulo 3

Metodologia Experimental

Nas próximas seções apresentamos os elementos principais da metodologia experimental aplicada à área de prosódia. Alguns experimentos serão descritos em detalhe, precedidos de uma apresentação sucinta das teorias e observações que os motivaram para que o leitor possa acompanhar todas as fases do ciclo experimental e possa formar um senso crítico. As seções começam por uma discussão geral para entrar em detalhes a partir de um estudo experimental. O título da seção evoca o tema principal, aquele que tomará mais tempo da discussão e especulação de alternativas, mas toda seção conterá todos os aspectos do ciclo experimental, incluindo uma rápida apresentação das observações que motivaram cada estudo. Outras teorias serão elencadas resumidamente aqui, além daquelas mais gerais apresentadas no capítulo anterior.

Para toda análise acústica que requeira um grau de a profundamento em fonética acústica experimental, o leitor encontra informação detalhada no livro de Barbosa e Madureira (2015). Todos os dados usados neste e nos demais capítulos foram obtidos rodando os scripts *SGDetector* e *ProsodyDescriptor*, disponíveis no repositório neste endereço: https://github.com/pabarbosa/prosody--scripts.

3.1 Hipóteses Científicas em Prosódia Experimental

Toda hipótese de pesquisa decorre da teoria científica que procura explicar os fatos observáveis e isso não é diferente na área da fala e da linguagem. As hipóteses formam uma ponte entre a teoria

e a metodologia que será empregada para confirmá-las, refutá-las ou refiná-las, por isso devem ser formuladas de tal forma que possam ser testadas por técnicas de medida e de inferência estatística.

Uma hipótese adequadamente formulada expressa, sob a forma de uma asserção, o que deve ser testado. Por exemplo, se admitirmos por uma teoria geral de percepção da fala, que o acento numa língua é percebido por se destacar do contexto imediato e que esse destaque é realizado por meio de parâmetros como a duração silábica, podemos emitir a hipótese de que a duração da sílaba tônica é maior do que a da sílaba átona. Autores como Massini (1991) e Barbosa (1996) mostraram que essa hipótese se confirma, desde que se assegurem condições de igualdade de contexto, uma vez que diversos fatores afetam a duração silábica. Por exemplo, em enunciados como "Não quero que ela apareça", a sílaba final é átona e dura mais do que a sílaba tônica anterior, por motivos específicos. Nesse caso, a átona dura mais por um efeito chamado de "alongamento final" (GAITENBY, 1965; OLLER, 1973; KLATT, 1975) que estende a duração de segmentos que precedem uma pausa. Além disso, a duração também depende da natureza dos segmentos e, no caso da sílaba átona desse exemplo, a duração também é maior por conta do segmento [s], que é dos mais longos do português brasileiro.

Veremos na seção 3.2 que, para testar uma hipótese, devemos ter condições experimentais em que o contexto imediato seja o mesmo, como no constraste entre os enunciados "Parece que casou sábado" vs. "Parece que caso sábado", em que, de fato, desde que pronunciadas com mesma entoação e com o mesmo ritmo, a sílaba "-sou" do primeiro enunciado (pronunciada como [zo]) é mais longa do que a sílaba -"so" (pronunciada como [zU]) do segundo.

Observe que, admitindo a mesma estrutura prosódica nos dois enunciados num determinado locutor, a única coisa que difere entre eles é a troca entre as palavras "casou" e "caso". As primeiras sílabas das duas palavras podem também ser comparadas para avaliar diferenças

duracionais entre tônica e pré-tônica pois, na palavra “caso”, a primeira sílaba é tônica e, na palavra “casou”, a primeira sílaba é átona. Em estudo comparando apenas as vogais das sílabas em entrevistas e trechos lidos, mostramos que as duas categorias de átonas se comportam da mesma forma (BARBOSA; ERIKSSON; ÅKESSON, 2013), contrariando resultados com frases isoladas em que a pós-tônica é mais curta do que a pré-tônica.

Passamos a detalhar dois exemplos de experimentos para indicar a forma como as hipóteses científicas são construídas. No primeiro exemplo, utilizaremos duas teorias concorrentes para investigar a existência do desfazimento do chamado encontro acentual, enquanto, no segundo exemplo, utilizaremos uma teoria do papel crucial da transição C-V para o processamento da sílaba, no intuito de mostrar que nossos sistemas de produção e percepção sonora estão vinculados e ancorados temporalmente nesse evento silábico.

3.1.1 Hipóteses em Pesquisa sobre Encontro Acentual

Na Fonologia Métrica de Liberman e Prince (1977), o aspecto relacional do acento pode ser indicado por uma estrutura chamada de “grade métrica”. Esse tipo de representação em grade, que representa em coluna o “grau” de saliência silábica, fornece uma ferramenta para explicar a necessidade de preservar a alternância de proeminências acentuais, alternância que caracteriza o chamado ritmo linguístico e que é garantida por uma “regra de ritmo”. A regra de ritmo se impõe na teoria quando a relação fraco-forte que existe em iambos como *thìrtéen* da grade 3.1 se inverte, soando como um troqueu (padrão forte-fraco) quando inserida em sequências como *thírtèen men*, em que a palavra *men* porta o acento frasal. O papel da regra do ritmo é o de desfazer o encontro acentual (*stress clash*) entre elementos adjacentes na grade, como se vê pelas duas colunas mais altas na grade

3.1, em que os marcadores de posição ‘x’ ocupam lugar sobre a sílaba correspondente, indicando um grau de acento que é proporcional à altura da coluna de ‘x’.

Tabela 3.1 – Grade 1, com choque acentual.

		x
	x	x
x	x	x
thir	teen	men

Pode-se ver que, na linha média da grade 3.1, os dois *x* consecutivos não têm nenhum *x* numa coluna que os separasse e que permitisse um “relaxamento” da “tensão” (termos dos autores) criada pela adjacência das duas colunas de *x* mais à direita que, assim, marcam uma contiguidade de dois níveis de saliência acentual, configurando o que os autores chamam de choque acentual[1]. Esse choque é desfeito, segundo eles, pela regra do ritmo, que age para criar a relação da grade 3.2, que soaria como se o acento estivesse na primeira sílaba da primeira palavra (observe agora que os dois *x* na linha média são intercalados pelo *x* de uma sílaba na linha inferior).

Tabela 3.2 – Grade 2, com choque acentual desfeito.

		x
x		x
x	x	x
thir	teen	men

Se adotarmos exemplo similar em português, a mesma regra do ritmo atuaria para modificar a relação métrica numa sequência como

1 Preferimos o termo “encontro acentual” por não entendermos que essa contigui dade seria sempre desfeita.

"café quente", produzindo uma palavra "café" que soaria como paroxítona, algo bastante improvável. No entanto, um desfazimento de cho- que acentual parece se dar na pronúncia fossilizada da expressão "Jesus Cristo", tão facilmente ouvida na canção de Roberto Carlos. Voltaremos a essa questão depois. Por ora, passemos a examinar a previsão da teoria dinâmica do ritmo apresentada na seção 2.3.3.

Nessa seção, vimos que o modelo dinâmico gera durações que aumentam até a realização do acento frasal, que ocorre em torno de unidade VV lexicalmente acentuada em palavra que o locutor enunciou como proeminente. Assim, se a sentença "Tomam|os um café quent|e" for produzida com dois acentos frasais, um na primeira palavra e outro na última, teremos um grupo acentual final, o segundo do enunciado, que começa depois da realização do primeiro acento frasal em "-ma-" e vai até "quen-", que é caracterizado por um movimento ascendente de duração. Por conta desse movimento de crescimento de duração, haverá um reforço de duração na segunda sílaba de "café" e não um reforço da duração de 'ca-". Observe que a previsão teórica do modelo dinâmico é oposta àquela prevista pela Fonologia Métrica em caso de desfazimento do encontro acentual. Isso ocorre porque, no modelo de osciladores acoplados, as posições de acento lexical são apenas pontos de ancoragem eventual de acento frasal.

Em estudo anterior (BARBOSA, 2002), mostramos que quatro locutores do português paulista realizam situações de encontro de acentos lexicais da forma hipotetizada pelo modelo de osciladores acoplados, isto é, com aumento de duração na sílaba mais à direita, contígua ao segundo acento lexical da sequência de duas palavras em análise, delimitadas abaixo por colchetes. Usamos pares de frases em que a primeira é uma frase-controle, sem encontro acentual, e a segunda é a frase experimental, para a qual ocorre encontro de acentos lexicais entre as palavras-chave "comi" e "bolo", entre "bordeaux" e "xucro", entre "falou" e "baixo" e entre "bebê" e "calvo", conforme abaixo, onde se indica em negrito a unidade onde incidiu o acento frasal.

- Eu [comi bol**or**] sexta-feira à n**oit**e. vs. Eu [comi b**ol**o] sexta-feira à n**oit**e;
- O [bordeaux chin**ês**] derramou-se pela m**es**a. vs. O [bordeaux x**ucr**o] derramou-se pela m**es**a;
- Parece que [falou 'baix**ou**'], e não 'ca**iu**'. vs. Parece que [falou 'b**aix**o'], e não '**alt**o';
- Um l**ind**o [bebê carm**im**]. vs. Um l**ind**o [bebê c**alv**o].

Para o locutor paulista analisado e utilizando-se um teste de ANOVA[2] com nível de significância de 5%, não foram encontradas diferenças significativas na duração média ao comparar tanto as primei- ras sílabas da palavra-chave quanto ao comparar as segundas sílabas (observe que são sílabas idênticas do ponto de vista fonológico). Na comparação entre sílabas, apenas o segundo par de frases teve diferença significativa na segunda sílaba ([do]) com valor de duração média de 151 ms na frase-controle e de 167 ms na frase experimental ($p < 0.02$). Na comparação com as unidades VV, em todos os pares a segunda VV é sempre mais longa que a primeira, na palavra-chave.

Com base nesse e em outros experimentos conduzidos (BARBOSA, 2002; BARBOSA; ARANTES, 2003; BARBOSA; ARANTES; SILVEIRA, 2004; MADUREIRA et al., 2004), a hipótese de desfazimento acentual da Fonologia Métrica foi refutada e a do modelo de osciladores acoplados confirmada. Observe como as hipóteses nas duas teorias conduziram a uma metodologia experimental que foi capaz de decidir entre uma e outra, uma vez que faziam previsões exatamente opostas. Além do mais, no caso do inglês americano, a cuidadosa investigação de Grabe e Warren (1995) revelou que existe um padrão de alternância de sílabas fortes e fracas que independe de qualquer noção

2 O teste ANOVA avalia a significância da diferença de média entre dois ou mais grupos de valores, desde que se obedeçam determinadas condições para a sua realização. Detalhes sobre esse tipo de teste no capítulo 6, seção 6.1.

de choque acentual, resultado confirmado em estudo ulterior de Kimball e Cole (2014).

Passemos agora a exemplificar um segundo experimento, sobre sincronização fala-metrônomo.

3.1.2 Hipóteses em pesquisa sobre o *p-center*

Estudos de autores como Fraisse (1982, p. 153) mostraram que a solicitação de produção espontânea de batidas repetidas do dedo indicador sobre a mesa em experimentos realizados desde a década de 1930 revela períodos com valores em torno de 600 ms que são representativos desse tipo de controle por nosso sistema motor.

Uma vez que a atividade motora da fala e a da batida do dedo indicador seriam controladas pelo mesmo mecanismo temporal gerado no córtex cerebelar (LEINER; LEINER; DOW, 1991), é de se esperar que a oscilação silábica produza períodos dessa ordem de grandeza quando somos solicitados a produzir sílabas repetidamente, como, de fato, ocorre (BARBOSA et al., 2005). Isso nos leva a pensar que podemos produzir essa repetição silábica em sincronismo com um metrônomo ou sequência de tons puros[3], ficando a questão de que lugar da sílaba ocorreria esse sincronismo, questão de pesquisa que norteou estudos na década de 1970. Esses estudos chamaram esse lugar de *perceptual-center* ou simplesmente *p-center* (MORTON; MARCUS; FRANKISH, 1976; MARCUS, 1976; POMPINO-MARSCHALL, 1989, 1991), definido como o momento no sinal acústico em que o ou- vinte se ancora para perceber uma sequência sonora como ocorrendo a intervalos regulares no tempo.

Vimos na seção 2.1 que a transição C-V é uma candidata para esse ponto de ancoragem temporal na fala, o que nos faz hipotetizar que a sílaba se sincronizaria com a batida de um tom puro exata-

3 O tom puro corresponde a um som periódico simples, formado por uma única frequência.

mente na transição C-V, isto é, no início da vogal, ponto em que há mudança brusca de energia. Consequentemente, não haveria distância entre o instante de tempo da transição C-V e esse tom.

