

ANÁLISE EXPERIMENTAL DO ACENTO

Neste capítulo será apresentada uma análise experimental do acento que busca dar suporte empírico à análise teórica apresentada nos capítulos anteriores. O capítulo se divide em três partes. A primeira apresentará considerações a respeito da metodologia experimental adotada. A segunda parte será reservada ao experimento de produção de fala; a terceira, aos três experimentos de percepção. Em cada uma das últimas partes, para cada experimento, iremos apresentar a metodologia adotada na coleta de dados (objetivos, hipóteses, estímulos, participantes, procedimentos), seguida dos resultados obtidos (distribuição dos dados, análise estatística, discussão dos resultados).

A análise e modelagem estatística dos dados dos quatro experimentos foi realizada por meio do R, uma linguagem e ambiente computacionais voltados para análise estatística e representação gráfica (R Development Core Team, 2011). Para maior intercâmbio de informações, os modelos testados na análise estatística serão apresentados diretamente na notação do R. O Apêndice B contém uma breve explicação de tal notação.

O protocolo de coleta de dados adotado nesta pesquisa foi aprovado pelo Comitê de Ética em Pesquisa – COEP, UFMG.

6.1 A METODOLOGIA EXPERIMENTAL NO ESTUDO DO ACENTO

Um dos objetivos deste estudo é apresentar uma análise experimental do acento lexical do português brasileiro, seguindo a abordagem proposta pela Fonologia de Laboratório (PIERREHUMBERT; BECKMAN; LADD, 2000). A Fonologia de Laboratório surgiu na década de 80 como uma proposta inovadora de utilizar métodos de laboratório para analisar a estrutura sonora da linguagem humana. Contribuiu para o estabelecimento de um compromisso entre fonética-fonologia – que até então eram áreas de pesquisa independentes –, propondo uma reestruturação de suas tarefas e objetos de estudos. Essa metodologia de pesquisa tem caráter interdisciplinar, reunindo pesquisadores com trajetórias acadêmicas variadas, comprometidos com modelos teóricos diversos e, não raro, em desacordo. Como destacam Pierrehumbert, Beckman e Ladd (2000), a Fonologia de Laboratório não se identifica com um modelo teórico específico, mas se define antes por suas opções metodológicas. Partindo da visão da linguagem humana como um fenômeno natural – e consequentemente da fonologia como uma ciência natural –, a Fonologia de Laboratório buscou fortalecer as bases científicas da fonologia, por meio da utilização de dados empíricos, do aprimoramento metodológico, da modelagem explícita e da acumulação de resultados.

A literatura sobre o acento no PB é composta predominantemente por estudos fonológicos tradicionais, como mostramos na seção anterior. Em outras línguas, o acento vem, cada vez mais, sendo examinado sob um ponto de vista que conjuga a análise experimental com reflexões sobre a organização gramatical (e.g. Cummins; PORT, 1997 e no inglês; Mooshammer; Fuchs, 2002, no alemão; Ortega-Llebaria; Prieto, 2007, no espanhol; Sluijter; Heuven, 1996, no holandês). No PB, podemos destacar os trabalhos de Barbosa (2006) e Moraes (1998) que têm conduzido estudos experimentais abrangentes sobre aspectos prosódicos do PB, abrindo caminho para pesquisas recentes que abordam o acento de forma experimental, e.g. Arantes (2010), Consoni (2006), Meireles (2007).

A análise experimental realizada neste estudo se baseou em experimentos de produção e de percepção de fala cujos resultados pudessem contribuir:

- c) com uma compreensão mais ampla dos correlatos acústicos e articulatórios do acento e sua interação com aspectos prosódicos (experimento de produção);
- d) com o debate sobre a organização e o processamento mental das informações acentuais (experimento de percepção).

6.2 EXPERIMENTO DE PRODUÇÃO

O experimento de produção teve como objetivo avaliar as propriedades acústicas e articulatórias envolvidas na realização do acento. Pretende-se testar a hipótese de que parâmetros acústicos e articulatórios produzidos nas sílabas tônicas são diferentes das sílabas pretônicas e postônicas.

6.2.1 Participantes

Foram avaliados neste experimento cinco participantes, todos do sexo masculino, com nível universitário de educação formal, na faixa de 20 a 30 anos. Todos são falantes nativos do PB na variedade de Minas Gerais, não tendo se ausentado da região por mais de um ano. Nenhum deles relatou histórico de problemas de fala ou audição. Foram privilegiados participantes do sexo masculino, com pouca gordura e pelo facial na região do pescoço, para obter sinal eletroglotográfico de melhor qualidade, seguindo recomendação de Vieira (1997).

6.2.2 Procedimentos

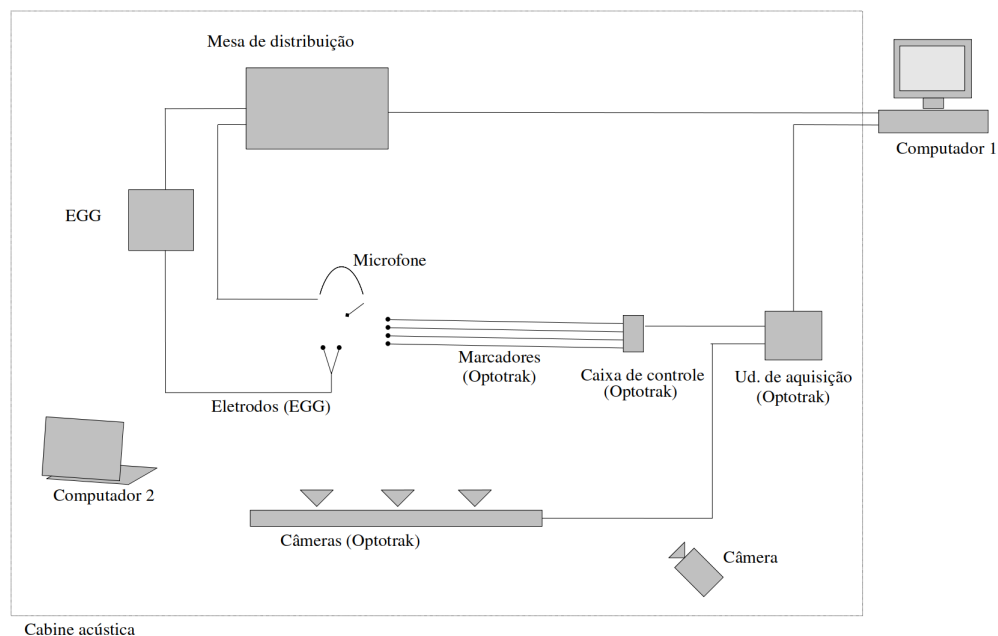
O experimento consistiu na gravação da fala, simultaneamente à aquisição de sinal glótico e captura de movimentação da face. Para captação do sinal de fala, foi utilizado microfone de cabeça cardioide, modelo Sennheiser HSP 4. A aquisição do sinal foi realizada em um computador, por meio do *software* Audacity (AUDACITY Team, 2008), a uma taxa de amostragem de 44100Hz e a 16 bits por amostra, adequando-se à extensão em frequência dos eventos acústicos da fala.

Para captação do sinal glótico, foi utilizado um eletroglotógrafo (EGG) de dois canais, modelo EG-2 da Glottal Enterprises. O EGG consiste basicamente em dois eletrodos metálicos de contato posicionados na altura da cartilagem tireoide que medem a variação de impedância elétrica transversal resultante da abertura e fechamento das cordas vocais (VIEIRA, 1997). Quando as cordas vocais se fecham, a impedância cai, como resultado de diferenças na condutância do ar e dos tecidos da região tireoide (ROTHENBERG; MASHIE, 1988). O sinal do EGG representa a variação da área e da extensão longitudinal do contato das cordas vocais ao longo do tempo (KRISHNAMURTHY, 1983). O sinal obtido pelo EGG foi adquirido no mesmo computador e programa que o sinal de fala – cada um ocupando um canal de uma faixa estéreo.

A gravação foi realizada em cabine acústica do CEFALA (Centro de Estudos da Fala, Acústica, Linguagem e Música, <http://www.cefala.org>), na Escola de Engenharia da UFMG, utilizando simultaneamente três tipos de equipamento de coleta de dados: microfone, eletroglotógrafo (EGG) e sistema de captura de movimento Optotrak¹, como mostra o esboço do ambiente experimental mostrado na FIG. 14, a seguir. A aquisição dos sinais do EGG e da fala foi realizada no computador 1. O sinal da fala também foi capturado pelo sistema Optotrak, em módulo de aquisição próprio, simultaneamente à aquisição do movimento dos marcadores, unicamente para posterior sincronização entre sinal de EGG, sinal da fala, posição dos marcadores.

¹ Por limitações de tempo, a avaliação dos dados do Optotrak será tema de trabalho futuro. A análise dos dados de movimentação da face poderá contribuir para a compreensão dos mecanismos de coordenação de gestos e sua relação com as propriedades acústicas envolvidas na proeminência acentual.

FIGURA 14 – Projeto experimental utilizado na coleta dos dados de produção



O EGG é composto por um par de eletrodos, ligados a uma unidade de processamento. O microfone capta o som da fala. O equipamento de captura de movimento Optotrak é formado por um conjunto de três câmeras alinhadas, que captam o deslocamento tridimensional dos marcadores, estes, por sua vez, conectados a uma caixa de controle. A unidade de aquisição do Optotrak controla os marcadores por meio da caixa de controle, recebe o sinal das câmeras e gera sequências de coordenadas em três eixos ao longo do tempo. Como mostra a FIG. 14, no projeto experimental elaborado, os sinais do EGG e do microfone passam por uma mesa de distribuição, sendo, juntos, transmitidos no formato estéreo, cada um em um canal, para o computador 1, posicionado do lado de fora da cabine, para minimizar ruídos. A mesa de distribuição foi utilizada para sincronização dos sinais do EGG e da fala. O sinal do sistema Optotrak também é enviado para o computador 1, responsável pela sincronização do sinal do Optotrak com o sinal da mesa de distribuição. Os três sinais (microfone, EGG e Optotrak) foram adquiridos simultaneamente e com mesma taxa de amostragem (44100 Hz), correspondendo, portanto, aos mesmos eventos da produção oral dos participantes. O computador 2, portátil, foi utilizado unicamente para apresentar os estímulos a serem produzidos pelo participante. O experimento foi documentado em vídeo pela gravação com câmera digital.

Antes de iniciar o experimento, solicitava-se aos participantes que se sentassem em uma cadeira, localizada no centro da sala. Os eletrodos do EGG, presos, com uma tira, na região central do pescoço, seguindo os sinalizadores de bom posicionamento do próprio equipamento; os eletrodos eram previamente limpos e empregado gel condutivo, para melhorar a captura do sinal. Finalmente, era posicionado o microfone de cabeça, com encaixe ao redor das orelhas e a cápsula a 2 cm de distância do canto esquerdo da boca.

6.2.3 Estímulos

Os estímulos foram construídos de forma a possibilitar um controle rígido das variáveis independentes. Foram selecionadas quatro variáveis independentes, que passaremos a descrever.

As sílabas, ao serem comparadas com outras sílabas em uma mesma palavra, são a unidade mais relevante para o estudo do acento. Foram testadas quatro sílabas diferentes, níveis da variável que referiremos como “sílabas”: [ga], [pa], [pi] e [ku]. A escolha dos sons para formar as sílabas-teste obedeceu primeiramente a dois critérios:

- a) apenas vogais [a, i, u], visto que as médias [ɛ, e, ɔ, o] apresentam restrição de ocorrência em determinados ambientes (cf. SILVA, 2002);
- b) apenas consoantes oclusivas, para facilitar a segmentação – a oclusão é formada principalmente por um momento marcado em que cessa a emissão de energia acústica, resultando em uma barra vertical branca no espectrograma, o que torna a tarefa de segmentação mais acurada (LADEFOGED, 1996).

As quatro sílabas [ga], [pa], [pi] e [ku] foram especificamente selecionadas por contarem com maior número de palavras em conformidade com os critérios adotados para controle das sentenças, que serão explicitados mais adiante. Além da variável “sílabas”, outra variável independente adotada foi o tipo de relação de proeminência estabelecida entre a sílaba-teste e a sílaba acentuada da palavra. A variável “tipo” contém oito níveis:

- a) sílaba acentuada final (T1)
- b) sílaba acentuada não final seguida por uma postônica (T2)
- c) sílaba acentuada não final seguida por duas postônicas (T3)
- d) sílaba postônica final única na palavra (O4)
- e) sílaba postônica final antecedia por postônica medial (O5)
- f) sílaba postônica medial (O6)
- g) sílaba pretônica adjacente à acentuada (E7)
- h) sílaba pretônica anterior à sílaba adjacente à acentuada (E8)

A variável “tipo” foi assim construída de forma que os níveis pudessem ser facilmente agrupados quanto à relação de proeminência da sílaba teste: acentuada (T), átona postônica (O) e átona pretônica (E). Os oito níveis da variável “tipo” foram avaliados individualmente nos oito níveis listados acima, assim como agrupados pela relação de proeminência em apenas três níveis, correspondentes aos níveis marcados com T, E e O.