Observe como essas asserções determinam o modo de conduzir a metodologia que deve conceber: (1) um modo de realizar essa sincronização com locutores do PB; (2) um modo de medir a distância entre a batida do metrônomo sonoro e o início da vogal; (3) um teste estatístico para avaliar se, ao menos em média, essa distância é nula. Visto que o foco de nosso sistema cognitivo na transição C-V está relacionado a transições bruscas de energia entre consoante e vogal, é importante testar o grau de sincronismo ao variar a discrepância dessas energias variando modos de articulação da consoante e altura da vogal. Tudo isso fizemos num experimento sobre p-center em PB (BARBOSA et al., 2005).

Para testar a hipótese principal sobre o sincronismo fala-metrônomo em torno da transição C-V, concebemos uma tarefa de produção de uma sequência de sílabas que o participante, um estudante paulista de cerca de 20 anos, tinha que fazer em simultaneidade com uma sequência de tons puros tocada via fone de ouvido. A primeira etapa do experimento foi aferir a taxa de elocução confortável de produção de uma sequência silábica pelo participante. Isso foi feito pedindo apenas que ele produ- zisse uma sequência de sílabas [pa], como achasse melhor, o que ele fez com um intervalo médio entre inícios de vogal de 556 ms (108 bpm).

Para a tarefa propriamente dita, cada sílaba a ser produzida repetidamente foi apresentada visualmente num cartão, numa ordem aleatória ao início de cada produção. Foram 21 sílabas CV distintas produzidas pela combinação das consoantes /p f r s ʃ l m/ com as vogais /i ɛ a/. Em torno de dez sílabas idênticas foram usadas para as análi-ses, sendo descartadas as cinco primeiras, pois foram consideradas um tempo de adaptação à tarefa. Para a realização do experimento, uti- lizamos um metrônomo analógico, de marca Matrix, modelo MR-

500, com taxas-limite de 40 a 208 bpm e a fala do participante foi capturada com microfone unidirecional com amostragem a uma taxa de 22,05 kHz com sinal do metrônomo gravado simultaneamente.

A figura 3.1 mostra a sequência de sílabas [pɛ] produzida em sincronismo com o metrônomo à taxa de 108 bpm. Observe como os pulsos do metrônomo, visíveis na parte negativa do gráfico, ficam em torno da transição CV.

A distância entre cada pulso e a sílaba correspondente foi medida calculando a fase φ definida pela equação 3.1, em que $t_{trans.CV}$ é o instante de tempo em que se dá a transição CV, no início acústico da vogal, $t_{p-center}$ é o instante da batida do pulso mais próximo do metrônomo e M é o período do metrônomo.

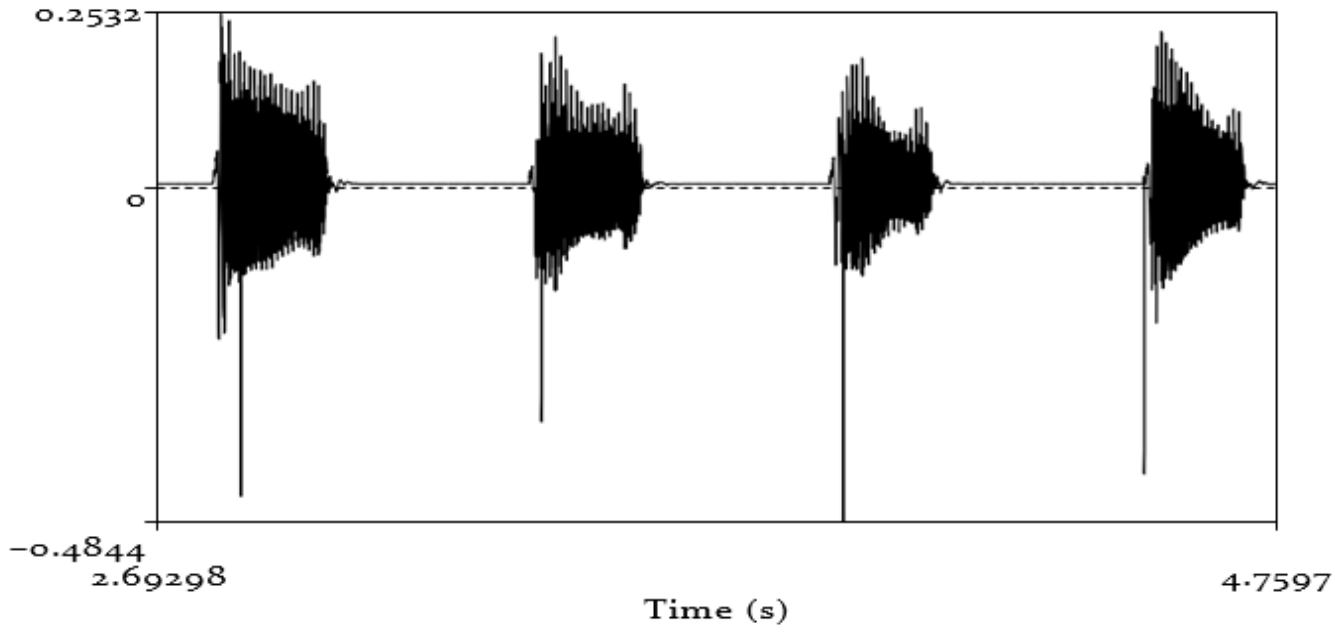

Figura 3.1 – Formas de onda de sequência de sílabas produzidas em sincronismo com o metrônomo a 108 bpm. Pulsos do metrônomo visíveis na região negativa do gráfico.

(3.1)

$$\varphi = \frac{(t_{trans.CV} - t_{p-center}) \cdot 360^{o}}{M}$$

Embora os resutados do experimento tenham revelado uma grande diferença na realização da tarefa, dependendo da sílaba considerada e da taxa do metrônomo, é importante ressaltar essa "atração" da produção do participante pela transição C-V, exceção feita quando de presença de consoante com muita energia como [ʃ]. A figura 3.2, reproduzida do capítulo 2 de Barbosa (2006), ilustra os valores médios (e desvios-padrão) da diferença de fase para todas as sílabas com o metrônomo a 108 bpm. Um teste t de amostra única[4] foi realizado para cada sílaba, adotando-se como hipótese nula que a média de $\varphi = 0$. Essa hipótese se manteve para as sílabas [pɛ], [pi], [fa], [sa], [la], [lɛ], [xa], [xɛ], [mi], isto é, todas elas tiveram seus inícios de vogal síncronos com as batidas do metrônomo.

Seis dos tipos silábicos que seguem a hipótese nula apresentam uma discrepância de energia total entre consoante e vogal seguinte das mais altas do rol de sílabas analisadas. Além disso, quando a taxa do metrônomo se lentifica, passando a 80 bpm, o sincronismo mantém um padrão semelhante à produção espontânea em relação a seu afastamento da transição C-V.

4 Este teste avalia a significância da diferença entre a média de um conjunto de dados e uma única média teórica. No caso aqui, 0. Uma apresentação formal desse teste será feita no capítulo 6, seção 6.1.

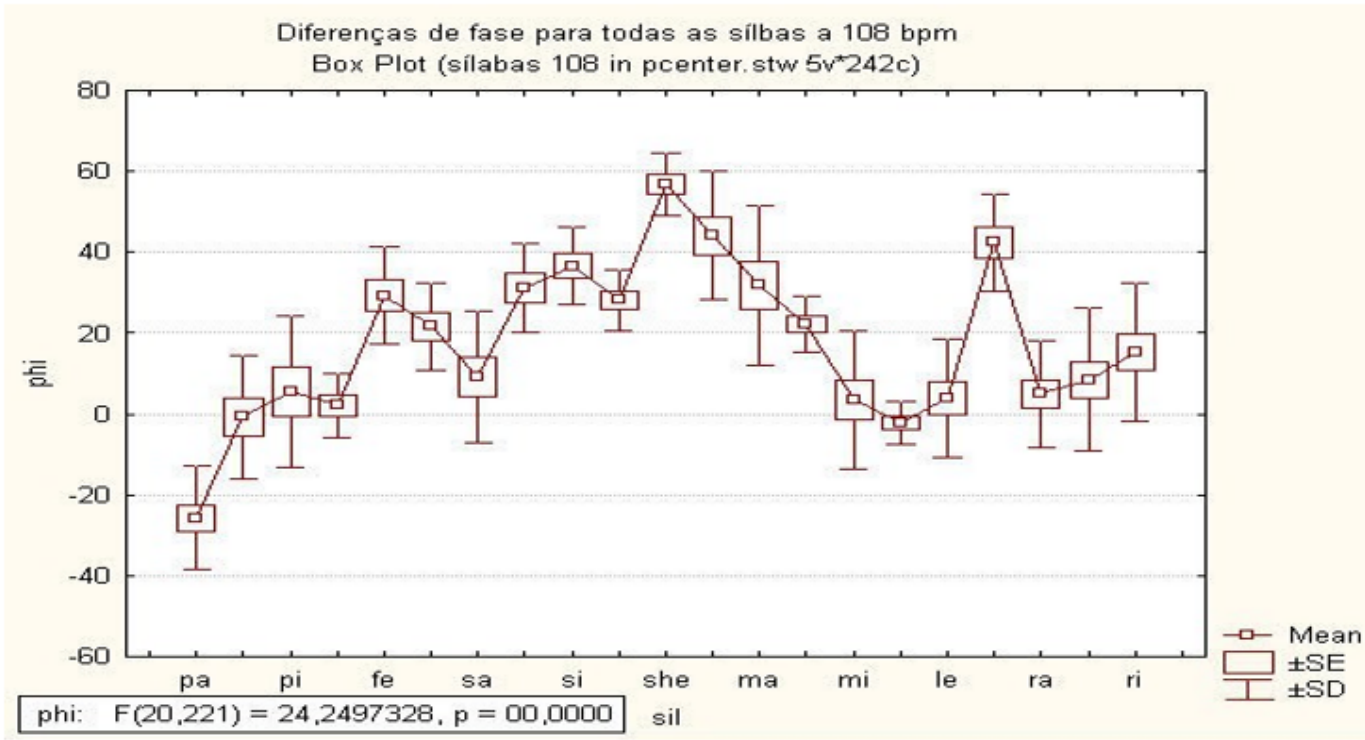

Figura 3.2 – Média e desvio-padrão (em graus) da posição do p-center em relação à transição C-V para as sílabas do experimento de sincronização fala-metrônomo com metrônomo a 108 bpm. O valor o para φ corresponde à posição da transição C-V.

Sobre os três pontos gerados pela hipótese principal, mostramos com esse exemplo (1) um modo de realizar a sincronização com locutores do PB que foi estendida a outros locutores no trabalho de Melo (2016), (2) um modo de medir a distância entre a batida do metrônomo e o início da vogal com a equação de fase acima e (3) usamos o teste t de variável única para avaliar se a hipótese de distância nula pode ser mantida.

Quanto ao último ponto, o experimento revelou um comportamento mais complexo com dependência do tipo de consoante e vogal e da taxa de metrônomo. Mas, em regra geral, os experimentos sugerem que há uma regularidade silábica que é organizada por uma sequência de transições CV. Novos experimentos podem ser feitos para confirmar esses resultados com participantes com e sem experiência musical, para ver como essa experiência afetaria o desempenho na tarefa de sincronização. Indivíduos de outras faixas etárias e outros dialetos e línguas testariam a universalidade do fenômeno. A

produção de sílabas distintas na sequência testaria como o sujeito se adaptaria à mudança de padrão de discrepância de energia em cada sílaba. Outras estruturas silábicas avaliariam se, de fato, a consoante de coda em sílabas CVC não afetaria o fenômeno de sincronização mantendo a atração pela transição C-V. O leitor pode ver que a teoria do *p-center* gera uma série de questionamentos concretos, uma característica de todo estudo experimental bem conduzido.

Vimos nesta seção e na anterior que as hipóteses foram formuladas de tal forma que conduziram à montagem de experimentos em que as observações vinculadas às teorias puderam ser verificadas a partir de medidas acústicas e de testes estatísticos inferenciais. Enquanto no experimento sobre encontro acentual apenas uma das teorias concorrentes explica o que acontece em PB em caso de encontro acentual, o experimento sobre *p-center* confirma parcialmente o sincronismo fala-metrônomo na transição C-V, pois é afetado por condições específicas. Prosseguindo com a metodologia, abordaremos os protocolos para a realização de experimentos na área.

3.2 Protocolos de Investigação em Prosódia Experimental

Os equipamentos para gravação e reprodução para estudos prosódicos são os mesmos dos usados para qualquer estudo fonético. Por isso, recomendamos a seção "Instrumentos de gravação e reprodução da fala" do livro de Barbosa e Madureira (2015) que traz a recomendação de que o microfone a ser usado seja unidirecional com uma resposta em frequência relativamente uniforme na faixa entre 30 e 16000 Hz. Para a reprodução sonora em testes de percepção da prosódia, o uso de fones de ouvido de alta qualidade é recomendado e existe uma gama grande de produtos no mercado.

O objetivo desta seção é orientar o leitor quanto a protocolos para

a realização de experimentos que dizem respeito especialmente a (1) escolha do participante[5]; (2) escolha do material a ser gravado e seleção de material para um teste de percepção; (3) protocolos experimentais para gravação de corpora e preparação de instruções para testes de percepção; (4) técnicas para obter material comparável em estilos de elocução distintos; (5) técnicas úteis para testes de percepção, como a deslexicalização.

3.2.1 Escolha do Participante

A escolha do participante depende da pesquisa que se faz e das hipóteses vinculadas à teoria adotada, mas, considerando a disponibilidade de cada um e a manutenção do bem-estar de cada pessoa, deve-se levar em conta os seguintes aspectos gerais.

Caso a pesquisa não diga respeito ao estudo de alguma patologia de fala afetando a prosódia, os participantes da pesquisa não devem ter problemas fonoarticulatórios ou auditivos. A depender do grau de importância para a pesquisa, essa constatação pode ser feita por autodeclaração ou com o auxílio de um fonoaudiólogo.

O participante deve ser capaz de realizar a tarefa que se pede e nem sempre isso é óbvio. Assim, um pequeno teste antes da coleta de dados é importante. Por exemplo, numa determinada leitura, um par- ticipante pode ter problema de fluência e, dependendo do caso, deve ser dispensado do experimento. Claro que, se as consequências da fluência em leitura para a prosódia da fala for o tema do trabalho, a escolha do participante se guiará justamente pelos níveis de fluência. Nesse caso em particular, uma ferramenta ou protocolo que avalie essa fluência é necessário para a classificação de cada um num determinado nível. A variação de fluência é inevitável em estudos de prosódia

5 Usamos o termo “escolha” de forma intercambiável com “seleção”, uma vez que,embora “seleção” assinale a obediência a um conjunto de critérios, uma parte de aleatório deve sempre ser considerada para um experimento que comporta um teste estatístico inferencial. Assim, entre dois participantes que obedecem a determinados critérios de inclusão, escolhe-se um deles para o experimento.

de língua se- gunda (L2) ou estrangeira (LE), exigindo, nesse caso, a avaliação dessa fluência ou da proficiência para que o aporte desse fator nos resultados possa ser avaliado adequadamente. Em tarefas com narrativas ou com jogos, é necessário pré-avaliar a habilidade do participante com essas tarefas, incluindo o uso dos equipamentos e das ferramentas do proto- colo experimental. Também se deve levar em conta que há pessoas que não têm muita habilidade em manusear o mouse; outras, a depender da faixa etária, não têm a capacidade de fazer determinadas tarefas por falta de treino ou maturidade motora ou cognitiva. Damos alguns exemplos.

Num experimento que fizemos, era necessário que um texto fosse lido de forma persuasiva, mas nem todos os participantes contactados foram capazes de fazer isso de forma adequada. Noutro experimento ainda, foi preciso simular uma atitude sarcástica a partir de um cenário imaginado. Mais uma vez, alguns participantes não foram capazes de fazer isso satisfatoriamente e foram descartados. A avaliação da adequação dos enunciados produzidos pela realização desses tipo de tarefa pode ser feita num teste de percepção em que se pergunta se determi- nado enunciado veicula persuasão, sarcasmo ou outra atitude ou afeto.

Mesmo quando o participante tem condições físicas e tem habilidade para fazer as tarefas, é preciso verificar se os parâmetros prosódico-acústicos são adequadamente mensuráveis em sua fala. Em caso de vozes soprosas, roucas ou com muita laringalização, por exemplo, haverá muitas falhas na medida de Fo, impossibilitando trabalhar com esse parâmetro. Para tanto, é essencial fazer um teste de gravação com verificação subsequente em programas de análise acústica para ver a continuidade dos traçados de Fo e se as fronteiras dos segmentos no espectrograma de banda larga são claramente delimitáveis na pessoa gravada. Se houver muitas falhas na obtenção desse traçado e na de- limitação de fronteiras na fala, o melhor é escolher outro participante. Nem sempre uma fala que soa bonita, agradável, é boa para

análise acústica.

Numa tarefa de percepção, por outro lado, uma tarefa de familiarização é indispensável para avaliar se as instruções são bem executadas e se a tarefa testará realmente o que se deseja. A seção 3.2.9, devotada a experimentos de percepção da prosódia, examinará esses cuidados com mais detalhe.

3.2.2 Distratores, Aleatorização e Deslexicalização

Em protocolos experimentais que envolvam o uso de frases ou palavras isoladas, é imprescindível garantir duas coisas. A primeira delas é que se intercalem frases (nos experimentos com frases) ou palavras distratoras (nos experimentos com palavras), para que o participante não infira os objetivos do experimento, pois isso afeta a forma de pronunciar ou o desempenho num teste de percepção. O número de distratores deve ser maior do que o das frases experimentais, para que seu efeito seja efetivo. Por exemplo, se um modo de realizar a entoação de questões for o tema do experimento, o número de frases interrogativas pode suscitar um comportamento desviante para evitar a monotonia ou por incomodar o participante, no caso de ele ter dificuldades com interrogativas. Para remediar isso, frases de outras modalidades como assertivas e imperativas devem ser inseridas em quantidade apropriada entre as interrogativas do experimento. Essas frases distratoras serão descartadas depois, não tomando tempo algum da análise. O mesmo se dá em testes de percepção, pois o ouvinte deve realizar uma tarefa de tal forma que o que faz num momento não influencie o que vai fazer depois.

O segundo aspecto imprescindível em leitura é que nas repetições dos trechos pelo mesmo participante e por participantes diferentes, a sequência tenha ordens distintas. Isso evita que o comportamento não desejado gerado numa leitura por conta do que se leu antes seja

reproduzido em todas as repetições pela mesma pessoa ou em todas as pessoas. Por exemplo, se numa frase um participante usa de uma ênfase numa palavra e encontra numa frase seguinte a mesma estrutura sintática, pode tender a fazer uma ênfase semelhante, enquanto, se ti- vesse lido essa frase muito mais adiante, não teria feito assim. Com a mudança de ordem das frases garante-se que o efeito indesejado não se repetirá. Esse efeito em que uma tarefa determina o comportamento na seguinte se chama efeito de "prompt". Por isso, o procedimento de aleatorização de frases ou palavras a serem lidas deve ser sempre adotado. A cada repetição, mesmo num mesmo participante, uma ordem aleatoriamente distinta deve ser usada. Se o material estiver escrito em cartões ou preparado em slides, a aleatorização simples pode ser feita, respectivamente, como se embaralhassem cartas ou com uma função de aleatorização do programa de apresentação de slides. Em ambos os casos, cada frase e palavra deve estar num cartão ou slide distintos.

No caso de testes de percepção, a aleatorização dos estímulos do teste é feita por instrução do programa que se usa para fazer o teste. O Praat, por exemplo, tem quatro métodos de aleatorização, a depender do que se deseja obter: (1) simples com ou (2) sem reposição, (3) por blocos evitando ou (4) não evitando o mesmo estímulo ouvido imediatamente antes. Convido o leitor a examinar esses procedimentos no *Help* do programa com a chave de busca *Randomization strategies*.

Ainda no caso de testes de percepção de prosódia da fala, um outro tipo de distrator pode ser necessário: fazer com que o ouvinte se concen- tre nos aspectos prosódicos e não nos segmentais ou ainda, evitar que reconheça uma língua ou um locutor conhecido. O reconhecimento de uma língua pode prejudicar o experimento, pois pode induzir um com- portamento específico no participante. Reconhecer um locutor pode facilitar de forma não desejada um teste de percepção. Por exemplo, num teste de reconhecimento de estilos de elocução entre jornalístico e político, se o participante ouve o enunciado sem nenhuma modificação e reconhece que quem fala é o Bóris Casoy, res-

ponderá que o estilo é jornalístico. Para fazer com que o participante se concentre na forma como o locutor fala, em sua prosódia, existem técnicas de deslexicali- zação.

Um dos métodos usados mais remotamente é o da inversão do sinal de áudio que o leitor pode ouvir como exemplo no arquivo **VentoSu- lInvertido**. A vantagem é que não se entende o que foi dito, mas, fora isso, apresenta duas desvantagens principais. O locutor pode ainda ser identificado pelo tom da voz e a curva melódica fica invertida, impos- sibilitando o reconhecimento da entoação do enunciado.

Outro método usado até hoje é o da filtragem passa-baixas, que usa um filtro digital para manter apenas as frequências do sinal de fala abaixo de determinada frequência de corte. O leitor pode ouvir dois exemplos com cortes distintos nos arquivos **VentoSulFiltrado200** e **VentoSulFiltrado400**. Note que é muito difícil reconhecer o que foi dito, embora, talvez, na filtragem a 400 Hz, se soubéssemos o assunto antes, pudéssemos inferir algo. Para que esse método funcione, deve-se preservar a curva de Fo e, para tanto, saber a frequência máxima de Fo no trecho que a pessoa fala, sob o risco de alterar a curva de Fo e tornar equívoca a percepção da entoação. Neste locutor a frequência máxima é de 210 Hz, sendo o corte a 200 Hz algo que deve ser evitado. Aqui parece não ter prejudicado tanto, pois a entoação e ritmo da fala parecem semelhantes nos dois áudios. A filtragem passa-baixa pode ser feita no Praat selecionando o objeto de áudio (Sound no Praat), escolhendo no menu *Filter* a opção *Filter (pass hann band)* para então escrever os limites de frequência a ser preservada entre 0 e a frequência de corte.

Ainda outro método de deslexicalização é o PURR (*Prosody Unveiling through Restricted Representation*), método que propõe preservar a prosódia do enunciado substituindo o sinal original por uma versão que usa uma onda senoidal. Baseando-se nessa filosofia, Petra Wagner implementou um script para o Praat que usa a função $sinc(x) = sen(x)/x$ que é alterada para se iniciar a cada pulso glotal

do sinal de fala com valor de frequência previamente extraído por um algoritmo específico do Praat. A intensidade é também preservada, no entanto, nenhuma transição formântica entre consoantes e vogais é mantida nesse método. O mesmo exemplo de áudio modificado por esse método pode ser ouvido no áudio **VentoSulPurr**. O script pode ser obtido neste lugar: <PURR-2004>. Na seção 3.2.9 daremos exemplos de experimentos que usaram esse método.

A figura 3.3 mostra as curvas de Fo do áudio original, que pode ser ouvido no site do livro com o nome **VentoSulOriginal**, e dos áudios deslexicalizados pelos métodos de inversão do sinal, de filtragem passa- baixas nas duas frequências e do algoritmo PURR.

3.2.3 Escolha e Cuidados com o Material para Gravar

Tendo em vista que a maior parte da pesquisa em prosódia envolve algum aspecto de controle experimental, é necessário obter gravações dos participantes ou fazer com que se submetam a um teste de percepção. Tanto uma tarefa quanto outra não deve passar de 30 minutos seguidos. Havendo necessidade de mais material gravado ou de um teste de percepção mais longo, é preciso organizar esses períodos de coleta em sessões em que esse tempo-limite seja respeitado. Evita-se num caso o cansaço vocal e, no outro, a sobrecarga cognitiva.