Conjugando as duas variáveis independentes, foram geradas 32 sílabas-teste diferentes, correspondentes a quatro níveis da variável “sílabas” (ga, pa, pi, ku) x oito níveis da variável “tipo” (T1, T2, T3, O4, O5, O6, E7, E8). A partir das 32 sílabas-teste, foram construídas 32 palavras, por sua vez, inseridas em 32 sentenças controladas, que foram usadas como estímulos no experimento. A leitura de sentenças, ao invés de palavras isoladas, permite maior controle da estrutura prosódica dos estímulos. Ademais, procurou-se construir sentenças da maneira mais natural possível, de forma a reduzir a artificialidade ocasionada pela experimentação em laboratório. Por essa razão, foi evitado especialmente o uso de sentenças-veículo – paradigma em que todas as palavras-teste são inseridas em uma mesma sentença –, sendo adotado em seu lugar um paradigma de sentenças controladas – ou seja, sentenças variadas, adequadas semanticamente à palavra-teste, mas controladas no nível prosódico e segmental. Na construção das sentenças, o ambiente sonoro anterior e seguinte às sílabas-teste foi controlado sistematicamente, como mostra o QUADRO 3, abaixo. Sílabas acentuadas encontram-se em negrito e sílabas-teste, sublinhadas. Asteriscos indicam logatomas de estrutura fonotática provável, criados para preencher lacunas no quadro quando não puderam ser encontradas palavras que seguissem o controle do contexto sonoro pretendido.

QUADRO 3 – Sílabas testadas e seu contexto adjacente conforme as variáveis sílabas e tipo

Tipo/Sílaba	[ga]	[pa]	[pi]	[ku]
T1	avô <u>gagá</u> tagarela	vatapá garantido	Pipi divertida	Pirarucu buliçoso
T2	aquela <u>gata</u> malhada	conta <u>paga</u>	Zipide*	outro <u>cubo</u>
T3	vila <u>gátara</u> * nova	agora <u>paga</u> -se	lipídico	povo <u>cúbule</u> *
O4	<u>maga</u> taróloga	<u>mapa</u> galático	Mississipi dilúvico	<u>cuco</u> burríssimo
O5	<u>málaga</u> taça	<u>sátrapa</u> gago	<u>príncipe</u> dito	<u>búbuco</u> * bule
O6	<u>ágatas</u>	<u>galápaga</u>	Eurípides	<u>súcubo</u>
E7	Dinorá <u>Gatári</u> *	tafetá <u>pagável</u>	sushi <u>pedido</u>	zebu <u>cubute</u> *
E8	mesma <u>gataria</u>	mega <u>pagamento</u>	Denise <u>Pidinale</u> *	velho <u>cubulento</u> *

Como se pode verificar no QUADRO 3, os ambientes sonoros adjacentes à sílaba-teste foram controlados da seguinte forma:

- a) repetiu-se a mesma vogal da sílaba-teste nas sílabas anterior e seguinte;
- b) repetiu-se a mesma consoante inicial da sílaba-teste no início da sílaba anterior;
- c) utilizou-se a mesma consoante no início da sílaba seguinte, sempre uma oclusiva diferente da consoante da sílaba-teste.

Por exemplo, como se pode ver na primeira coluna do QUADRO 3, todas as palavras contendo a sílaba-teste [ga] apresentam sílaba anterior e seguinte com vogal [a], igual à sílaba-teste; sílaba anterior iniciada pela mesma consoante, [g], igual à sílaba-teste; sílaba seguinte iniciada pela mesma consoante, a oclusiva [t].

Ademais, as sentenças foram controladas quanto a:

- a) número de sílabas antes e depois da sílaba-teste, de forma a haver sempre três sílabas antes e quatro depois;
- b) posição das outras sílabas acentuadas antes e depois da sílaba acentuada da palavra-teste, formando um pé binário antes e um terciário depois.

No Apêndice A, encontram-se listadas as 32 sentenças formuladas para o experimento. Os estímulos, em ordenação aleatória, foram apresentados por meio do *software* LibreOffice Impress, do pacote LibreOffice (2010), no computador 2. Os informantes foram solicitados a ler as 32 sentenças, apresentadas, uma após a outra, em *slides* individuais. Cada uma das sentenças foi lida em modalidade neutra e, adicionalmente, em modalidade com foco na palavra-teste. Dessa forma, foi possível comparar a realização de cada uma das sílabas-teste em estruturas com níveis de ênfase distintos (neutra x foco), a fim de se isolar características específicas da proeminência focal de características acentuais. Foi inserida no experimento, assim, uma terceira variável independente com dois níveis (neutro x foco) e que chamaremos de “foco”. A realização do foco na palavra-teste foi induzida por meio de um paradigma pergunta-resposta, em um breve treinamento que antecedeu o início do experimento. Neste treinamento, o participante escutava um locutor (A) dizendo uma sentença afirmativa neutra. Em seguida, escutava uma pergunta, feita por outro locutor (B), cujo questionamento dizia respeito justamente à palavra-teste. Por fim, o participante escutava novamente o primeiro locutor respondendo à pergunta, com ênfase na palavra-teste, elemento da sentença que era alvo do questionamento.

Exemplo do paradigma apresentado no treinamento:

(Locutor A) *Aquela casa tombada está recebendo reparos* [sentença neutra]

(Locutor B) *Aquela casa ALUGADA?*

(Locutor A) Não, aquela casa TOMBADA está recebendo reparos. [sentença com foco]

Este paradigma (neutra-pergunta-foco) prosseguia durante o experimento. Após o treinamento, os participantes eram instruídos a assumir o papel do locutor (A): primeiro deveriam pronunciar a sentença mostrada no *slide* “normalmente” (i.e. forma neutra). Depois os participantes ouviam o locutor (B) fazendo uma pergunta que questionava uma das palavras da sentença. Nesse momento, o controlador do experimento mudava o *slide* da apresentação, que passava a apresentar a mesma sentença com a palavra-teste em letras maiúsculas, facilitando a identificação da palavra a receber o foco. Os participantes deveriam então pronunciar a sentença novamente, respondendo à pergunta (i.e., forma com foco).

Exemplo do paradigma apresentado na aquisição dos estímulos:

(participante) O avô gagá tagarela está dizendo bobagens [sentença neutra]

(locutor B) O avô LELÉ tagarela?

(participante) Não, o avô GAGÁ tagarela está dizendo bobagens [sentença com foco]

As perguntas do locutor (B) ao longo do experimento, bem como as sentenças utilizadas no treinamento foram gravadas previamente por um falante natural de Belo Horizonte, do sexo masculino, com a mesma faixa etária e escolaridade dos participantes testados. As palavras utilizadas nas perguntas eram da mesma classe gramatical, mesmo número de sílabas e mesmo padrão acentual das palavras-teste, além de serem semanticamente plausíveis como substitutas das mesmas.

Não foram dadas instruções específicas sobre a forma com que os participantes deveriam responder à pergunta do locutor, pois os participantes demonstraram facilidade em perceber as alterações induzidas pelo foco na sentença resposta, repetindo-as naturalmente. Com adição desta variável, o número de estímulos duplicou, passando a 64 (32 sentenças x dois níveis de “foco”).

A tarefa de leitura foi repetida quatro vezes por cada informante, com cerca de três dias de diferença entre cada sessão. Foram obtidos, portanto, 1280 dados (64 sentenças x quatro sessões x cinco informantes).

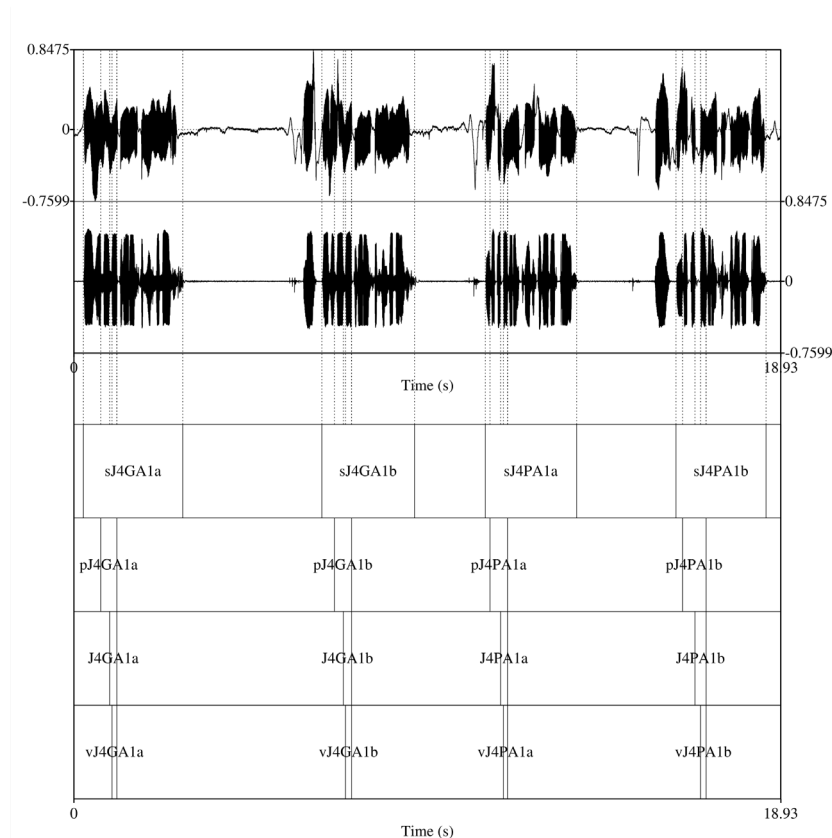
6.2.4 Pré-processamento dos dados

Uma vez obtidos os sinais apresentados acima, foi realizado um pré-processamento dos dados, que passaremos a descrever. O sinal de fala foi segmentado e anotado manualmente por meio da ferramenta TextGrid, do Praat, adequada à etiquetagem e anotação do sinal (BOERSMA; WEENINK, 2008). Para cada sessão

de coleta, foi gerado um arquivo com extensão *wave*, estéreo, contendo o sinal do EGG no canal esquerdo e o sinal de voz, no direito. Esse arquivo foi nomeado com a inicial do informante e o número da sessão (e.g. J1 para o informante J***, sessão 1). Foi então criado um arquivo de extensão TextGrid, pareado com o arquivo *wave* estéreo de cada sessão e com o mesmo nome (e.g. J1.wav; J1.TextGrid).

A FIG. 15, abaixo, mostra como exemplo um trecho da segmentação e anotação no arquivo TextGrid alinhado com o *wave* estéreo (de cima para baixo: sinal do EGG, sinal da fala, camadas de anotação/segmentação sentença, palavra, sílaba e vogal).

FIGURA 15 – Estrutura de camadas utilizada na segmentação do sinal obtido no experimento de produção



Como mostra a FIG. 15, a segmentação foi realizada em quatro camadas, todas de intervalo: na primeira, foram separadas as 64 sentenças testadas em cada sessão. Em seguida, em cada sentença, foi isolada a palavra-teste e, nesta, a sílaba-teste. Por fim, nas sílabas-teste, foi segmentada a vogal-teste. Dessa forma,

foi obtida no TextGrid uma estrutura em quatro camadas, a serem usadas posteriormente na medição por *scripts*.

As camadas do TextGrid foram anotadas usando os códigos apresentados no QUADRO 4, que representam as variáveis independentes em estudo, acrescidas de um código para indicar o nível da segmentação em uma das quatro camadas. Um exemplo de código seria vJ4GA1a (i.e. vogal-teste, informante J***, sessão 4, sílaba [ga], posição/distância acentuada/oxítônica, neutro).