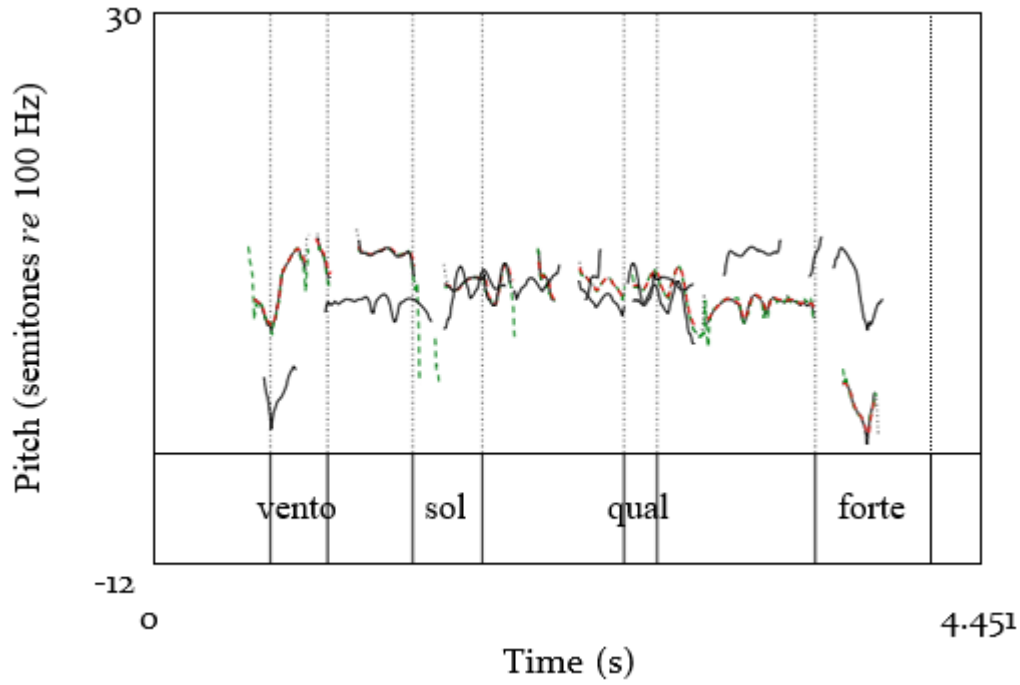

Figura 3.3 – Curvas de Fo do enunciado "O vento sul e o sol discutiam qual dos dois era o mais forte." após três procedimentos de deslexicalização. A curva original e aquelas pelo método de filtragem passa-baixas (curva pontilhada vermelha - 400 Hz - e curva tracejada verde - 200 Hz) e o método PURR coincidem, enquanto a por inversão tem a curva invertida também. Apenas algumas palavras são mostradas para facilitar a visualização de trechos da curva melódica.

Como discorremos no primeiro capítulo, a fala de laboratório abrange muito mais do que a leitura de textos: é toda fala obtida com algum grau de intervenção do pesquisador (XU, 2010). Em qualquer experimento em que se deseja compreender a forma e a função prosódicas em situações espontâneas, é inevitável que um procedimento de controle experimental não entre em jogo. A leitura de frases isoladas foi, durante muito tempo, o tipo de material gravado mais usado na investigação fonética tanto segmental quanto prosódica. Mas, se de um lado permite investigar com eficácia o que se altera na forma em contrastes de determinada função comunicativa nos eixos paradigmático ou sintagmático, por outro lado dificilmente o contraste se dá espontaneamente. Vejamos um exemplo.

Num experimento sobre os correlatos acústicos do acento lexical em PB, contrastamos dois estilos de elocução, fala lida e fala de entre-

vista, para ver se os correlatos se mantinham os mesmos para assinalar o grau de tonicidade das vogais de oxítonas, paroxítonas e proparoxítonas com número de sílabas variando de 2 a 6 (BARBOSA; ERIKSSON; ÅKESSON, 2013). Neste estudo usamos do seguinte procedimento: gravamos as entrevistas entre amigos próximos para assegurar mais material. Em seguida, transcrevemos as entrevistas ortograficamente e escolhemos trechos das próprias entrevistas com frases mais apropriadas para leitura a serem lidas pelas mesmas pessoas duas semanas depois. Examinamos três parâmetros prosódicos em vogais em posição tônica, pré-tônica e pós-tônica nos três padrões acentuais lexicais do PB: duração, desvio-padrão da Fo e intensidade relativa (ênfase espectral). Seguindo esse procedimento, foi possível comparar diretamente os parâmetros em dois estilos de elocução, fala lida e fala de entrevista. Os resultados mostraram que os parâmetros mantiveram a mesma hierarquia de importância em revelar o acento lexical, sendo a duração bem superior aos demais parâmetros prosódicos.

Por conta de muitas vezes haver necessidade desse contraste entre fala não planejada de antemão (como na entrevista, em narrativas) e fala lida, a leitura é ainda muito usada nos estudos prosódicos, desde a frase isolada até textos de diversas naturezas.

Tanto para estudos em uma língua pouco investigada, quanto para estudos de estilos de elocução em que o papel de determinadas funções prosódicas bem como as formas prosódicas a elas associadas são pouco conhecidas, é importante utilizar material a ser lido nos experimentos. Num trabalho sobre o estilo jornalístico em PB e em francês da França (MAREÜIL; BARBOSA, 2018), utilizamos um texto curto, de cerca de 100 palavras nas duas línguas que foi lido em estilo habitual e jornalístico por quatro profissionais do jornalismo em Campinas e em Paris. O texto lido foi exatamente o mesmo, somente os enunciados foram produzidos de forma a reproduzir uma leitura habitual e uma leitura jornalística por pessoas acostumadas a fazer isso em sua profissão. Pelo estudo dos grupos acentuais realizados pe-

los participantes, chegamos à conclusão de que a proporção de uso de proeminências iniciais aumenta no estilo jornalístico, sobretudo em francês, que a taxa de elocução diminui de 10 a 43% nesse estilo e que a frequência fundamental mediana é superior no estilo jornalístico, mais nos homens em francês e nos dois sexos em PB.

A frase isolada, por sua vez, pode ser usada para aumentar a compreensão de aspectos básicos da prosódia, como efeitos segmentais, realização de diferentes tons de fronteira, realização de diferentes tipos de foco, diferenças na realização de modalidades frásticas como assertiva, interrogativa, imperativa e mesmo estudos dos efeitos de diferentes atitudes e emoções.

Os efeitos segmentais se referem tanto à interferência da produção da prosódia nos segmentos fônicos (consoantes e vogais) quanto a dos segmentos fônicos na prosódia. É, portanto, uma interferência de mão dupla. Desse modo, verificam-se modificações tanto nos parâmetros prosódico-acústicos (Fo, duração e intensidade) diante da ocorrência de segmentos fônicos com determinadas características, quanto nas propriedades acústicas dos segmentos fônicos por modificações na estrutura prosódica de um enunciado. Um dos mais conhecidos fenômenos desse tipo de interferência mútua é a chamada micromelodia, uma modificação local na curva de Fo que se verifica sob o efeito da ocorrência de um fone vozeado ou não vozeado. Assim, no contraste “Fizeram a sobremesa usando a nata disponível.” vs. “Fizeram a sobremesa usando nada de caro.”, a realização de [t] em “nata” aumenta localmente a vibração das pregas vocais na vogal seguinte, fazendo com que a curva de Fo no início dessa vogal comece com valor mais alto. Exatamente o movimento contrário, ocorre depois do [d] em “nada”, ou seja, no início da vogal que segue essa consoante a curva de Fo se inicia com valor mais baixo.

Um exemplo do segundo tipo de efeito segmental que pode ser estudado pelo contraste de frases isoladas é a mudança de intensidade, de duração e de primeiro formante do [s] na palavra “saco” em condi-

ções distintas de foco em: “Ele comprou um SACO de estopa.” (após alguém ter falado que era uma caixa) vs. “Ele comprou um saco de estopa.” (após uma pergunta sobre o que a pessoa fez naquele dia). Observe que o uso de frases isoladas permite assegurar o mesmo contexto fonético para aquilo que se quer verificar, o efeito da prosódia sobre o segmento [s].

É também fundamental, para se construir um conhecimento de como se realiza acusticamente uma fronteira prosódica, a comparação entre unidades linguísticas de sentenças distintas. Veja o seguinte contraste entre diferentes tipos de fronteira prosódica ao fim da palavra “cedo”, sendo as duas primeiras terminais e as duas seguintes não terminais. O locutor é de identidade paulista com cerca de 50 anos na época da gravação.

- Logo cedo? Paulo foi pra São Paulo.
- Logo cedo. Paulo foi pra São Paulo.
- Logo cedo Paulo foi pra São Paulo.
- Logo cedo, Paulo foi pra São Paulo.

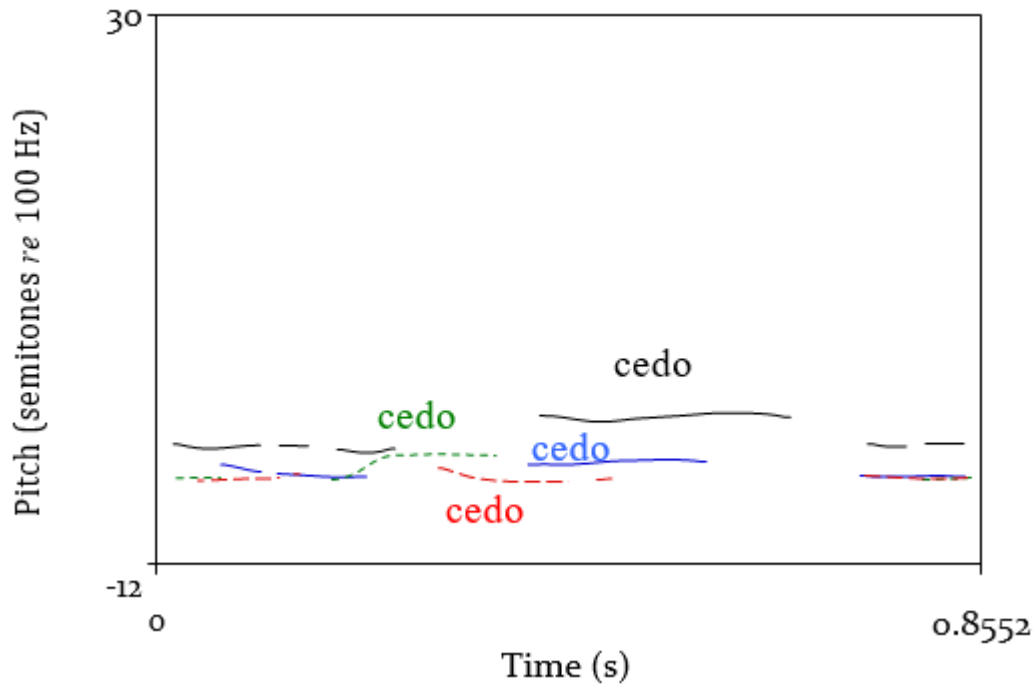

Figura 3.4 – Curvas de Fo do trecho “Logo cedo Paulo” de quatro enunciados de locutor paulista extraídos de “Logso cedo. Paulo foi pra São Paulo.” (vermelho, tracejado); “Logo cedo? Paulo foi pra São Paulo.” (verde, pontilhado); “Logo cedo Paulo foi pra São Paulo.” (cheia, mais alta); “Logo cedo, Paulo foi pra São Paulo.” (cheia, mais baixa).

Na Figura 3.4, ilustramos o contraste das curvas de Fo dos quatro exemplos para o trecho “Logo cedo Paulo”. Observe nos exemplos de fronteiras não terminais indicadas pelos contornos de linhas cheias que a curva sobe e fica nivelada em “cedo” caindo depois, no início de “Paulo”. Já nos exemplos de fronteiras terminais, indicadas pelos contornos de linhas tracejada e pontilhada, se pode ver em “logo cedo” assertivo que a curva desce durante “cedo” enquanto em “logo cedo” interrogativo a subida da curva está contida na palavra “cedo”. Nesses dois últimos exemplos há uma pausa silenciosa entre “cedo” e “Paulo”.

Muito frequentemente, o conhecimento de como são os perfis de F0 numa situação controlada como essa, com pelo menos três funções distintas, permite identificar esses mesmos perfis ou componentes deles na fala espontânea, como se vê no exemplo da Figura 3.5. Nesse exemplo chama-se a atenção para os perfis de F0 das palavras que precedem fronteiras não terminais, “opção” e “público”. Pode-se

ver que, tanto na leitura quanto na entrevista, são perfis majoritariamente ascendentes. O perfil é de uma subida mais acentuada em "opção", na leitura em contraste com a entrevista, e mais semelhante na segunda palavra, embora tenha uma descida curta ao final por conta de uma laringalização durante as pós-tônicas de "público", na entrevista.

A entrevista foi sobre tópico relacionado aos estudos e à vida profissional, entre amigos próximos. E a leitura, de trechos selecionados da entrevista transcritos ortograficamente e lidos pela mesma pessoa duas semanas depois, três vezes em ordem aleatória entre os trechos selecionados.