QUADRO 4 – Códigos utilizados para anotação do sinal estéreo

Tipo	Símbolo	Significado
(nível)	s	sentença
	p	palavra-teste
		sílaba-teste
inicial do participante	v	vogal-teste
	E1-E4	participante E***, sessões 1 a 4
	J1-J4	participante J***, sessões 1 a 4
	G1-G4	participante G***, sessões 1 a 4
	R1-R4	participante R***, sessões 1 a 4
+ número da sessão de gravação	V1-V4	participante V***, sessões 1 a 4
	GA	[ga]
	PA	[pa]
	PI	[pi]
	KU	[ku]
sílaba	1	acentuada em oxítônica
	2	acentuada em paroxítônica
	3	acentuada em proparoxítônica
	4	postônica final adjacente
	5	postônica medial
	6	postônica final não adjacente
	7	pretônica adjacente
	8	pretônica não adjacente
posição e distância da tônica	a	neutra
	b	com foco

A segmentação e anotação foram realizadas nos arquivos estéreo de cada sessão, em que os sinais de fala e do EGG se encontram alinhados temporalmente, exceto no caso em que era necessário segmentar uma fricativa, quando apenas o canal do sinal de fala era visualizado. A segmentação do início e fim de todos os sons foi feita com

referência ao EGG, exceto no início e fim dos sons não vozeados que fossem precedidos ou seguidos por outro som não vozeado ou pausa, em que a segmentação foi feita com relação ao oscilograma ou ao espectrograma. Os seguintes ambientes sonoros foram encontrados na segmentação das sentenças, palavras-teste, sílabas-teste e vogais das sílabas-teste: início e fim de vogais orais, de consoantes oclusivas e fricativas; apenas início de consoantes africadas e nasais. Nos casos em que houve cancelamento da vogal final da sentença, considerou-se, para a segmentação, o fim da estrutura formântica das consoantes finais ou sua soltura, a depender da consoante envolvida. O QUADRO 5, a seguir, detalha os critérios de segmentação adotados para cada tipo de som e o instrumento de visualização utilizado como referência para a segmentação.

QUADRO 5 – Critérios adotados na segmentação do sinal de fala

Sons	Início	Fim
Vogais	(EGG) primeiro cruzamento de zero positivo no primeiro ciclo glótico completo correspondente à vogal	(EGG) último cruzamento de zero negativo no último ciclo glótico completo correspondente à vogal
Sons pré- ou pós-vocálicos	(EGG) cruzamento de zero positivo após o último ciclo glótico correspondente à vogal anterior	(EGG) cruzamento de zero positivo antes do primeiro ciclo glótico correspondente à vogal seguinte
Sons precedidos ou seguidos de som não vozeado ou pausa	Oclusivas e africadas não vozeadas: (Oscilograma) início do aumento de energia correspondente à soltura Fricativas: (Espectrograma) início da região de ruído Nasais: (Espectrograma) início da barra de vozeamento	Fricativas: (Espectrograma) fim da região de ruído

Uma vez segmentadas e anotadas as quatro camadas do TextGrid correspondente ao *wave* estéreo, realizou-se a separação dos dois canais do arquivo *wave* em dois arquivos mono, um correspondente ao sinal do EGG, outro, ao sinal de fala, por meio do *script* Monofy (ARANTES, 2010b), trabalhando ainda no Praat. Os arquivos mono foram mantidos com o mesmo nome dos estéreo, mas alocados em pastas diferentes.

Separados os dois canais, prosseguiu-se com o pré-processamento dos sinais. Quanto ao sinal do EGG, como sugerido na metodologia de Vieira (1997), foi realizada uma filtragem passa-faixa para retirada da flutuação de base no sinal, que pode ser provocada pela movimentação do pescoço, da cabeça ou dos eletrodos ou como resultado de rede elétrica. A filtragem foi realizada em MATLAB, com um filtro passa-faixas Chebyshev II 62 Hz – 7000 Hz, com *notch* em ≈ 60 Hz e atenuação de 45 dB na banda de rejeição. Este tipo de filtro foi selecionado por alterar em menor grau o sinal analisado, comparativamente a outros filtros, e por não introduzir *ripple* na banda de passagem, como o Chebyshev tipo I. As frequências de corte, 62 Hz a 7000Hz foram selecionadas, pois

se encontram acima da frequência de transmissão do sistema elétrico brasileiro e das taxas de movimentação da cabeça e laringe dos participantes, sendo ainda adequadas para falantes do sexo masculino. Foi utilizado um filtro passa-faixa, e não passa-altas, pois o sinal do EGG não apresenta eventos significativos acima de 7000Hz. Seguindo a metodologia de Vieira (1997), o procedimento de filtragem contemplou as seguintes etapas, com objetivo de evitar a defasagem: filtragem do sinal; inversão do sinal no eixo horizontal; filtragem do sinal invertido; re-inversão do sinal; conferência.

Após a filtragem, de volta ao Praat, o *script* Slicer (ARANTES, 2010c) foi usado para segmentar automaticamente cada vogal-teste a partir do arquivo mono filtrado com sinal de EGG. Foram gerados 1280 arquivos *wave* mono, correspondentes a cada vogal-teste, nomeados como no TextGrid (e.g. vJ4GA1a.wav). Tais arquivos foram usados nas medidas do sinal de fala, por meio de *scripts* do Praat, e do sinal do EGG, por meio do programa cedido pelo Prof. Maurílio Vieira e cujos princípios foram descritos em Vieira (1997).

6.2.5 Medição

Com relação ao processo de medição, iremos primeiramente explicitar as ferramentas utilizadas para realizar cada medida. Descreveremos, então, cada uma das medidas realizadas, divididas em medidas acústicas e medidas articulatórias. Por fim, retomaremos as relações entre as medidas acústicas e articulatórias apresentadas no cap. 3 e sistematizaremos possíveis relações de cada medida com a acentuação.

Ferramentas de medição

O sinal de fala obtido após a fase de pré-processamento foi analisado por meio do programa Praat (Boersma; Weenink, 2007). Foram utilizados *scripts* para automatizar as medidas de intensidade (ARANTES, 2008).

A partir do sinal do EGG, foram realizadas as demais medidas: duração, F0² e as medidas intrinsecamente articulatórias quocientes de contato (*contact quotient*, CQ), índice de velocidade (*speed index*, SI), fase de fechamento (*closing phase*, Cp) e fase de abertura (*opening phase*, Op). Tais medidas seguem a metodologia de Vieira (1997) e foram obtidas por meio de programa cedido pelo pesquisador.

² Duração e F0 não são medidas intrinsecamente fonatórias, mas podem ser facilmente obtidas a partir do sinal do EGG – no caso de duração, de forma equivalente; no caso de F0, com maior precisão que se obtida a partir do sinal da fala.

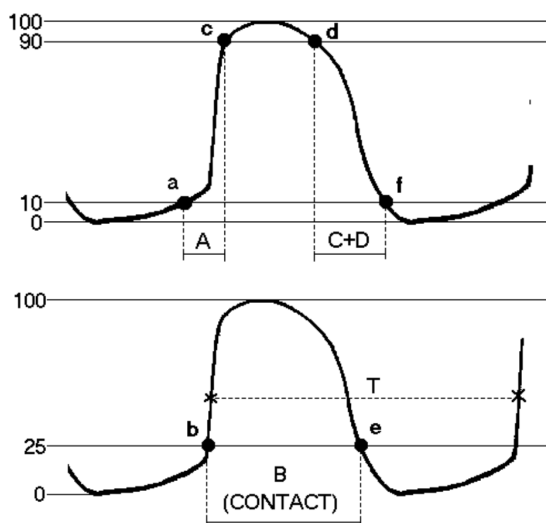
Medidas acústicas

As medidas acústicas realizadas foram duração, F0 e intensidade. A duração pode ser definida como a extensão temporal de um evento – convencionalmente é dada em segundos (s), mas a unidade milissegundo (ms) é preferida por vezes, por ser mais adequada ao nível de detalhe dos eventos da fala. A intensidade, que depende da pressão sonora, corresponde à potência sonora dividida pela unidade de área e é dada em W/m^2 . Aqui, como na maioria dos estudos sobre fala, a medida de intensidade é dada em dB, que é uma unidade em escala logarítmica. A frequência, dada em Hz, diz respeito ao número de repetições de um padrão em um segundo. A medida de frequência que utilizaremos se refere à frequência fundamental de vibração das cordas vocais (F0).

Medidas articulatórias

Foram adotadas quatro medidas articulatórias, quociente de contato (CQ), índice de velocidade (SI), fase de fechamento (Cp) e fase de abertura (Op), obtidas a partir do sinal do EGG. A FIG. 16, a seguir, reproduzida de Vieira (1987), apresenta os pontos de referência no sinal do EGG utilizados para medição do sinal glótico.

FIGURA 16 – Pontos de referência utilizados nas medidas da articulação glótica a partir do sinal do EGG



FONTE: VIEIRA, 1997, p. 105-106 (reprodução autorizada pelo autor)

Os pontos de referência indicados na FIG. 16 são definidos pelo autor como:

- a) T: período fundamental
- b) A: fase de fechamento das cordas vocais
- c) B: contato entre as cordas vocais
- d) C+D: soma das duas partes que compõem a fase de abertura das cordas vocais

A partir de um algoritmo de detecção de F0 e de tais pontos de referência, as quatro medidas específicas do sinal glótico que adotamos foram calculadas automaticamente pelo programa de Vieira (1997), da seguinte forma:

$$CQ = (B/T) \times 100\%$$

$$Cp = (A/T) \times 100\%$$

$$Op = ((C + D)/T) \times 100\%$$

$$SI = (Cp - Op)/(Cp + Op)$$

O quociente de contato (CQ), ou quociente de fechamento, corresponde à razão entre o contato das cordas vocais e o período fundamental, sendo o valor transformado em percentual. Valores de referência giram em torno de 50%.

As fases de fechamento (Cp) e de abertura (Op) foram calculadas respectivamente como a razão entre a trajetória de fechamento ou de abertura das cordas vocais e o período fundamental. Os valores são dados em porcentagens. Valores de referência giram em torno de 5 a 10% para Cp e de 33 a 39% para Op (no caso de Op, foi adotado o valor de referência obtido experimentalmente por Vieira (1997)).

O índice de velocidade (SI) é uma razão da diferença entre Cp e Op pela porção do sinal correspondente à soma das duas medidas. Varia de -1 a 1, sendo que o valor 0 corresponde a Cp e Op idênticos. O valor de referência é $\approx -0,65$, devido à tendência de que a fase de fechamento seja menor que a de abertura.

Relações entre medidas acústicas e articulatórias

Em geral, o aumento em F0 é acompanhado por um aumento em intensidade, mas as duas propriedades podem ser alteradas de forma independente (FANT, 1981). Considerando a alteração isolada de cada uma dessas propriedades, sabemos que F0 é diretamente proporcional à pressão subglótica, além de ser afetada pela tensão das cordas vocais. O aumento na pressão subglótica, por sua vez, pode se dar por dois mecanismos: pela ação dos pulmões e sistema respiratório ou pela resistência exercida pela glote à passagem da corrente de ar (ROTHERBERG, 1973).

Com relação à intensidade, Fant (1981) relata duas estratégias básicas: pode ser aumentada pelo aumento na pressão subglótica ou pela adução das cordas vocais causada pela compressão medial. No primeiro mecanismo, o

incremento de intensidade afeta as regiões de frequência de forma aproximadamente homogênea, não sendo esperadas, portanto, alterações em ênfase espectral. Já no segundo mecanismo, a energia espectral é redistribuída e, com isso, espera-se que o aumento na intensidade seja acompanhado por um aumento em ênfase espectral e uma diminuição em Cp. Além disso, de modo geral, o aumento da intensidade levaria ao aumento de CQ, devido a um aumento do contato relativo das cordas vocais.

Espera-se que as sílabas acentuadas apresentem maior duração, maior intensidade e menor F0 (nas sentenças neutras; nas com foco, F0 seria maior nas acentuadas) que as sílabas átonas. Quanto às medidas específicas do EGG, espera-se, nas sílabas acentuadas, um maior CQ (devido a um contato vocal mais prolongado), menor SI e maior Op (devido a uma fase de abertura mais demorada) com relação às átonas.