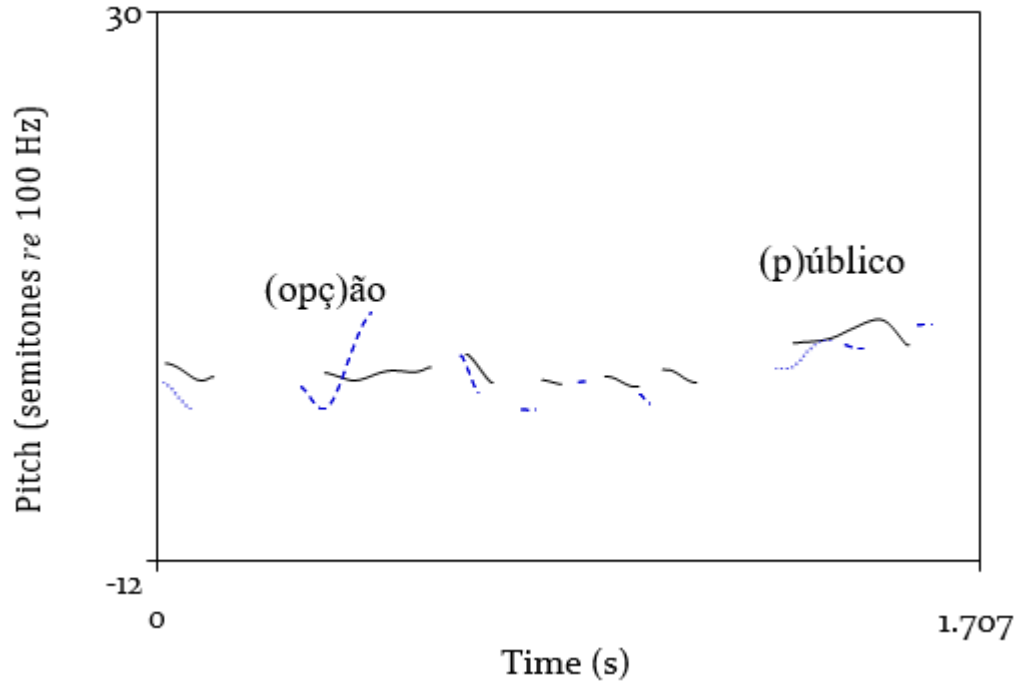

Figura 3.5 – Curvas de F0 do trecho "de opção, prestei concurso público [. . .]." de enunciado mais amplo tirado de uma entrevista entre amigos (linha cheia) e de leitura duas semanas depois do mesmo trecho pelo mesmo locutor paulista de cerca de 25 anos.

O mesmo se dá em diferentes tipos de foco estreito. Reconhecer na fala espontânea um foco contrastivo, por exemplo, é uma tarefa facilitada se se conhece como se realiza em laboratório. No exemplo acima, em que usamos a palavra "saco" para exemplificar o efeito da

prosódia sobre o segmento [s], pode-se montar um protocolo experimental para o estudo dos diferentes perfis de F0 em enunciados contrastando a mesma palavra com diferentes tipos de foco: foco informacional, foco contrastivo, ausência de qualquer foco ou em posição pós-focal. Para tanto, bastaria instruir o participante a primeiramente ler a frase que será a controle, sem qualquer tipo de foco. Para as demais, instrui-se da seguinte forma: na com foco informacional se apresenta para leitura a frase que contém a palavra "saco" grafada em maiúsculas e se pede para que leia o que está em maiúsculas com ênfase. Para o foco contrastivo, por sua vez, diz-se ao participante que a palavra em maiúsculas sublinhada, por exemplo, deve ser lida para deixar claro ao interlocutor que não é um "monte" e sim um "saco" de estopa. E para obter a frase com a palavra "saco" de forma desfocada, colocam-se as maiúsculas na palavra seguinte, "estopa", instruindo o participante para dar ênfase nessa outra palavra. As frases devem ser repetidas algumas vezes (pelo menos cinco) e ser intercaladas com frases distratoras.

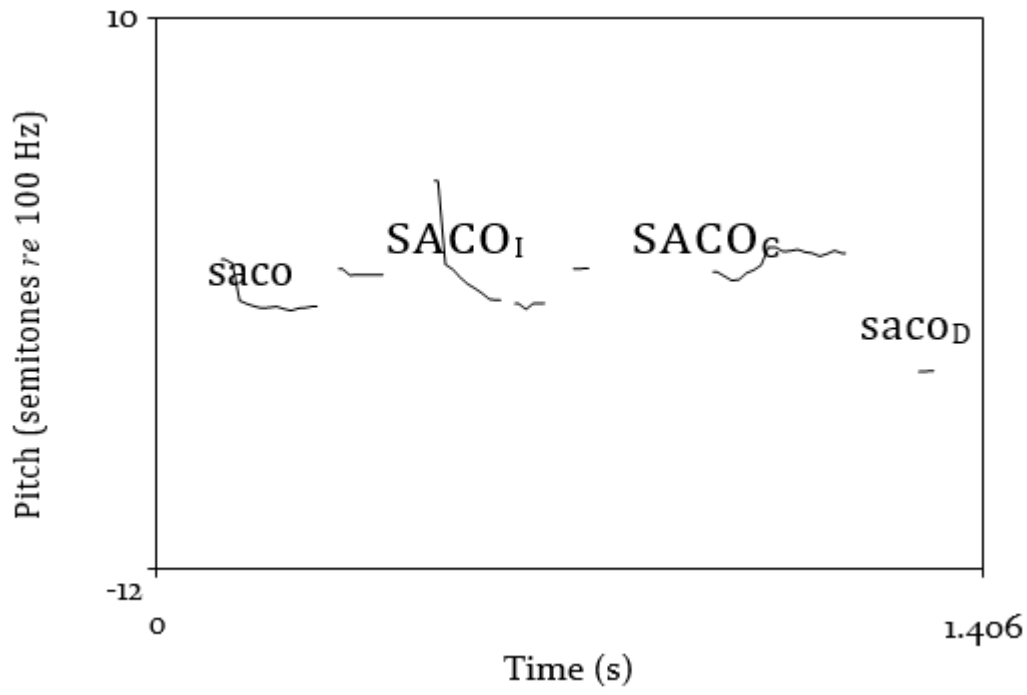

Figura 3.6 – Curvas de F0 extraídas da sentença "Ele comprou um saco de estopa." com quatro condições de foco, da esquerda para a direita: sem foco, com foco informacional, com foco contrastivo e desfocado.

Na Figura 3.6 pode-se ver que, no primeiro perfil de F0, sem foco algum, após o efeito elevador de F0 do [s], o contorno fica nivelado na tônica [a] e sobe na pós-tônica. Na mesma tônica, o perfil é descendente na palavra com foco informacional, enquanto é ascendente no foco contrastivo. A palavra desfocada é muito curta e foi dita com alguma soprosidade, tendo apenas alguns valores válidos e baixos de F0. Esses padrões poderão ser verificados, em seguida, na fala espontânea, a partir do conhecimento adquirido pelo pesquisador na situação controlada apresentada aqui. Vale a pena, sobretudo, observar os movimentos ascendente e descendente nos dois tipos de foco, pelo fato de a mudança na direção do movimento determinar diferenças importantes na percepção de proeminência melódica.

O efeito de atitudes e emoções na fala é o de afetar normalmente a prosódia como um todo nos enunciados, não tendo efeitos majoritariamente de domínio restrito como no caso de foco estreito mesmo que, como vimos, o foco estreito em "estopa" afete as condições de realização da palavra "saco". O problema do estudo de atitudes e emoções é a possibilidade de obter resultados satisfatórios em fala não espontânea, pois é algo que requer capacidade de interpretação. Vejamos o caso do estudo de Silva (2019) na seção seguinte.

3.2.4 Protocolo Experimental: Atuação de Atitudes

O corpus de fala montado por Silva (2019) em sua tese de doutorado foi formado por gravações de dez sentenças, produzidas por onze participantes brasileiros, sendo seis mulheres e cinco homens. Essas sentenças foram produzidas em atitudes de ironia sarcástica, sarcasmo puro e fala neutra, com enunciados correspon-dentes validados por teste de percepção. Por esse procedimento, 236 enunciados foram considerados válidos e retidos para as análises prosódico-

-acústicas. A validação é necessária justamente para reter apenas os enunciados que realmente passam as atitudes requeridas no estudo, tendo em vista que nem sempre é simples fazer essa interpretação. Silva utilizou para isso de um texto escrito apresentando um cenário em que a sentença experimental era usada para marcar a atitude naquela situação específica num contexto dialógico. Vejamos o exemplo para produzir sarcasmo puro e ironia sarcástica. Dois participantes, A e B, liam e procuravam "sentir" as situações seguintes para produzir de forma apropriada as duas atitudes nas sentenças em negrito.

No caso de sarcasmo puro foi esta a situação usada para a sentença "Você deveria passar protetor antes de sair de casa":

> A: Eu já estou vermelha no corpo inteiro novamente. E isso porque eu apliquei uma espessa camada de protetor solar antes de sair para a praia hoje.
>
> B: Sério? Qual você usa?
>
> A: Aquele que eu comprei na cidade recentemente. Você estava lá, lembra?
>
> B: Mas aquele lá só tem fator de proteção 5! Isso é muito pouco! **Você deveria passar protetor antes de sair de casa.** (E quero dizer um protetor de verdade.)

enquanto, para a ironia sarcástica, a situação foi a seguinte:

> A: Finalmente o inverno acabou! Vou à praia esta tarde.
>
> B: Está apenas 5º C! Não acho que esteja bom lá ainda!
>
> A: Mas o tempo está tão lindo! Só por acaso vou pegar o guarda-sol e meu calção de banho.
>
> B: Claro... Porque está super quente. **Você deveria passar protetor antes de sair de casa**.

No seu primeiro estudo de produção, Silva investigou se a expressão das atitudes modifica um ou mais entre 17 parâmetros prosódico-acústicos calculados globalmente nos enunciados. Os parâmetros acústicos foram descritores estatísticos da frequência fundamental, intensidade global e relativa, duração e qualidade de voz, extraídos dos enunciados validados pelo uso de um script para o Praat. Esses parâmetros serão trabalhados nos capítulos 4 e 5.

Quatro dos cinco parâmetros relacionados à frequência fundamental foram significativamente distintos para a ironia sarcástica quando comparada à fala neutra, sendo que a mediana, o valor máximo e o valor mínimo de Fo[6] tiveram médias menores em relação à atitude neutra, resultado condizente com o obtido para línguas germânicas como o alemão e o inglês e o oposto dos encontrados para línguas românicas como o italiano e o francês, mostrando que aspectos atitudinais dependem mais de convenções sócio-culturais, conforme explicado em detalhe na tese de Silva (2019).

Para a qualidade de voz, o resultado mais interessante é que, em relação à atitude neutra, tanto a ironia sarcástica quanto o sarcasmo puro tiveram redução da relação harmônico-ruído, o que aponta para um aumento do ruído espectral nesses enunciados, isto é, mais soprosidade na fala. Observe que as modificações estudadas afetam os enunciados como um todo. Esse aspecto foi mais importante do que modificações locais (em palavra específicas, por exemplo). Observe ainda que, embora sejam contrastadas nesse estudo frases, elas só se encontram isoladas, fora de contexto, na condição neutra. Esse aspecto pode explicar uma leitura encontrada mais rápida nas frases experimentais (as com sarcasmo puro e ironia sarcástica) obtidas a partir da situação colocada para os participantes. O uso de um mesmo texto em diferentes estilos pode ser uma alternativa para explicar essa diferen-

6 Seis descritores apresentados no capítulo 5.

ças nos contrastes examinados aqui.

Em dois outros estudos mostrados na próxima seção apresentamos protocolos em que se faz um paralelo entre diferentes estilos de fala.

3.2.5 Protocolo Experimental: Entrevista Informal pareada a leitura

No estudo de Barbosa, Eriksson e Åkesson (2013), pareamos fala lida e entrevista informal de forma a obter material comparável para o estudo dos correlatos acústicos do acento lexical nas vogais do PB. Embora a duração silábica (e consequentemente seu principal componente, a vogal) já tenha sido apontada como parâmetro mais importante em apontar o acento lexical em PB, algumas questões haviam ficado em aberto do ponto de vista experimental, uma vez que os estudos se restringiram à fala lida (BARBOSA, 1996; MASSINI, 1991; MORAES, 1987; FERNANDES, 1976). As principais questões dizem respeito (1) ao grau de importância dos diferentes parâmetros prosódico-acústicos para assinalar o acento lexical, (2) se a hierarquia entre esses graus se mantém em outros estilos de fala e (3) a eventuais diferenças entre homens e mulheres no uso dos parâmetros relativos ao acento lexical.

Assim, as hipóteses científicas foram assim enunciadas: (1) a duração é o principal parâmetro que assinala o acento lexical, independentemente do estilo; (2) a intensidade global é o parâmetro na sequência de importância, tendo em vista os estudos de Massini (1991), Moraes (1987), Fernandes (1976); (3) a F0 teria uma importância menor, tendo sobretudo uma função entoacional; (4) não há diferença no uso e hierarquia dos parâmetros do acento entre homens e mulheres.