A seguir, sistematizamos os resultados esperados das medidas acústicas e articulatórias em relação ao acento. Considerando a hipótese inicial de que parâmetros acústicos e articulatórios produzidos nas sílabas tônicas são diferentes das sílabas pretônicas e postônicas, espera-se que, em comparação com as átonas, as sílabas acentuadas tenham

a) (Medidas acústicas)

- Duração maior
- F0 menor (sentenças neutras) ou maior (nas sentenças com foco)
- Intensidade maior

b) (Medidas articulatórias)

- CQ maior
- SI menor
- Cp menor
- Op maior

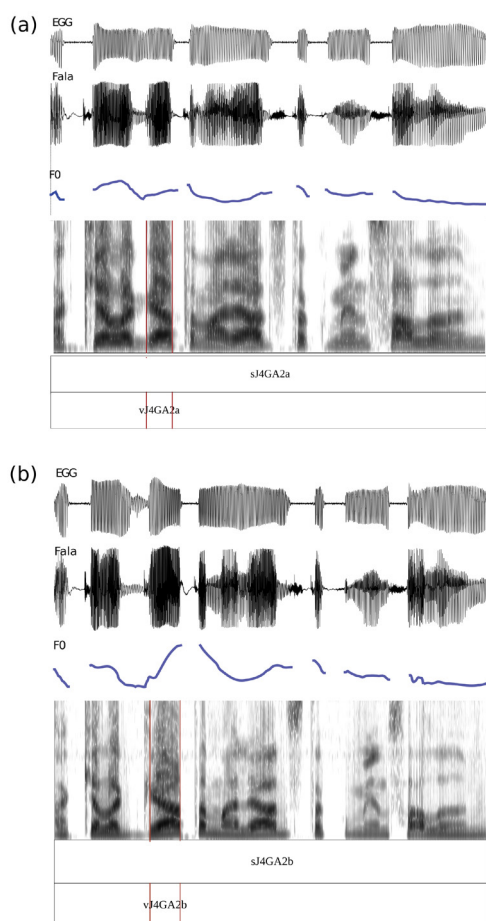
Quando da análise dos dados, serão levadas em consideração diferenças inerentes às sílabas testadas – portanto, não relacionadas à acentuação e fora do campo de interesse desse estudo. Vogais altas apresentam valores mais altos de F0 que as vogais baixas, o que seria uma característica encontrada universalmente (WHALEN; LEVITT, 1995). Apenas para algumas línguas, F0 é maior nas vogais posteriores que nas anteriores. SI, ao contrário, diminui com a posterioridade da vogal (MARASEK, 1996).

6.2.6 Resultados

A FIG. 17, a seguir, apresenta, a título de ilustração, dois exemplos de sinais de fala e de EGG obtidos para uma sentença neutra (“Aquele gata malhada está comendo salame”) e sua versão correspondente com foco. São mostrados também a curva de F0, o espectrograma do sinal de fala e as anotações realizadas por meio do TextGrid.

FIGURA 17 – Exemplos de dados obtidos no experimento de produção

(a) modalidade neutra e (b) com foco, sentença “Aquele gata malhada está comendo salame”



Na análise dos dados, buscou-se avaliar a distribuição dos padrões acentuais por participante e o tipo de sílaba, assim como possíveis correlações entre as variáveis do som da fala e do sinal EGG, para cada tipo de padrão acentual.

As variáveis dependentes avaliadas neste experimento são duração (dur), intensidade (int), frequência fundamental (F0), quociente de contato (CQ), índice de velocidade (SI), fase de fechamento (CP) e fase de abertura (Op), que correspondem a parâmetros possivelmente relevantes na realização da proeminência acentual.

As variáveis independentes são:

tipo: tipo de relação de proeminência entre a sílaba testada e a sílaba acentuada da palavra, com níveis T1, T2, T3, O4, O5, O6, E7, E8; foram avaliados também os agrupamentos T, O, E

foco: grau de ênfase conferida à palavra da sílaba, com níveis neutros ou com foco.

Para cada uma das variáveis dependentes, foram realizados testes a fim de identificar o modelo mais adequado aos dados. Iremos descrever, como exemplo, o procedimento adotado para a variável duração. O procedimento foi realizado isoladamente para os dados referentes aos níveis neutro e com foco da variável independente foco. Note-se que, antes de iniciar a análise estatística, foi realizada, para cada item testado, uma média dos valores obtidos nas quatro sessões de gravação, para cada participante, para cada frase e para cada tipo de foco.

Primeiramente, os dados foram ajustados a um modelo linear³ tendo como fator fixo o tipo, como mostrado em (i).

(i) $\text{dur} \sim \text{tipo}$

Em seguida, foi testado um modelo linear com efeitos mistos⁴, em que a variável “sujeito” foi inserida como fator aleatório, como mostrado em (ii).

(ii) $\text{dur} \sim \text{tipo} + (1 \mid \text{sujeito})$

Por fim, foi testado um modelo linear com efeitos mistos⁵, inserindo a variável “sílaba” como fator aleatório, como mostrado em (iii).

(iii) $\text{dur} \sim \text{tipo} + (1 \mid \text{sujeito}) + (1 \mid \text{sílaba})$

Os três modelos foram comparados por um teste de razão de verossimilhança⁶, que respeita a distribuição qui-quadrado.

(i) vs. (ii): $\chi^2[1] = 109,38 \text{ p} < 0,001$

(ii) vs. (iii): $\chi^2[1] = 260,40 \text{ p} < 0,001$

A diferença entre os modelos é estatisticamente significativa, o que justifica a utilização do modelo com os dois fatores aleatórios, que, apesar de apresentar mais coeficientes, se ajusta melhor aos dados que os demais.

³ R: função *lm*, pacote *stats*

⁴ R: função *lme*, pacote *nlme*

⁵ R: função *lmer*, pacote *lme4*

⁶ R: função *anova*, pacote *stats*

O mesmo procedimento foi seguido para cada uma das variáveis dependentes: além de duração, F0, intensidade, CQ, SI, Cp, Op. Para todas elas, o modelo misto com dois fatores aleatórios se mostrou mais adequado e foi, portanto, o adotado na análise.

Identificado o modelo ao qual ajustar os dados, para cada variável dependente, realizamos testes simultâneos de comparações múltiplas de médias⁷, para determinar a significância da interação entre os níveis das variáveis independentes inseridas como fatores fixos no modelo.

Na avaliação dos resultados, que passaremos a realizar, apresentaremos, para cada variável dependente, um gráfico com a distribuição dos dados, seguido de observações gerais e do teste de comparações múltiplas, que determina a significância de algumas das tendências observadas na distribuição dos dados.

Nas FIG. 18 a 24, a seguir, é mostrada, em boxplots, a distribuição dos dados obtidos no experimento para cada variável dependente, disposta no eixo vertical dos gráficos. No eixo horizontal, aparecem as variáveis tipo de proeminência, com seus oito níveis (numerados de 1 a 8), e acento, em dois níveis (a = neutro e b = foco). As cores das caixas dos gráficos correspondem ao agrupamento da variável tipo de proeminência em três níveis: acentuada (T) = rosa, postônica (O) = azul e pretônica (E) = amarelo.

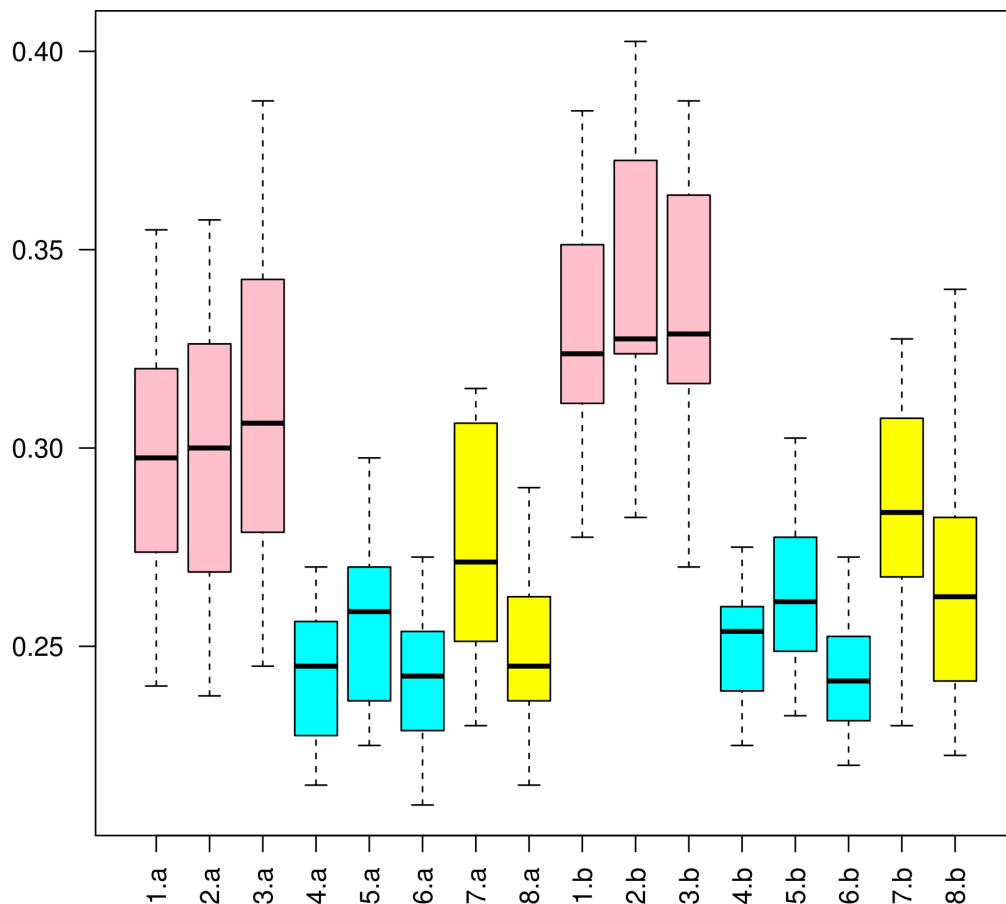
Segundo Crawley (2007), os boxplots mostram a distribuição dos dados quanto a sua localização e dispersão, assim como a assimetria da distribuição. Em cada retângulo do gráfico, os limites inferior e superior correspondem respectivamente ao primeiro e ao terceiro quartis, representando, portanto, como estão distribuídos os 50% centrais dos dados. A linha mais escura no interior da caixa indica a mediana ou o centro dos dados. O deslocamento do centro dos dados (mediana) com relação ao centro do retângulo corresponde à assimetria dos dados. As linhas tracejadas mostram o que for menor entre o valor máximo dos dados e 1,5 vezes a distância entre o primeiro e terceiro quartis (aproximadamente dois desvios-padrão), representando a variabilidade dos dados. Os eventuais pontos que se estendem além da linha tracejada são outliers. Adicionalmente, incluímos nos boxplots a média, representada por um círculo sólido cinza, e o erro padrão, representado por um traço sólido cinza, ambos sobrepostos aos retângulos de cada grupo.

Depois de cada boxplot, serão apresentadas tabelas com resultados de teste estatístico, com a finalidade de determinar se as distribuições dos dados observadas no gráfico são significativamente diferentes umas das outras. A estatística utilizada em tais comparações é o z-score, que é um teste de hipótese baseado no desvio padrão. Testa se as médias de duas distribuições são iguais (hipótese nula). De forma simplificada, o valor de p corresponde à probabilidade de que a hipótese nula seja verdadeira, considerando o z-score encontrado. Adotaremos significância a $p \leq \alpha = 0,05$ (que equivale a um nível de confiança de 95%).

⁷ R: função *glht*, pacote *multcomp*

A seguir, a FIG. 18 mostra a distribuição dos dados para a variável duração.

FIGURA 18 – Distribuição dos dados de duração (s) de acordo com tipo de proeminência, localização da proeminência e foco



<i>Tipo de proeminência</i>		<i>Localização da proeminência</i>	
	acentuada	1	acentuada final
	postônica	2	acentuada seguida por uma postônica
	pretônica	3	acentuada seguida por duas postônicas
<i>Foco</i>		4	postônica final única
		5	postônica final antecedida por postônica medial
a	neutra	6	postônica medial
b	com foco	7	pretônica adjacente à acentuada
		8	pretônica anterior à sílaba adjacente à acentuada

Na FIG. 18, podemos identificar as seguintes características de distribuição da propriedade duração: as sílabas acentuadas (rosa) apresentam maior variabilidade que as pretônicas (amarelo) e estas, por sua vez, que as postônicas (azul). Há grande assimetria nas acentuadas com foco (1b, 2b e 3b). São muito próximos os centros das acentuadas neutras entre si; das acentuadas com foco entre si; das postônicas final única e medial entre si, o que seria uma pista de que as amostras não são significativamente diferentes umas das outras.

Podem ser observadas tendências claras com relação aos valores de duração: as sílabas acentuadas apresentam maior duração que as átonas, tanto nas sentenças neutras, quanto nas com foco. Entre as átonas, as sílabas pretônicas se mostraram mais longas que as postônicas.

Essas observações gerais foram avaliadas por testes simultâneos de comparações múltiplas de médias, que são mostrados na TAB. 3, a seguir.