Para responder a essas hipóteses, montamos um corpus com as seguintes características e participantes: 5 homens (de 21 a 30 anos) e 5 mulheres (de 18 a 26 anos) do Estado de São Paulo, todos univer-

si- tários com Graduação completa que deram uma entrevista informal a algum amigo muito próximo, que fazia a gravação. Cada participante teve, assim, seu próprio entrevistador, com todas as gravações feitas na sala do Grupo de Estudos de Prosódia da Fala, e usando um microfone Shure SM58 conectado à placa de som ProTools com amostragem de 22050 Hz. Em seguida, fizemos a transcrição completa das entrevistas. Do material transcrito dos 10 entrevistados, selecionamos 15 trechos com uma sintaxe compatível com a de texto escrito e montamos 15 cartões escritos sem eventuais hesitações. Esses trechos foram se-lecio- nados a partir de palavras neles contidas que assegurassem, no total, uma proporção de oxítonas, paroxítonas e proparoxítonas semelhante à encontrada no PB (CINTRA, 1997). Os trechos foram lidos três ve- zes pelas mesmas pessoas, duas semanas depois, com leitura em ordem aleatória. De cada trecho foi escolhida uma palavra para ser analisada, que também foi transcrita isoladamente para ser lida. Assim, no total, para cada participante, foram analisadas 15 palavras distintas nos três estilos de elocução: Entrevista Informal entre amigos próximos (EI), Leitura de trechos da Entrevista transcrita (LE) e Leitura de Palavras isoladas (LP).

Para o total de 150 palavras (10 participantes x 15 palavras por participante), a distribuição do padrão acentual lexical foi de: 70% de paroxítonas, 20% de oxítonas e 10% de proparoxítonas, compatível com a distribuição desse padrão em PB segundo Cintra (1997). Quanto à extensão das palavras, foi de 2 a 6 sílabas, sendo 84% de 3 e 4 sílabas. As palavras selecionadas se encontravam na proporção de 62% em posição medial na frase, sendo 82% dessas em situação de proeminência nos dois estilos. Para cada palavra produzida foram medidas nas vogais as variáveis dependentes que seguem.

- Duração em ms. O número e a variedade das vogais dispensam o procedimento de normalização duracional explicado no capítulo

4: 1610 vogais para os homens e 1728 vogais para as mulheres.

- Mediana e desvio-padrão da Fo em Hz e em semitons, uma medida logarítmica que simula a sensação de *pitch* (mais no capítulo 4);
- Ênfase espectral em dB. A ênfase espectral (EE) foi definida por Traunmüller e Eriksson (2000) pela equação EE = L - L_o, em que L é a energia de todo o espectro e L_o é a energia da banda baixa do espectro da frequência 0 até o valor limite de 1,43 x média da Fo na vogal. Essa medida é correlato do esforço vocal.

Para mostrar as diferenças entre os valores médios dos parâmetros acima entre os estilos e os níveis acentuais em cada gênero, utilizamos uma ANOVA de dois fatores e um teste *post hoc*.[7] Para avaliar o quanto um parâmetro explica as diferenças de médias entre os níveis de acento e entre os estilos, usamos um teste estatístico chamado de tamanho do efeito (*effect size*).

O cálculo do tamanho do efeito mostrou que a duração é o correlato principal do acento lexical independentemente de estilo e que, diferentemente dos trabalhos anteriores, vogais pré-tônicas e pós-tônicas não diferem em duração a não ser na situação de palavra isolada (ainda apenas para os homens), via teste post hoc. Mostramos ainda que o acento lexical explica uma maior percentagem da variância dos parâmetros “duração” e “desvio-padrão” de F0 do que o estilo de elocução. As vogais pós-tônicas têm menos ênfase espectral nos três estilos com relação à tônica. Assim, uma queda em ênfase espectral é sinal de que a vogal precedente é acentuada lexicalmente, ainda que a F0 varie mais em tônicas e pós-tônicas, especialmente nas palavras isoladas. Em mulheres, a variação de F0 é mais importante para explicar acento lexical do que a ênfase espectral.

7 O primeiro teste é mais geral e avalia a significância da diferença de média entre os conjuntos de valores para as duas variáveis independentes, grau de acento e estilo. Quanto ao segundo teste, ele avalia entre quais conjuntos de dados existe a diferença. Detalhes sobre esses tipo de testes no capítulo 6, seção 6.1.

3.2.6 Protocolo Experimental: Estilos de Elocução

Dois estudos experimentais que fizemos permitem levantar uma discussão sobre outros aspectos da pesquisa experimental em prosódia. O primeiro deles abordou a questão de imitação de fala (BARBOSA; MAREÜIL, 2018). É sabido que a imitação da fala retém apenas os aspectos mais salientes do locutor ou a representação fonológica do enunciado (COLE; SHATTUCK-HUFNAGEL, 2011). Sendo assim, pensamos que a imitação do estilo telejornalístico seria marcado por uma tendência para a proeminência inicial, pois essa tem sido observada tanto no estilo telejornalístico francês quanto em PB em palavras com grande carga semântica (e.g., BIlhões de reais), mas não somente nessas, quando se pensa nas frases de abertura das notícias.

No que tange o estilo profissional de locução em geral, resultados de um estudo prévio sobre locução de rádio (CAMPOS, 2012) indicaram que as principais mudanças quando da imitação desse estilo por um profissional do rádio em comparação com sua própria entrevista informal foi um aumento de 12% da mediana de F0 e um aumento da taxa de acentos de *pitch*. Propusemo-nos então a comparar imitações do estilo jornalístico em duas línguas/culturas, o francês da França e o português brasileiro, nas variedades mais disseminadas: a locução em telejornais de Paris e aquela no eixo Rio-São Paulo, respectivamente, por representarem a norma de pronúncia dos respectivos países. Para tanto examinamos quais parâmetros prosódico-acústicos são mo- bilizados para assinalar dois tipos de imitação que lançam mão seja da memória de longo termo, seja daquela de curto termo.

Como hipóteses de trabalho, tendo em vista resultados de estudos prévios (CAMPOS, 2012; CASTRO, 2008), esperamos, na imitação do estilo de telejornal: (1) um aumento na mediana e no desvio-padrão de F0; (2) um aumento na proporção de proeminência inicial; (3) a con-

vergência de alguns parâmetros, especialmente taxa de elocução, com os parâmetros do modelo imitado consecutivamente; (4) uma igualdade de resultados para ambas as línguas.

Para essa pesquisa selecionamos quatro jornalistas em cada país, de Paris e de Campinas, dois homens e duas mulheres em cada caso. O número diz respeito, sobretudo em Paris, à dificuldade de conseguir profissionais disponíveis e com tempo para colaborar na pesquisa. Cada um deles leu um texto de três maneiras em sua língua. O texto foi *La Bise et le soleil* e sua correspondente tradução para o PB, "O vento sul e o sol". As três maneiras foram: (1) de forma neutra - NE; (2) no estilo telejornalístico de cada país, de acordo com a internalização de cada um do que é o estilo telejornalístico - JM; (3) no estilo de uma telejornalista de cada país, em que, após ouvi-la, fazia-se a locução sobre outro assunto tentando imitá-la (imitação consecutiva) - JC. As leituras foram uniformemente divididas em 36 grupos acentuais (AP) em PB e 39 em francês com pelo menos 3 sílabas no grupo acentual em cada língua. Textos e grupos acentuais analisados podem ser vistos abaixo, primeiro em PB e depois em francês. Somente os grupos acentuais entre colchetes foram analisados.

[O vento] sul e o sol [discutiam] qual dos dois era [o mais forte], quando passou [um viajante] [envolto] [num casaco]. [Ao vê-lo], [apostaram] que [aquele] que [primeiro] [conseguisse] [obrigar] [o viajante] [a tirar] [o casaco] [seria] [considerado] [o mais forte]. [O vento] sul [começou] [a soprar] [com muita força], mas quanto [mais soprava], [mais o viajante] [se embrulhava] [no seu casaco], [até que] [o vento] sul [desistiu]. O sol brilhou então [com toda intensidade], e [imediatamente] [o viajante] tirou [o casaco]. [O vento] sul teve assim [de reconhecer] [a superioridade] do sol.

La bise et [le soleil] [se disputaient], chacun [assurant] [qu'il était] [le plus fort], [quand ils ont vu] [un voyageur] [qui s'avançait], [enveloppé] [dans son manteau]. [Ils sont tombés] d'accord [que celui] [qui arriverait] [le premier] [à faire ôter] [son manteau] [au voyageur] serait [regardé] [comme le plus fort]. Alors, la bise s'est mise [à souffler] [de toute sa force] mais [plus elle soufflait], [plus le voyageur] serrait [son manteau] [autour de lui] et [à la fin,] la bise [a renoncé] [à le lui faire ôter]. Alors [le soleil] a commencé] [à briller] et [au bout d'un moment], [le voyageur], [réchauffé], [a ôté] [son manteau]. Ainsi, la bise [a dû reconnaître] [que le soleil] était [le plus fort] des deux.

Dentro de cada grupo acentual realizado pelos quatro jornalistas em cada língua, medimos as variáveis seguintes:

- Proporção de proeminências iniciais em cada leitura. Essa proeminência inicial foi toda palavra com saliência em borda esquerda que não fosse a posição de acento lexical. Por exemplo, em "discutiam", foi contado como proeminência inicial se havia essa saliência na sílaba "dis" em PB e, no caso, do francês nas sílabas *se* ou *dis* do grupo acentual *se disputaient* ;
- Descritores estatísticos da F0: mediana, máximo, amplitude de variação (máximo - mínimo), desvio-padrão bruto e normalizado

pelo valor da mediana, semi-amplitude entre quartis[8] bruta e normalizada pela mediana, com todos os valores em semitons;
- Tempo total de leitura e tempo total de pausa silenciosa (soma dos valores das durações das pausas silenciosas);
- Intensidade relativa calculada pela fórmula da ênfase espectral, conforme visto na seção anterior.

Para a análise estatística, usamos o teste não paramétrico de dois fatores de Scheirer Ray Hare (SHR), equivalente à ANOVA de dois fatores, com os fatores SUJEITO (4 níveis) e ESTILO (3 níveis) e, em seguida, utilizamos o teste *post hoc* não paramétrico de Wilcoxon[9]. Para todos os casos o nível de significância α foi fixado em 1%. A razão do uso do teste não paramétrico é que, em nenhum dos casos, os resíduos passaram no teste de normalidade e a razão de se usar um nível de significância mais baixo é diminuir a chance de erro do tipo I, por termos um número baixo de participantes.

Os resultados significativos para o PB revelaram uma proporção de proeminência inicial de cerca de 57% nos estilos neutro e imitação de cor contra 67% no estilo imitação consecutiva. Para F0, a mediana é maior em 2 semitons no estilo imitação de cor em relação aos demais, tendo uma amplitude de variação também maior de 2 semitons nos estilos de imitação em relação ao neutro, mas apenas nos grupos acentuais contendo proeminência inicial. Para o tempo de leitura, 10% a mais de duração na imitação de cor com relação ao neutro e até 31% a mais na imitação consecutiva.

Já para o francês, os resultados significativos revelaram uma proporção de proeminência inicial de cerca de 50% no estilo neutro contra 65% nos dois estilos de imitação. Quanto à F0, encontramos uma mediana maior de 3 semitons nos estilos de imitação nos homens com

8 Medida não paramétrica do desvio-padrão, definida como a metade da diferença entre os quartis 1 e 3.

9 Esses testes serão apresentados no capítulo 6, seção 6.1.

relação ao neutro e do mesmo montante no estilo de imitação consecutiva em relação ao neutro, mas apenas para uma das jornalistas. A amplitude média de variação de F0 é maior em 3 semitons nos estilos de imitação, mas apenas para a mesma jornalista que teve mediana de F0 maior na imitação consecutiva. Houve 32% a mais de duração em dois jornalistas no estilo de imitação consecutiva, lentificando assim a fala. Esse resultado é inesperado, uma vez que a jornalista cujo modo de falar lhes foi apresentada como modelo para imitar fala muito rapidamente. Nas duas línguas houve 2 a 3 dB a mais de ênfase espectral nas imitações em relação ao neutro, sinalizando maior esforço vocal na imitação independentemente de língua.