TABELA 3 – Resultados dos testes de comparações múltiplas para a variável duração

Duração							
neutro			foco				
Níveis	valor z	p (> z)		Níveis	valor z	p (> z)	
T ~ O	18,915	<1e-04	***	T ~ O	28,901	<1e-04	***
T ~ E	11,339	<1e-04	***	T ~ E	18,154	<1e-04	***
O ~ E	-9,657	<1e-04	***	O ~ E	-14,079	<1e-04	***
T1 ~ T2	-0,099	1,000		T1 ~ T2	-3,147	0,0139 *	
T1 ~ T3	-2,344	0,140		T1 ~ T3	-1,735	0,4641	
T2 ~ T3	-2,245	0,176		T2 ~ T3	1,413	0,6952	
O4 ~ O5	-2,640	0,065		O4 ~ O5	-2,379	0,1286	
O4 ~ O6	0,395	0,999		O4 ~ O6	1,685	0,4993	
O5 ~ O6	3,034	0,020	*	O5 ~ O6	4,064	0,0004	***
E7 ~ E8	-2,640	0,065		E7 ~ E8	-2,379	0,1286	

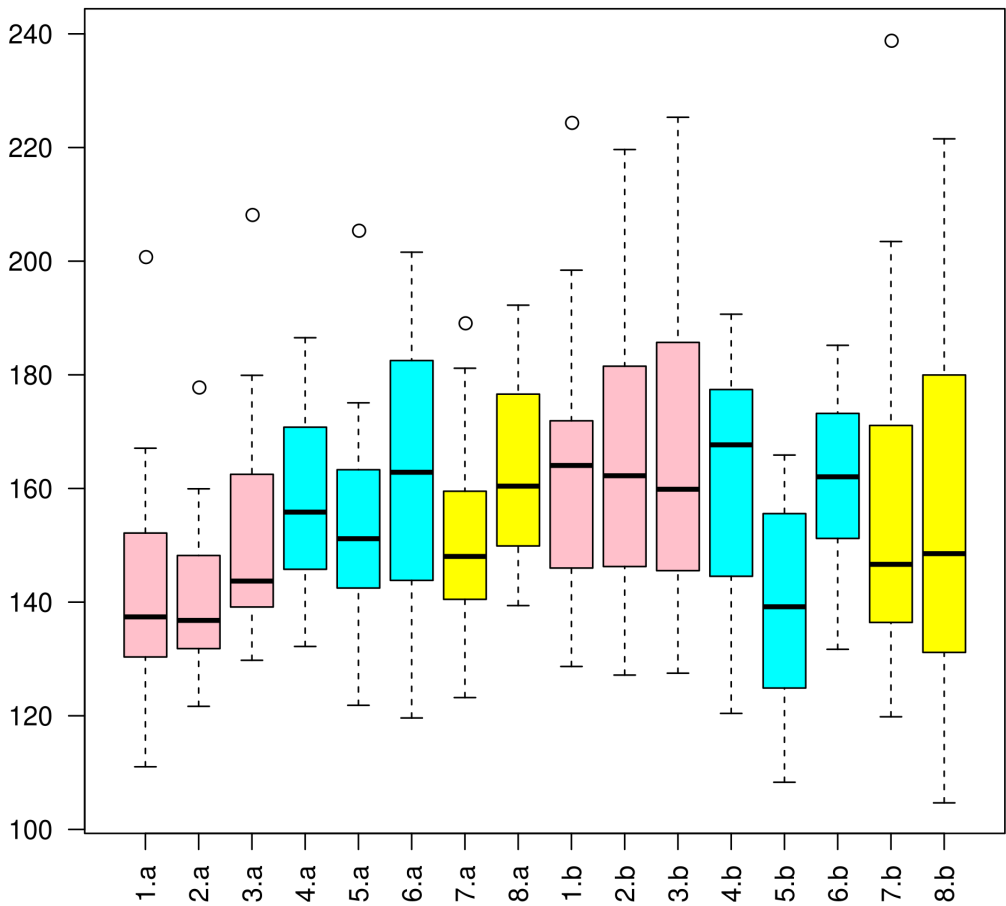
Códigos de significância: *** 0,001 ** 0,01 * 0,05




A TAB. 3 mostra que, nas sentenças neutras – considerando a significância a $p \leq 0,05$ –, a variável duração se mostrou significativa para diferenciar acentuadas (T) de postônicas (O), acentuadas (T) de pretônicas (E), postônicas (O) de pretônicas (E), assim como as sílabas postônicas finais (O5) e mediais (O6) em uma proparoxítona.

Nas sentenças com foco, foram observados resultados semelhantes, exceto por, nesse caso, a duração ter adicionalmente se mostrado significativa para diferenciar sílabas acentuadas finais (T1) de sílabas acentuadas seguidas por uma postônica (T2).

A seguir, a FIG. 19 mostra a distribuição dos dados para a variável F0.

FIGURA 19 – Distribuição dos dados de F0 (Hz) de acordo com tipo de proeminência, localização da proeminência e foco



Tipo de proeminência		Localização da proeminência	
	acentuada	1	acentuada final
	postônica	2	acentuada seguida por uma postônica
	pretônica	3	acentuada seguida por duas postônicas
Foco	a neutra	4	postônica final única
		5	postônica final antecedida por postônica medial
	b com foco	6	postônica medial
		7	pretônica adjacente à acentuada
		8	pretônica anterior à sílaba adjacente à acentuada

Na FIG. 19, podemos identificar as seguintes características de distribuição da propriedade F0: há grande variabilidade em algumas categorias, mas aparentemente assistemática. Há certa assimetria nas acentuadas neutras (1a, 2a e 3a) e, nas sentenças com foco, na acentuada final (1b). São muito próximos os centros das acentuadas neutras 1a e 2a entre si; das acentuadas com foco entre si; das postônicas final única (4a) e final antecedida por medial (5a) entre si, o que seria uma pista de que tais amostras não são significativamente diferentes entre si.

Podem ser observadas algumas tendências com relação aos valores de F0: as sílabas acentuadas apresentam menor F0 que as átonas nas sentenças neutras, o que se inverte nas sentenças com foco, em que as sílabas acentuadas apresentam maiores valores de F0 que as átonas. Essas observações gerais foram avaliadas por testes simultâneos de comparações múltiplas de médias, que são mostrados na TAB. 4, a seguir.

TABELA 4 – Resultados dos testes de comparações múltiplas para a variável F0

F0							
neutro				foco			
Níveis	valor z	p (> z)		Níveis	valor z	p (> z)	
T ~ O	-6,567	<1e-04	***	T ~ O	3,221	0,011	*
T ~ E	-4,449	<1e-04	***	T ~ E	2,427	0,114	
O ~ E	2,934	0,028	*	O ~ E	-1,240	0,806	
T1 ~ T2	-0,033	1,000		T1 ~ T2	-0,642	0,991	
T1 ~ T3	-2,919	0,029	*	T1 ~ T3	-0,743	0,980	
T2 ~ T3	-2,886	0,032	*	T2 ~ T3	-0,101	1,000	
O4 ~ O5	1,348	0,739		O4 ~ O5	3,726	0,002	**
O4 ~ O6	-1,183	0,838		O4 ~ O6	0,149	1,000	
O5 ~ O6	-2,530	0,088		O5 ~ O6	-3,577	0,003	**
E7 ~ E8	1,348	0,739		E7 ~ E8	3,726	0,002	**

Códigos de significância: *** 0,001 ** 0,01 * 0,05

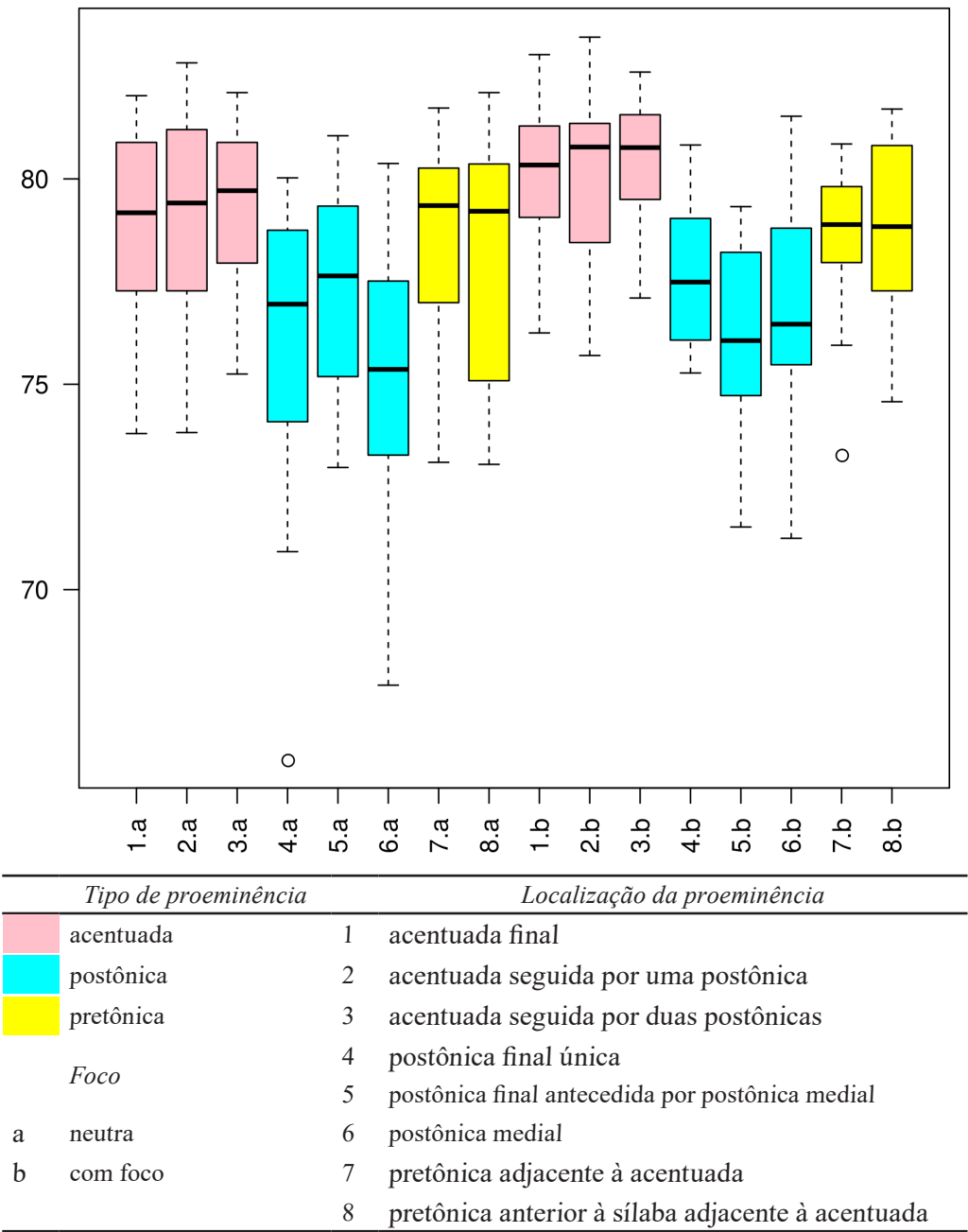
A TAB. 4 mostra que, nas sentenças neutras – considerando a significância a $p \leq 0,05$ –, a variável F0 se mostrou significativa para diferenciar acentuadas (T) de postônicas (O), acentuadas (T) de pretônicas (E), postônicas (O) de pretônicas (E), assim como para diferenciar acentuadas finais (T1) de acentuadas seguidas por duas postônicas (T3) e acentuadas seguidas por uma postônica (T2) de acentuadas seguidas por duas postônicas (T3).

Nas sentenças com foco, foram observados resultados muito diferentes dos observados nas sentenças neutras: F0 mostrou-se relevante para distinguir apenas as sílabas postônicas únicas (O4) de postônicas finais antecedidas de postônica

medial (O5), estas de postônicas mediais (O6) e pretônicas adjacentes (E7) de pretônicas anteriores à adjacente (E8).

A seguir, a FIG. 20 mostra a distribuição dos dados para a variável intensidade.

FIGURA 20 – Distribuição dos dados de intensidade (dB) de acordo com tipo de proeminência, localização da proeminência e foco



Na FIG. 20, podemos identificar as seguintes características de distribuição da propriedade intensidade: destaca-se das demais a variabilidade da postônica medial neutra (6a). Há assimetria nas pretônicas neutras (7a, 8a) e, nas sentenças com foco, na acentuada seguida por uma postônica (2b). São muito próximos os centros das acentuadas neutras entre si; das pretônicas neutras entre si; das acentuadas com foco entre si; das pretônicas com foco entre si, o que seria uma pista de que tais amostras não são significativamente diferentes entre si.

Podem ser observadas as seguintes tendências com relação aos valores de intensidade: tanto as sílabas acentuadas quanto as pretônicas apresentam maior intensidade que as postônicas, o que acontece nas sentenças neutras e com foco.

Essas observações gerais foram avaliadas por testes simultâneos de comparações múltiplas de médias, que são mostrados na TAB. 5, a seguir.

TABELA 5 – Resultados dos testes de comparações múltiplas para a variável intensidade

Intensidade							
neutro				foco			
Níveis	valor z	p (> z)		Níveis	valor z	p (> z)	
T ~ O	8,295	<1e-04	***	T ~ O	10,861	< 1e-04	***
T ~ E	1,779	0,433		T ~ E	4,218	0,0002	***
O ~ E	-6,842	<1e-04	***	O ~ E	-7,417	< 1e-04	***
T1 ~ T2	-0,480	0,998		T1 ~ T2	0,190	1,0000	
T1 ~ T3	-0,789	0,974		T1 ~ T3	-0,787	0,9738	
T2 ~ T3	-0,309	1,000		T2 ~ T3	-0,978	0,9270	
O4 ~ O5	-1,936	0,332		O4 ~ O5	2,632	0,0667	
O4 ~ O6	1,369	0,725		O4 ~ O6	1,105	0,8763	
O5 ~ O6	3,305	0,008	**	O5 ~ O6	-1,527	0,6141	
E7 ~ E8	-1,936	0,332		E7 ~ E8	2,632	0,0666	

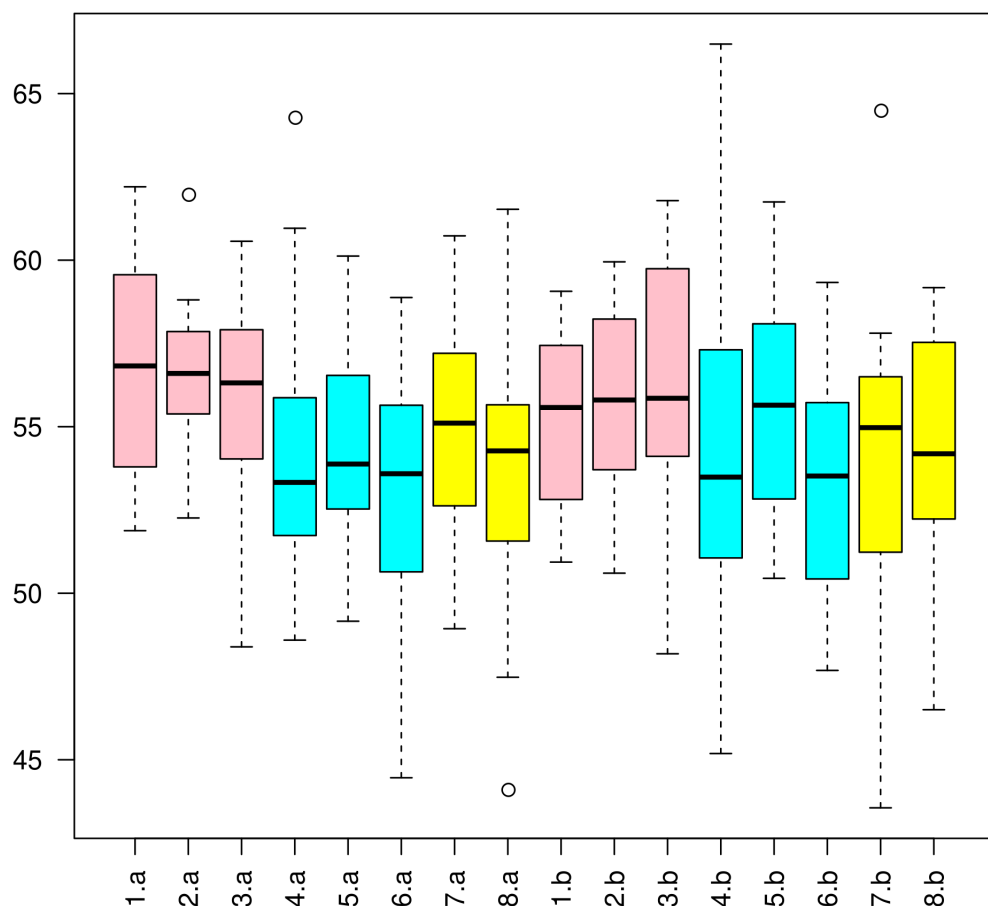
Códigos de significância: *** 0,001 ** 0,01 * 0,05

A TAB. 5 mostra que, nas sentenças neutras – considerando a significância a $p \leq 0,05$ –, a variável intensidade se mostrou significativa para diferenciar acentuadas (T) de postônicas (O) e postônicas (O) de pretônicas (E), assim como para diferenciar postônicas finais antecedidas de postônica medial (O5) de postônicas mediais (O6).

Nas sentenças com foco, foram observadas algumas diferenças: intensidade mostrou-se significativa para diferenciar acentuadas (T) de postônicas (O), acentuadas (T) de pretônicas (E), postônicas (O) de pretônicas (E).

A seguir, a FIG. 21 mostra a distribuição dos dados para a variável CQ.

FIGURA 21 – Distribuição dos dados de quociente de contato CQ (%) de acordo com tipo de proeminência, localização da proeminência e foco



<i>Tipo de proeminência</i>		<i>Localização da proeminência</i>	
<div style="display: inline-block; width: 15px; height: 15px; background-color: #FF9999; border: 1px solid black; margin-right: 5px;"></div> <div style="display: inline-block; width: 15px; height: 15px; background-color: #00FFFF; border: 1px solid black; margin-right: 5px;"></div> <div style="display: inline-block; width: 15px; height: 15px; background-color: #FFFF00; border: 1px solid black; margin-right: 5px;"></div>	acentuada	1	acentuada final
	postônica	2	acentuada seguida por uma postônica
	pretônica	3	acentuada seguida por duas postônicas
<i>Foco</i>		4	postônica final única
		5	postônica final antecedida por postônica medial
a	neutra	6	postônica medial
b	com foco	7	pretônica adjacente à acentuada
		8	pretônica anterior à sílaba adjacente à acentuada

Na FIG. 21, podemos identificar as seguintes características de distribuição da propriedade quociente de contato (CQ): há grande variabilidade em todas as

categorias, exceto, nas neutras, a acentuada seguida de uma postônica (2a) e, nas com foco, a acentuada final (1b) e também a acentuada seguida de uma postônica (2b). Há certa assimetria, nas sentenças neutras, na postônica final (4a), na postônica antecedida de postônica medial (5a) e na pretônica não adjacente (8a). Nas sentenças com foco, há assimetria na acentuada seguida de duas postônicas (3b), bem como nas pretônicas. São muito próximos os centros das acentuadas neutras entre si; das postônicas neutras entre si; das pretônicas neutras entre si; das acentuadas com foco entre si, o que seria uma pista de que tais amostras não são significativamente diferentes entre si.

Essas observações gerais foram avaliadas por testes simultâneos de comparações múltiplas de médias, que são mostrados na TAB. 6, a seguir.

TABELA 6 – Resultados dos testes de comparações múltiplas para a variável CQ

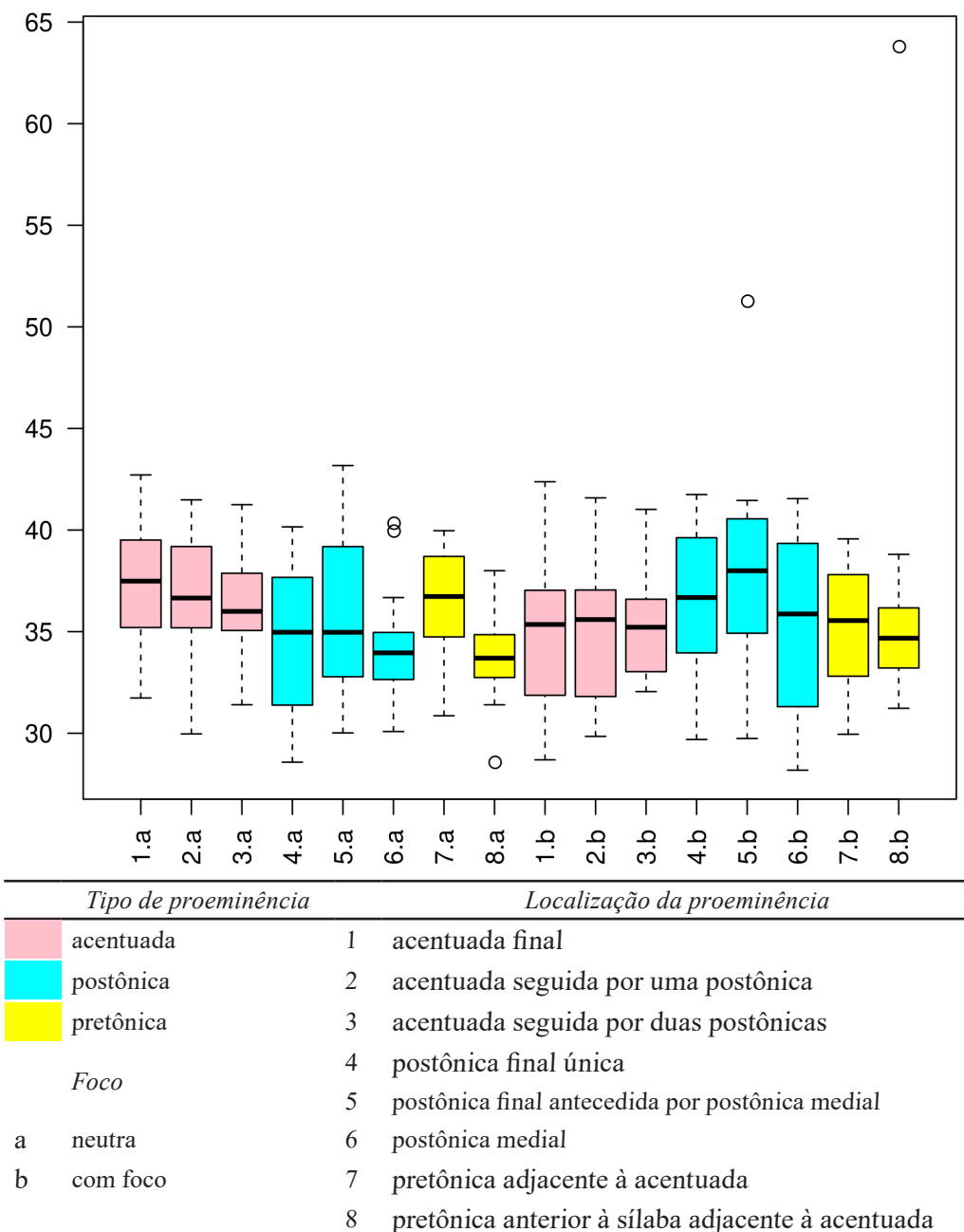
Intensidade						
neutro				foco		
Níveis	valor z	p ($> z $)		Níveis	valor z	p ($> z $)
T ~ O	4,681	<1e-04	***	T ~ O	0,334	1,932
T ~ E	3,241	0,010	*	T ~ E	2,505	0,094
O ~ E	-2,035	0,275		O ~ E	0,113	1,000
T1 ~ T2	0,253	1,000		T1 ~ T2	-0,293	1,000
T1 ~ T3	0,881	0,955		T1 ~ T3	-1,141	0,859
T2 ~ T3	0,628	0,992		T2 ~ T3	-0,848	0,962
O4 ~ O5	0,143	1,000		O4 ~ O5	-1,002	0,919
O4 ~ O6	1,225	0,815		O4 ~ O6	0,934	0,941
O5 ~ O6	1,081	0,887		O5 ~ O6	1,937	0,331
E7 ~ E8	0,143	1,000		E7 ~ E8	-1,002	0,918

Códigos de significância: *** 0,001 ** 0,01 * 0,05

Como podemos ver na TAB. 6, nas sentenças neutras – considerando a significância a $p \leq 0,05$ –, a variável CQ se mostrou significativa para diferenciar acentuadas (T) de postônicas (O) e postônicas (O) de pretônicas (E), assim como para diferenciar postônicas finais antecedidas de postônica medial (O5) de postônicas mediais (O6).

Nas sentenças com foco, foram observadas algumas diferenças: CQ mostrou-se significativo para diferenciar acentuadas (T) de postônicas (O), acentuadas (T) de pretônicas (E), postônicas (O) de pretônicas (E).

A seguir, a FIG. 22 mostra a distribuição dos dados para a variável SI.