Concluímos com esse trabalho que a proeminência inicial é um traço importante do estilo imitado nas duas línguas, sendo sinalizado também por maior valor de mediana de F0 e amplitude de variação, isto é, mais agudo e mais variável. A maior diferença entre as línguas diz respeito ao estilo de imitação consecutiva por conta das particularidades de cada jornalista: a francesa que falou rapidamente e a brasileira que falou lentamente (essa escolha não foi, em princípio proposital, nem muito menos caracteriza a locução de todo jornalista nesses países). As hipóteses (1) e (2) foram confirmadas, mas não a hipótese (3), pois não observamos nenhuma convergência no estilo imitado consecutivamente em francês. Quanto à hipótese (4), as diferenças foram mais relacionadas aos participantes ou gêneros e não tanto às línguas.

3.2.7 Protocolo Experimental: Variação da Taxa de Elocução

Em um dos experimentos que compuseram nosso trabalho sobre o ritmo da fala (BARBOSA, 2006), utilizamos uma passagem do livro "A Menina do Nariz Arrebitado" de Monteiro Lobato, lida por quatro participantes masculinos paulistas em três taxas de elocução para es-

tudar como se modificam as durações de sílabas fonéticas e grupos acentuais no PB sob a demanda de falar mais lentamente ou mais rapidamente.

Para esse corpus, as taxas de elocução foram eliciadas por instruções dadas pelo experimentador. Esse solicitou a cada participante que começasse a ler a passagem com uma taxa de conversação confortável (taxa normal). Em seguida, que lesse o mais lentamente possível (taxa lenta), mas preservando o sentido dos enunciados e assim evitando o estilo de ditado e, finalmente, que lesse o mais rapidamente possível sem cometer lapsos (taxa rápida). É claro que, apesar das instruções, não há impedimento de que taxas estatisticamente iguais fossem produzidas, o que, de fato, ocorreu depois da constatação de diferença com um teste de ANOVA. Para evitar isso, pode-se dar como modelo para escuta prévia um áudio de fala natural ou sintetizada em que haja taxas de elocução estatisticamente distintas para serem reproduzidas. Utilizamos esse procedimento no trabalho de doutorado (BARBOSA, 1994) obtendo cinco taxas de elocução distintas por participante. Não usamos esse procedimento aqui para garantir um certo conforto nas produções. O excerto de Monteiro Lobato segue abaixo.

Em seguida apareceu um papagaio real que tinha fama de orador. Subiu a tribuna de um poleiro de ouro e fez um belo discurso a respeito da arte de falar. Nesse discurso provou que os homens tinham aprendido a falar com os papagaios, e não os papagaios com os homens, como diz a ciência destes. Uma chuva de palmas acolheu suas palavras.

O mesmo não aconteceu, porém, com a poetisa Lagartixa, que principiou a recitar uma longa poesia e engasgou no meio, acabando o recitativo em choro e faniquito. Para destruir essa má impressão vieram três vagalumes mágicos que fizeram várias sortes, sendo muito apreciada a sorte de comer fogo.

Foram segmentadas todas as sílabas fonéticas (unidades VV) por detecção automática de início de vogal com o script *Beat Extrac- tor* (BARBOSA, 2006) seguido de correção manual e, em seguida, com o script *SG Detector*, foram obtidos os picos de duração normalizada considerando todo pico local como fronteira à direita de grupo acentual (para aprender como obter duração normalizada veja o capítulo seguinte). Com isso pudemos obter, por participante e por taxa, os valores das durações de unidades VV e dos grupos acentuais com os quais calculamos média e desvio-padrão, mostrados nas Tabelas 3.3 e 3.4.

A Tabela 3.3 revela uma extensão de médias de duração das uni- dades VV de 152 ms a 283 ms (correspondentes respectivamente a 6,6 e 3,5 unidades VV/s de taxa de elocução). O intervalo entre os per- centis 5% a 95% é de 95 ms a 570 ms para AP, o locutor mais lento, e de 87 ms a 292 ms para FA, o locutor mais rápido, revelando valores mínimos semelhantes e valores máximos de praticamente o dobro para AP em relação a FA. Isso representa uma grande variabilidade para o alongamento silábico entre diferentes locutores do PB e diz respeito a variações individuais.

Tabela 3.3 – Valores médios (e desvios-padrão) em milissegundos da duração das unidades VV no *corpus Lobato* em quatro participantes paulistas masculinos, para três taxas de elocução.

	Participante			
Taxa de Elocução	AP	AC	DP	FA
Lenta	283 (185)	243 (203)	190 (154)	189 (111)
Normal	235 (169)	223 (166)	188 (165)	165 (88)
Rápida	201 (144)	194 (138)	165 (119)	152 (74)

Tabela 3.4 – Valores médios (e desvios-padrão) em milissegundos da duração dos grupos acentuais no *corpus Lobato* em quatro participantes paulistas masculinos, para três taxas de elocução.

	Participante			
Taxa de Elocução	AP	AC	DP	FA
Lenta	1504 (694)	1518 (562)	1107 (527)	1154 (389)
Normal	1370 (496)	1353 (521)	1233 (588)	931 (363)
Rápida	1180 (543)	1348 (525)	1077 (546)	889 (395)

A Tabela 3.4 revela uma extensão de médias de duração de grupo acentual de cerca de 1 s a 1,5 s para diferentes taxas, revelando uma certa tendência em realizar um acento frasal a uma cadência semelhante. Grosso modo, essa ordem de grandeza é da leitura de um hemistíquio de um verso alexandrino, verso de 12 sílabas. Pode-se calcular das duas tabelas, de fato, que o valor médio do número de unidades VV em cada grupo acentual para os quatro locutores é de 6,5 unidades VV, muito próximo ao número de sílabas do hemistíquio, que é de 6. Esses aspectos revelam o caráter universal da sucessão de proeminências na fala.

3.2.8 Protocolo Experimental: Leitura e Narrativa

Há alguns anos gravamos o corpus Belém com o fim de cotejar a prosódia da leitura e da narrativa consecutiva de um texto sobre a origem dos pastéis de nata de Belém[10]. O trabalho que detalhamos aqui, ressaltando o seu protocolo experimental, é o de Barbosa e Silva (2012), trabalho que procurou responder duas perguntas: o que faz com que dois enunciados sejam percebidos como prosodicamente distintos e que parâmetros contribuem para que dois enunciados difiram no

10 O texto foi proposto pelo INESC de Lisboa quando de uma colaboração empreendida em 2009. O texto original, em português europeu, narra a história dos pastéis de Belém. A mesma equipe solicitou na época a um brasileiro que adaptasse o texto para o português brasileiro e foi a versão adaptada que usamos nos experimentos com o PB.

modo de falar?

Considerando que as duas funções básicas da prosódia são a marca- ção de fronteiras durante a fala, bem como o assinalamento de unidades proeminentes, nos propusemos a examinar as diferenças produzidas e percebidas relacionadas a essas duas funções. No trabalho de Barbosa e Silva (2012), examinamos unicamente parâmetros temporais, muito embora um deles diga respeito à entoação stricto sensu, pois medimos a taxa de acentos de *pitch* por segundo.

Para avaliar diferenças rítmicas, comparamos a fala lida e a narrada, por suas características distintas. A escolha da fala narrada se deve ao fato de ela apresentar elementos comuns com a conversa espontânea, um estilo de elocução muito frequente nas instâncias comunicativas. Alguns desses elementos são as hesitações causadas pelo macro e microplanejamento do discurso, que dizem respeito respectivamente ao conteúdo e organização sintática do que se vai dizer (LEVELT, 1989). Embora haja hesitações na fala lida, essas são bem menos frequentes do que na narracão de uma história lida, devido à maior demanda para a memória de trabalho[11].

O corpus usado foi formado pela leitura e pela narração consecutiva do texto dos pastéis de Belém adaptado para o PB. O texto tem cerca de 1600 palavras e foi, na época, lido por duas mulheres e um homem, sendo os três estudantes do curso de Linguística com cerca de 30 a 45 anos. Imediatamente após sua leitura, a história foi narrada pelos três participantes com suas próprias palavras.

O corpus foi primeiramente segmentado em excertos de 9 a 18 se- gundos por conta dos testes de percepção que foram feitos e para a avaliação do vínculo entre produção e percepção da prosódia. A esco- lha dessa extensão é fundamentada em testes anteriores com trechos mais curtos e na literatura, pois é necessário um trecho mais longo para que o ouvinte infira o modo de falar de uma pessoa. Cada

11 Trata-se do mecanismo cognitivo para reter informações enquanto fazemos uma tarefa. Ver (COWAN, 1997) para detalhes.

excerto foi segmentado em sílabas fonéticas cujas fronteiras foram definidas por inícios de vogais, como repetidamente explicado na literatura sobre prosódia (LEHISTE, 1970; CLASSE, 1939; BARBOSA, 2006, 2019). Essa segmentação foi feita semi-automaticamente em camada de anotação do software Praat em duas etapas. Para a primeira etapa usamos o script Beat Extractor para a detecção automática dos inícios de vo- gal e, na segunda etapa, corrigimos os erros de detecção manualmente, introduzindo manualmente a transcrição fonética para possibilitar a normalização da duração silábica.

A partir da segmentação realizada, um script feito para esse trabalho calculou 10 parâmetros prosódico-acústicos, entre eles: a taxa de elocução em unidades VV por segundo, os três primeiros descritores estatísticos (média, desvio-padrão e assimetria da distribuição) e a taxa por segundo dos picos de duração normalizada ao longo dos excertos, a taxa de picos da curva de Fo suavizada por segundo[12], os coeficientes de variação (desvio-padrão dividido pela média) da duração do grupo acentual, do número de unidade VV por grupo acentual e da duração da unidade VV e, por fim, a taxa de unidades VV não salientes (isto é, não são picos locais de duração normalizada).

Os excertos analisados foram associados em pares de áudio separados pelo áudio de um tom puro (de frequência de 1000 Hz) para a confecção de um teste de discriminação no Praat do qual participaram 10 estudantes de Linguística. A finalidade do tom puro é apenas assinalar a fronteira entre os dois áudios a serem avaliados quanto à diferença rítmica. Em seu conjunto, os excertos foram formados pela narrativa de uma das participantes e pela leitura dos outros. Foram utilizados 44 pares de excertos combinados aleatoriamente para o teste, que durou até cerca de 25 minutos para ser completado.

Cada excerto foi também deslexicalizado (vide seção 3.2.2 sobre

12 Essa taxa foi obtida dividindo o número de picos de F0 suavizado num trecho por sua duração. Esses picos são máximos locais da curva suavizada de F0 com a frequência de corte de 5 Hz. Em seguida interpola-se a curva e conta-se automaticamente o número de picos no excerto.

deslexicalização) usando o algoritmo desenvolvido por Vainio et al. (2009). Assim, cada par de excertos foi combinado na ordem AB e BA tanto na versão original quanto deslexicalizada (cada participante ouviu os áudios nas duas ordens, o que foi necessário para avaliar a consistência das respostas). A inversão da ordem permite avaliar o grau de consistência das respostas dos ouvintes, pois, sendo o mesmo par avaliado, espera-se que a resposta quanto ao grau de discriminação seja a mesma. Cada ouvinte avaliou primeiramente a versão deslexicalizada e depois, em ordem aleatória, a versão original a partir da seguinte instrução: "Avalie quão diferente é o modo de falar dos trechos de áudio no par separados por um tom numa escala de 1 a 5, sendo 1 'mesmo modo de falar' e 5 'modos de falar completamente diferentes', usando qualquer nível entre os dois a partir da sua percepção." O tom usado foi um sinal senoidal com amplitude descendente de 1000 Hz de cerca de 30 ms. As respostas de 1 a 5 foram posteriormente recodificadas linearmente de -1 a 1, sendo 0 considerado uma resposta neutra.

Quanto ao desempenho dos ouvintes, duas hipóteses foram consideradas: (1) que a consistência nas respostas seria maior na fala deslexicalizada e que haveria melhor desempenho para diferenças entre estilos diferentes, tendo em vista maior atenção na prosódia por conta da deslexicalização; (2) que a diferença de valores médios de pelo menos um parâmetro acústico seria capaz de prever as respostas dos ouvintes, por conta do esperado vínculo entre produção e percepção da prosódia.

Quanto à consistência das respostas, a média das respostas no mesmo par de excertos em diferentes ordens foi menor e menos va- riável para a versão original (diferença entre médias com $t_{gl=398} = 4,2$, $p < 10^{-4}$), contradizendo a primeira hipótese. Isto é, a informação lexi-cal ajudou na manutenção da mesma resposta para o par apresentado em ordem distinta. Quanto à resposta ao grau de diferença no modo de falar, não há diferença alguma quer se use a versão deslexicalizada quer se use a versão original.