FIGURA 22 – Distribuição dos dados de índice de velocidade SI (u.a.) de acordo com tipo de proeminência, localização da proeminência e foco

Na FIG. 22, podemos identificar as seguintes características de distribuição da propriedade índice de velocidade (SI): há grande variabilidade em todas as

categorias, exceto nas acentuadas neutras seguidas de uma postônica (2a) e de duas postônicas (3a), assim como na acentuada com foco seguida de duas postônicas (3b). Nesses três casos, note-se, contudo, a presença de *outliers* no que corresponde à cauda inferior dos dados. A assimetria pode ser descrita de forma semelhante, ocorrendo de forma generalizada nas categorias, exceto nos três casos de pouca variabilidade, mas com *outliers*. Quanto aos centros, aproximam-se as acentuadas neutras seguidas de uma (2a) e duas postônicas (3a); as postônicas neutras final (4a) e medial (6a); nas sentenças com foco, a acentuada final (1b) e seguida de duas postônicas. A proximidade do centro dessas amostras seria uma pista de que elas não são significativamente diferentes entre si.

Essas observações gerais foram avaliadas por testes simultâneos de comparações múltiplas de médias, que são mostrados na TAB. 7, a seguir.

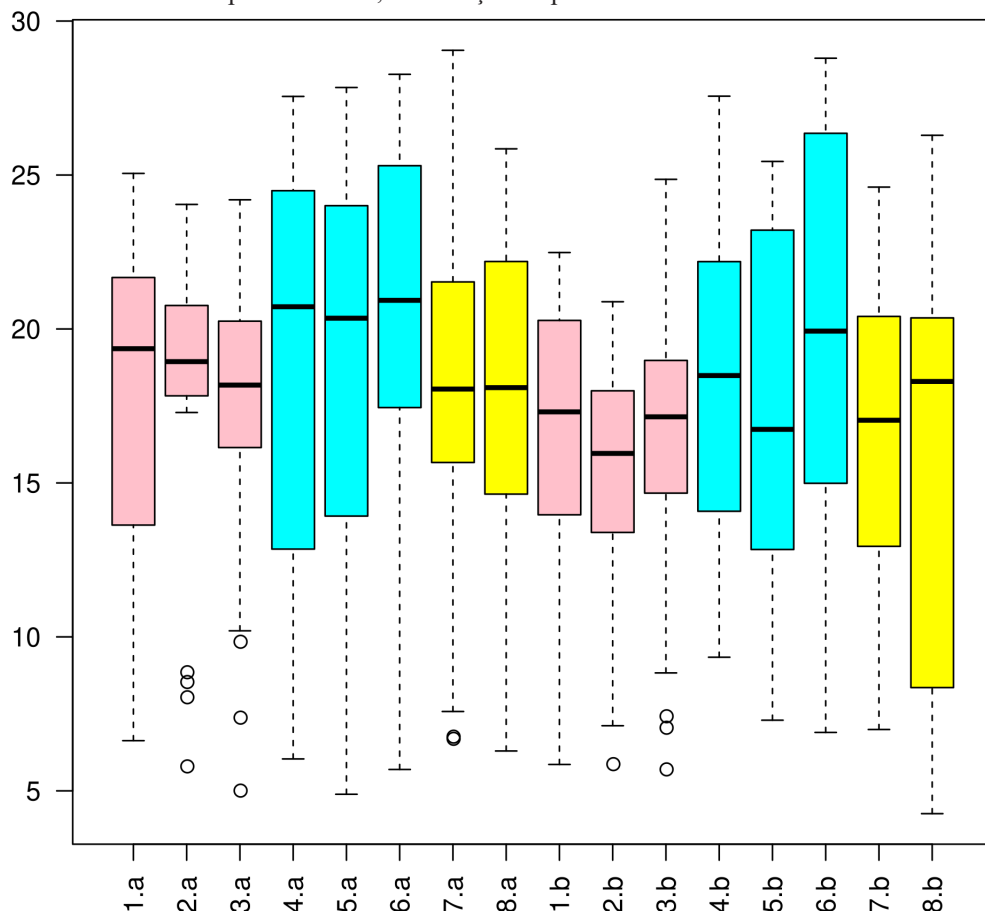
TABELA 7 – Resultados dos testes de comparações múltiplas para a variável SI

SI					
neutro			foco		
Níveis	valor z	p (> z)	Níveis	valor z	p (> z)
T ~ O	-2,176	0,205	T ~ O	-2,497	0,096
T ~ E	-0,118	1,000	T ~ E	2,706	1,000
O ~ E	2,070	0,256	O ~ E	0,269	0,054
T1 ~ T2	-0,998	0,920	T1 ~ T2	1,203	0,827
T1 ~ T3	-0,665	0,989	T1 ~ T3	0,469	0,998
T2 ~ T3	0,333	1,000	T2 ~ T3	-0,734	0,982
O4 ~ O5	1,606	0,557	O4 ~ O5	1,605	0,558
O4 ~ O6	1,089	0,884	O4 ~ O6	-0,688	0,987
O5 ~ O6	-0,495	0,998	O5 ~ O6	-2,292	0,158
E7 ~ E8	1,606	0,557	E7 ~ E8	1,605	0,558

Códigos de significância: *** 0,001 ** 0,01 * 0,05

Como podemos ver na TAB. 7, de modo geral – considerando a significância a $p \leq 0,05$ –, a variável SI não se mostrou significativa em nenhuma das comparações. Esse resultado pode estar relacionado à grande variabilidade observada em todas as categorias.

A seguir, a FIG. 23 mostra a distribuição dos dados para a variável Cp.

FIGURA 23 – Distribuição dos dados de fase de fechamento Cp (%) de acordo com tipo de proeminência, localização da proeminência e foco

<i>Tipo de proeminência</i>		<i>Localização da proeminência</i>	
<div style="display: inline-block; width: 15px; height: 15px; background-color: #FF9999; border: 1px solid black; margin-right: 5px;"></div> acentuada <div style="display: inline-block; width: 15px; height: 15px; background-color: #00FFFF; border: 1px solid black; margin-right: 5px;"></div> postônica <div style="display: inline-block; width: 15px; height: 15px; background-color: #FFFF00; border: 1px solid black; margin-right: 5px;"></div> pretônica		1	acentuada final
		2	acentuada seguida por uma postônica
		3	acentuada seguida por duas postônicas
<i>Foco</i> a neutra b com foco		4	postônica final única
		5	postônica final antecedida por postônica medial
	a	6	postônica medial
	b	7	pretônica adjacente à acentuada
		8	pretônica anterior à sílaba adjacente à acentuada

Na FIG. 23, podemos identificar as seguintes características de distribuição da propriedade fase de fechamento (Cp): há grande variabilidade em todas as

categorias, exceto na acentuada neutra seguida uma postônica (2a), em que há, contudo, muitos *outliers* no que corresponde à cauda esquerda dos dados. Há assimetria na maioria das categorias, exceto na acentuada neutra seguida de duas postônicas (3a) e, nas sentenças com foco, nas acentuadas, na postônica final (4b) e na pretônica não adjacente (7b). Quanto aos centros, aproximam-se as postônicas neutras; as pretônicas neutras; nas sentenças com foco, as acentuadas final (1b) e seguida de duas postônicas (3b).

Essas observações gerais foram avaliadas por testes simultâneos de comparações múltiplas de médias, que são mostrados na TAB. 8, a seguir.

TABELA 8 – Resultados dos testes de comparações múltiplas para a variável Cp

Cp							
neutro			foco				
Níveis	valor z	p (> z)		Níveis	valor z	p (> z)	
T ~ O	-2,924	0,028	*	T ~ O	-4,916	< 1e-04	***
T ~ E	-0,311	1,000		T ~ E	-1,120	0,869	
O ~ E	2,657	0,062		O ~ E	3,978	0,001	***
T1 ~ T2	-0,199	1,000		T1 ~ T2	1,199	0,829	
T1 ~ T3	0,338	1,000		T1 ~ T3	0,243	1,000	
T2 ~ T3	0,537	0,997		T2 ~ T3	-0,956	0,934	
O4 ~ O5	0,782	0,975		O4 ~ O5	0,750	0,979	
O4 ~ O6	-0,491	0,998		O4 ~ O6	-1,533	0,610	
O5 ~ O6	-1,263	0,792		O5 ~ O6	-2,283	0,162	
E7 ~ E8	0,782	0,975		E7 ~ E8	0,750	0,979	

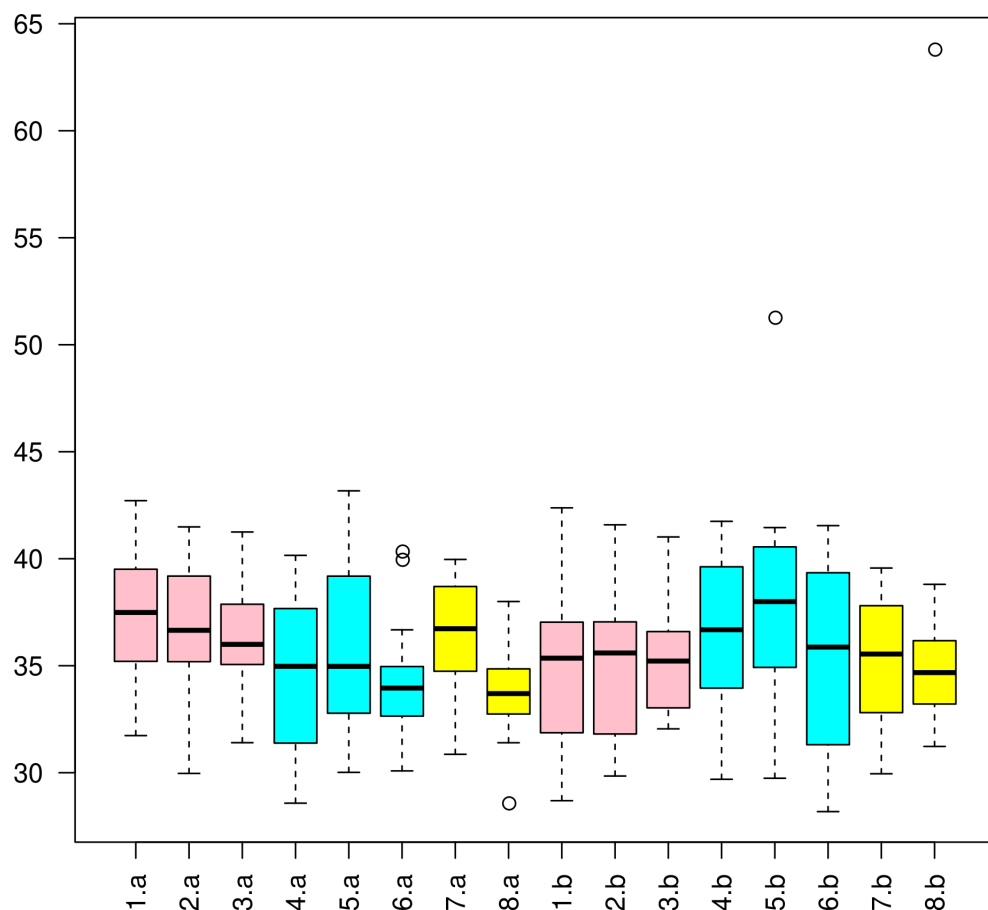
Códigos de significância: *** 0,001 ** 0,01 * 0,05

Como podemos ver na TAB. 8, nas sentenças neutras – considerando a significância a $p \leq 0,05$ –, a variável Cp se mostrou significativa apenas para diferenciar acentuadas (T) de postônicas (O).

Nas sentenças com foco, Cp mostrou-se significativa para diferenciar não só acentuadas (T) de postônicas (O) como também postônicas (O) de pretônicas (E).

A seguir, a FIG. 24 mostra a distribuição dos dados para a variável Op.

FIGURA 24 – Distribuição dos dados de fase de abertura Op (%) de acordo com tipo de proeminência, localização da proeminência e foco



<i>Tipo de proeminência</i>		<i>Localização da proeminência</i>	
<div style="display: inline-block; width: 15px; height: 15px; background-color: #FF9999; border: 1px solid black; margin-right: 5px;"></div> <div style="display: inline-block; width: 15px; height: 15px; background-color: #00FFFF; border: 1px solid black; margin-right: 5px;"></div> <div style="display: inline-block; width: 15px; height: 15px; background-color: #FFFF00; border: 1px solid black; margin-right: 5px;"></div>	acentuada	1	acentuada final
	postônica	2	acentuada seguida por uma postônica
	pretônica	3	acentuada seguida por duas postônicas
<i>Foco</i>		4	postônica final única
		5	postônica final antecedida por postônica medial
a	neutra	6	postônica medial
b	com foco	7	pretônica adjacente à acentuada
		8	pretônica anterior à sílaba adjacente à acentuada

Na FIG. 24, podemos identificar as seguintes características de distribuição da propriedade fase de abertura (Op): há certa variabilidade em todas as categorias.