Para avaliar a segunda hipótese, tomamos apenas pares de excertos com respostas com consistência inferior a 0,5 (a consistência foi definida como a diferença das respostas nas duas ordens de apresentação que teoricamente deveria ser 0) e com desvio-padrão entre respostas de diferentes ouvintes também menor do que 0,5, com o fim de utilizar apenas os 15 pares com respostas homogêneas e consistentes que permitissem uma melhor avaliação de sua relação com os parâmetros prosódico-acústicos. Usamos testes de regressão linear múltipla[13] para prever a resposta média dos ouvintes para cada par a partir da diferença dos valores médios dos 10 parâmetros acústicos extraídos dos excertos em cada par.

O melhor modelo explicou 71% da variância das respostas dos ouvintes (resp)[14]: $resp = -1,5 + 10,4.pr + 2,65.sr - 10,75.pr \times sr$ em que *sr* é a taxa de elocução e *pr* é a taxa de picos de duração normalizada, isto é, respectivamente um parâmetro relacionado à sucessão de sílabas fonéticas e outro relacionado à sucessão de sílabas fonéticas proeminentes. Esse resultado confirma a segunda hipótese e aponta para o papel crucial da taxa de elocução e da taxa de produção de sílabas proeminentes para a percepção do modo de falar.

3.2.9 Testes de Percepção da Prosódia

Conforme acabamos de ver, os testes de percepção são muito úteis para avaliar a relação entre produção e percepção da prosódia. Além de testes de discriminação que discutiremos mais detalhadamente aqui, há também os testes de classificação. Se o teste de discriminação requer, cognitivamente, uma avaliação de elementos presentes na memória de trabalho, o teste de classificação faz uso da memória de longo

13 Este teste avalia a correlação entre um conjunto de variáveis preditoras e uma variável predita.

14 Valor de p de pelo menos 0, 009 para todos os coeficientes da regressão ($F_{3,11} = 12,4$, $p < 0,0008$).

termo, pois requer a associação de um estímulo a uma classe que nos é conhecida em menor ou maior grau, em função de nossa experiência comunicativa.

Perceber elementos na fala envolve a capacidade de avaliar semelhanças e diferenças entre estímulos, evocando eventualmente duas categorias de memória. A percepção pode assim se dar de duas maneiras: (1) pela comparação da informação acústica armazenada temporariamente na memória de trabalho para os dois estímulos[15] ou (2) pela comparação da informação acústica do estímulo que se ouve com elementos de alguma categoria construída e armazenada na memória de longo termo para o estímulo que se ouviu anteriormente. Por conta disso, é necessário falar de categorias também na investigação em prosódia experimental.

Pelo relato testemunhal de Repp (1984), a pesquisa sobre percepção categórica na fala começou nos Haskins Laboratories depois da construção do primeiro sintetizador de fala, o *Pattern Playback* com o trabalho de Liberman et al. (1957), que criaram um contínuo acústico de sílabas sintéticas do tipo /Ce/ em que C = /b d g/. O desempenho de teste de discriminação do tipo ABX (ouvem-se os três estímulos e é preciso responder se X é A ou B) revelou que os ouvintes tinham mais facili- dade em responder quando os estímulos proviam claramente de duas categorias distintas do que quando provinham da mesma categoria. Por exemplo, se dois estímulos distintos eram exemplos de /be/, foi mais difícil discriminá-los do que se um era exemplo de /be/ e outro de /de/. Essas categorias de um dos três fonemas foram avaliadas antes com um teste de classificação. Sendo assim, propunha-se que o desempenho dos ouvintes no teste de classificação poderia prever seu desempenho no teste de discriminação. Essa relação estreita entre os desempenhos nas duas tarefas sempre foi considerado

15 A memória de curto termo retém informação acústica por cerca de no máximo 500 ms, mas essa informação pode ser categorizada de alguma forma e permane- cer na memória de trabalho até seu limite temporal, que é de cerca de 20 segundos (COWAN, 1997).

como necessária para se verificar se houve ou não percepção categórica. No entanto, essa necessidade foi contestada mais de uma vez em trabalhos como os de Pollack e Pisoni (1971), Schouten, Gerrits e Hessen (2003), Gerrits e Schouten (2004), que mostraram que além de não ser estreita a relação dos desempenhos nos dois testes, o desempenho do participante depende do tipo de teste, sendo ABX apenas um deles. Para um estudo de outros tipos de teste, que não consideraremos neste livro, ver também a tese de Gerrits (2001) e as excelentes recomendações e apanhado geral dos testes mais úteis no relatório de McGuire (2010).

3.2.9.1 Testes de Discriminação

O teste de discriminação pode ser feito em diversos paradigmas (ABX, AX, AXB, 2IFC, 4IAX and 4I-oddity), mas todos envolvem responder a uma pergunta sobre a similaridade ou dissimilaridade entre estímulos. O mais usado na área de prosódia é o teste do tipo AX, em que se pergunta se o segundo estímulo (X) é igual ou distinto do primeiro (A). Vários pares de estímulos como esses são habitualmente apresentados para os ouvintes e, como vimos, é preciso montar o experimento numa plataforma que possibilite a aleatorização dos estímulos, o controle do tempo entre a resposta dada e o próximo estímulo, o tempo para dar a resposta desde a apresentação do estímulo (tempo de reação), a inclusão de estímulos distratores que dificultem ao ouvinte inferir os objetivos do experimento e ter um comportamento enviesado, entre eles número de falsos alarmes[16] em demasia, entre outras coisas. Várias plataformas estão disponíveis gratuitamente, entre elas o Praat.

Para testes de discriminação de trechos de fala para estudos

16 Um falso alarme é identificar um fenômeno ou estímulo que não pertence a uma categoria como dessa categoria. Assim, por exemplo, se é solicitado a identificar fronteiras prosódicas num enunciado, o acerto é quando uma fronteira de fato é identificada como tal e um falso alarme uma posição em que não tem fronteira identificada como sendo de fronteira.

prosódicos recomenda-se que a extensão desses esteja entre 10 e 20 segundos, pois valores inferiores a 10 segundos não permitem adequada avaliação de um modo de falar e valores superiores a 20 segundos estão em geral além do limiar temporal da memória de trabalho para dados acústicos.

Além do exemplo da seção 3.2.8, comparamos trechos de fala de estilos jornalístico e político em quatro línguas (BARBOSA; MADUREIRA; MAREÜIL, 2017): português brasileiro e europeu, francês da França e alemão da Alemanha, além de leitura e narrativas de não profissionais. Como os ouvintes foram brasileiros com português nativo, foi necessário deslexicalizar os trechos de fala política e jornalística, para que a discriminação não se desse pelo reconhecimento de quem falava. Esse trabalho revelou uma discriminação entre os estilos profissionais, sendo o discurso político o mais facilmente reconhecido nas quatro lín- guas. Além do teste de discriminação, um dos experimentos realizados envolveu um teste de classificação.

3.2.9.2 Testes de Classificação

Os testes de classificação ou identificação possibilitam saber se o ouvinte é capaz de classificar um estímulo dentro de um conjunto fechado ou aberto de possibilidades. Se o conjunto for fechado, o teste se denomina teste de classificação de escolha forçada. Por exemplo, no caso de identificação de estilos de elocução entre três possibilidades: sermão religioso, discurso político ou fala telejornalística, o ouvinte é convidado a ouvir um estímulo e escolher imediatamente depois entre uma dessas três possibilidades.

Se uma das opções de resposta permite que o ouvinte diga que não é nenhum dos estilos propostos, tem-se um teste de classificação de escolha não forçada que permitiria avaliar a ambiguidade de determinados estímulos. A esse respeito, vale a pena incluir no teste de

classificação uma avaliação do quão típico representante da categoria escolhida é aquele estímulo, uma avaliação denominada em inglês *goodness of fit* (qualidade de ajuste). O resultado do teste permite ao experimentador uma reavaliação dos estímulos considerados maus representantes de sua classe.

Algo semelhante ao julgamento de qualidade de ajuste é a avaliação de alguma grandeza perceptiva numa determinada escala que avalie aspectos como "quão enfático soa a palavra x" numa escala de 1 a 5, "quão agudo soa a palavra x" numa escala gradativa de nada agudo a extremamente agudo, "quão agradável soa este enunciado" numa escala de 5 pontos como "nada agradável", "pouco desagradável", "nem agradável nem desagradável", "agradável" e "muito agradável". Há vários modos de pedir a um ouvinte para gradar uma qualidade para fins de investigação prosódica com algumas propostas que fazem um apanhado de escalas perceptivas no trabalho de Rietveld e Chen (2006, p. 286- 302). A importância de avaliar os resultados desse tipo de avaliação é a determinação de quais parâmetros acústicos expressariam uma determinada qualidade perceptiva. Se qualidades como "agudo" trazem imediatamente à lembrança o valor médio da Fo; outras, como "enfático", sugerem maior intensidade, Fo e duração, outras ainda, como "agradável", não são clara e decisivamente associadas a um parâmetro prosódico-acústico.

Outro aspecto fundamental para tirar o máximo proveito dos resultados de um teste de percepção é a análise da resposta dos participantes, avaliando sua sensibilidade ao teste, bem como a coerência entre suas respostas. A sensibilidade a um teste leva em conta não apenas a taxa de respostas de um determinado tipo, como também o viés de resposta de um participante. Suponhamos que se queira verificar se um trecho de fala telejornalística é, de fato, percebido como tal e se faça um teste de classificação. Suponha-se ainda que, entre os estímulos, haja cerca de 60% de estímulos de fala de um estilo de elocução bem distinto como distratores. Se um dos participantes tiver

respondido a todos os estímulos, nos dois estilos, como sendo todos de fala telejornalística, esse participante teria 100% de "acerto" para o estilo telejornalístico, embora com 60% de falsos alarmes, pois teria dito que todos os distratores são do estilo telejornalístico. Esse participante teria introduzido um viés, não tendo sensibilidade ao teste. Para corrigir esse tipo de resultado, é necessário considerar acertos e falsos alarmes, o que é proposto na Teoria da Detecção (GREEN; SWETS, 1966; MACMILLAN; CREELMAN, 2004) com o conceito de d' (d linha). Essa grandeza é definida pela equação 3.2 e mede uma resposta "líquida", isto é, que considera a diferença em unidades de *z-score*[17] entre proporções de acerto e de falso alarme.

$$d' = z(p_{acertos}) - z(p_{\text{falsos alarmes}}) \tag{3.2}$$

Sendo assim, um participante que respondesse a um estímulo com proporções idênticas de acerto e falso alarme teria um d' nulo, enquanto nosso participante hipotético, que tem 60% de falsos alarmes e 40% de acertos teria um d' negativo, mais precisamente de -0, 5. Valores de d' que considerem uma sensibilidade razoável a um estímulo devem ser pelo menos maiores ou iguais a 1. Participantes com valores de d' nulos e negativos devem, em princípio, terem seus resultados desconsiderados num determinado experimento, pois não foram sensíveis ao teste.

Outro aspecto a se considerar quanto aos resultados de um teste de percepção é a coerência de respostas dos participantes, se aplicável. É evidente que, num teste de identificação de fronteiras prosódicas, por exemplo, determinadas fronteiras fracas não são percebidas por todos os participantes e isso é um fato da percepção, pois a não coincidência das respostas não é um problema, é uma informação importan-

17 Medida normalizada de um valor na distribuição gaussiana, definido com arazão entre a diferença de um valor e a média da distribuição pelo desvio-padrão.

te sobre a saliência dessa fronteira.

Por outro lado, nos casos em que se quer determinar as características prosódico-acústicas de um estilo de elocução, de uma atitude ou de uma emoção, por exemplo, é necessário validar perceptivamente esses estímulos. A validação consiste em saber se veiculam, de fato, para os ouvintes aquele estilo, aquela atitude ou aquela emoção. Para tanto, testes de classificação devem ser feitos e, em seguida, deve-se medir a consistência dos participantes em suas respostas, a fim de saber se respondem não aleatoriamente e da mesma forma a um mesmo estímulo.

O teste estatístico que mede essa coerência é o teste de coerência entre juízes, proposto por Cohen para dois juízes e por Fleiss para várias juízes, como extensão do teste de Cohen. Recomendamos ao leitor que se inteire em livros como o de Rietveld e Hout (1993) sobre esse tipo de teste.

3.3 Prelúdio para o Próximo Capítulo

Tendo visto os aspectos metodológicos mais importantes da área de pesquisa em prosódia experimental, vamos apresentar as técnicas de medida de parâmetros prosódico-acústicos que permitirão ao leitor fazer seus próprios experimentos. Os aspectos vistos aqui serão retomados adiante, com estudos de caso e apresentação sucinta de técnicas para análise estatística inferencial.