Verifica-se uma distribuição assimétrica em (2a) exceto nos três casos de pouca variabilidade, mas com *outliers*. Quanto aos centros, aproximam-se as acentuadas neutras seguidas de uma (2a) e duas postônicas (3a); as postônicas neutras final (4a) e medial (6a); nas sentenças com foco, a acentuada final (1b) e seguida de duas postônicas. A proximidade do centro dessas amostras seria uma pista de que elas não são significativamente diferentes entre si.

Essas observações gerais foram avaliadas por testes simultâneos de comparações múltiplas de médias, que são mostrados na TAB. 9, a seguir.

TABELA 9 – Resultados dos testes de comparações múltiplas para a variável *Op*

SI						
neutro			foco			
Níveis	valor z	p ($> z $)		Níveis	valor z	p ($> z $)
T ~ O	3,860	<1e-04	***	T ~ O	-2,242	0,178
T ~ E	2,338	0,142		T ~ E	-0,850	0,962
O ~ E	-1,938	0,331		O ~ E	1,536	0,607
T1 ~ T2	0,887	0,953		T1 ~ T2	0,033	1,000
T1 ~ T3	1,432	0,682		T1 ~ T3	-0,556	0,996
T2 ~ T3	0,545	0,996		T2 ~ T3	-0,590	0,994
O4 ~ O5	-1,266	0,791		O4 ~ O5	-0,959	0,933
O4 ~ O6	0,539	0,996		O4 ~ O6	1,115	0,872
O5 ~ O6	1,788	0,427		O5 ~ O6	2,074	0,254
E7 ~ E8	-1,266	0,791		E7 ~ E8	-0,959	0,933

Códigos de significância: *** 0,001 ** 0,01 * 0,05

Como podemos ver na TAB. 9, de modo geral – considerando a significância a $p \leq 0,05$ –, a variável *Op* se mostrou significativa somente para diferenciar acentuadas (T) de postônicas (O).

O QUADRO 6, a seguir, resume os resultados de comparações múltiplas apresentados nas TAB. 3 a 9. As células coloridas indicam as comparações em que a variável dependente foi significativa, ou seja, os tipos de sílabas que é capaz de distinguir. Células em rosa correspondem às sentenças neutras e células em cinza, às sentenças com foco.

QUADRO 6 – Resumo dos testes simultâneos de comparações múltiplas de médias

	Dur		F0		Int		CQ		SI		Cp		Op	
	neut	foco	neut	foco	neut	foco	neut	foco	neut	foco	neut	foco	neut	foco
T ~ O														
T ~ E														
O ~ E														
T1 ~ T2														
T1 ~ T3														
T2 ~ T3														
O4 ~ O5														
O4 ~ O6														
O5 ~ O6														
E7 ~ E8														

O QUADRO 6 permite identificar um maior número de relações entre as variáveis correspondentes às medidas acústicas (duração, F0, intensidade) e o acento do que entre as variáveis correspondentes às medidas articulatórias (CQ, SI, Cp e Op) e o acento. Consideraremos primeiramente as sentenças neutras. De modo geral, à exceção de SI, todas as variáveis foram significativas para distinguir sílabas acentuadas de sílabas postônicas.

Especificamente com relação às medidas acústicas, vemos que duração e F0 são capazes de distinguir entre os três tipos de sílabas – acentuadas, postônicas e pretônicas –, ao passo que intensidade só distingue acentuadas de postônicas e postônicas de pretônicas – mas não as acentuadas das pretônicas. Duração e intensidade são responsáveis pela distinção entre a postônica final e a medial de uma proparoxítona (O5 ~ O6). Adicionalmente, F0 distingue entre sílabas acentuadas de oxítonas e sílabas acentuadas de proparoxítonas (T1 ~ T3), bem como sílabas acentuadas de paroxítonas e sílabas acentuadas de proparoxítonas (T2 ~ T3).

Quanto às medidas articulatórias, observamos que tanto CQ quanto Cp e Op distinguem entre sílabas acentuadas de sílabas postônicas, mas somente CQ distingue sílabas acentuadas de pretônicas.

Considerando as sentenças com foco, verifica-se que algumas propriedades mudam o papel contrastivo que desempenhavam nas sentenças neutras. Nas

medidas acústicas, vemos que duração mantém a capacidade de diferenciar acentuadas, postônicas e pretônicas, assim como a postônica final e a medial de uma proparoxítona (O5 ~ O6), mas passa a distinguir também a sílaba acentuada de uma oxítona da sílaba acentuada de uma paroxítona (T1 ~ T2). Intensidade ganha o papel de contrastar as acentuadas das pretônicas, mas perde o de contrastar a postônica final e a medial de uma proparoxítona (O5 ~ O6). F0 muda quase que radicalmente seu poder distintivo: mantém apenas o papel de contrastar sílabas acentuadas e postônicas e adicionalmente passa a diferenciar a postônica de paroxítonas e a postônica final de proparoxítonas (O4 ~ O5), a postônica final e a medial de uma proparoxítona (O5 ~ O6), pretônicas adjacente e não adjacente à sílaba acentuada (E7 ~ E8).

Nas medidas articulatórias, vemos que CQ e Op deixam de exercer qualquer papel contrastivo, ao passo que Cp, além de sílabas acentuadas e postônicas, passa a diferenciar também postônicas e pretônicas.

6.2.7 Discussão dos resultados

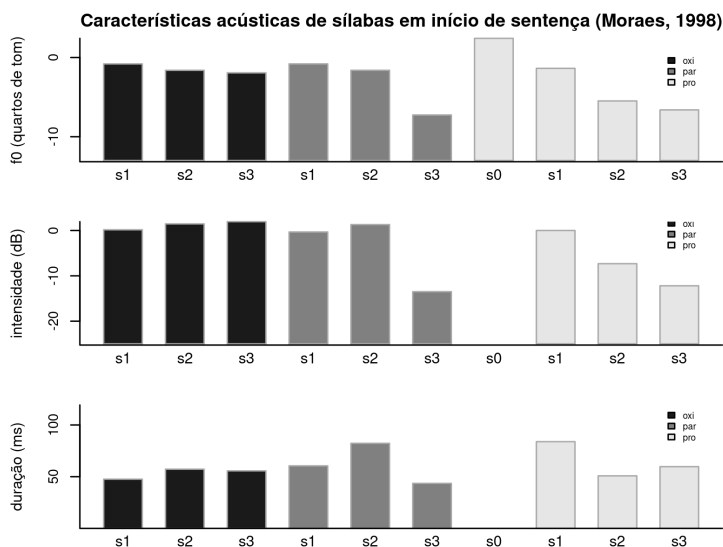
Medidas acústicas

A partir dos resultados obtidos neste experimento, verificamos que a propriedade acústica mais significativa na produção do acento no PB, considerando os participantes testados, é a duração, corroborando os estudos de Fernandes (1977), Ferreira Neto (2007), Major (1985) e Moraes (1998). Essa é a única propriedade capaz de distinguir entre os três tipos de padrão acentual, tanto em sentenças neutras como em sentenças com foco: sílabas acentuadas são mais longas que as pretônicas e que as postônicas, enquanto as pretônicas são mais longas que as postônicas. Adicionalmente, a duração é capaz de distinguir as sílabas postônicas de proparoxítonas, sendo que as sílabas postônicas finais são mais longas que as mediais.

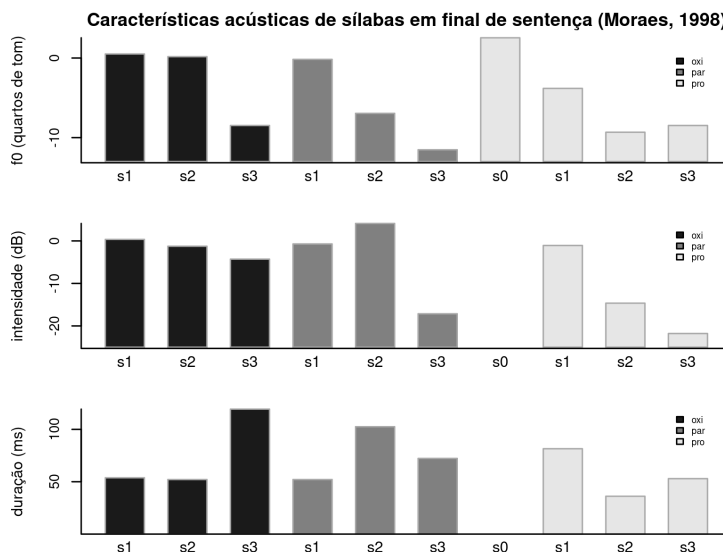
F0 também é relevante para a distinção dos três padrões acentuais, mas somente em sentenças neutras. Quando é alterada a estrutura prosódica associada à sentença, o que acontece nas sentenças com foco, F0 já não mais distingue os três padrões acentuais, mas passa a contrastar postônicas e pretônicas.

Mostrou-se significativa também a intensidade, mas com menor abrangência: sílabas acentuadas e pretônicas são mais intensas que postônicas. Ao lado da duração, intensidade também diferencia sílabas postônicas de uma proparoxítona: postônicas finais são mais intensas que as mediais.

Retomaremos agora os resultados de Moraes (1998), apresentados anteriormente nas FIG. 6 e 7 e recuperados a seguir por praticidade.

FIGURA 6 (retomada) – Propriedades acústicas do acento lexical em início de sentençaa) *F0* (quartos de tom); b) *Intensidade* (dB); c) *Duração* (cs)

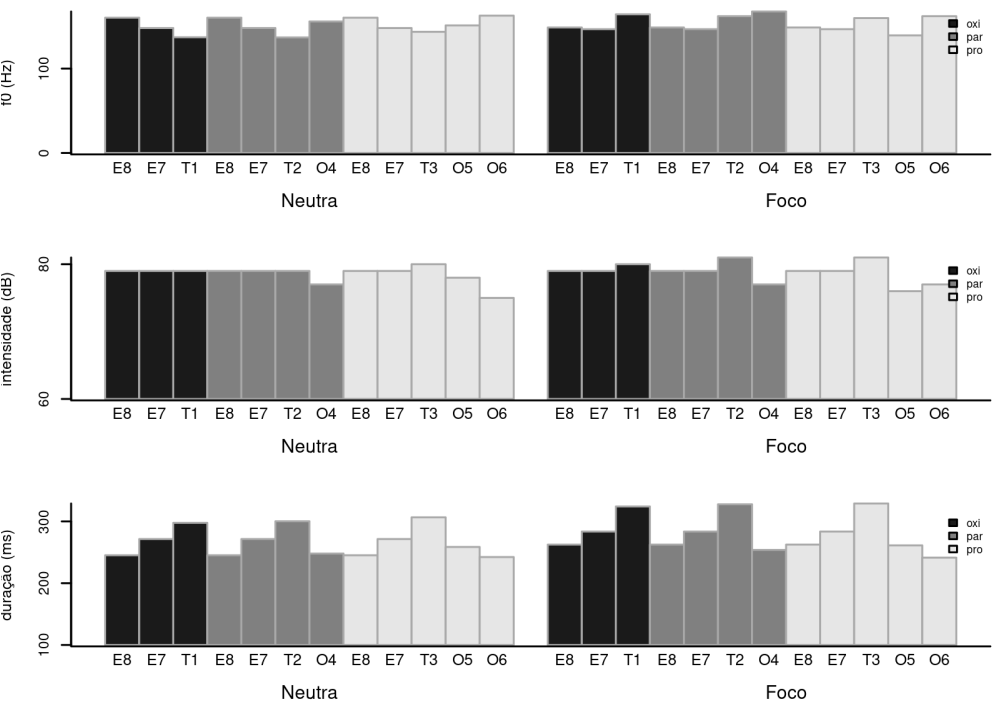
FONTE: MORAES, 1998, p. 180, adaptado. (Os dados de *f0* que consistiam em uma curva ao longo da sílaba foram substituídos por um único valor: a média de *f0* inicial e final naquela sílaba.)

FIGURA 7 (retomada) – Propriedades acústicas do acento lexical em final de sentençaa) *F0* (quartos de tom); b) *Intensidade* (dB); c) *Duração* (cs)

FONTE: MORAES, 1998, p. 181, adaptado. (Os dados de *f0* que consistiam em uma curva ao longo da sílaba foram substituídos por um único valor: a média de *f0* inicial e final naquela sílaba.)

A partir das medianas das FIG. 18 a 20, elaboramos gráficos semelhantes aos de Moraes, com a finalidade de facilitar a comparação dos resultados. Tais gráficos são mostrados a seguir, na FIG. 25. Note-se que os gráficos dos diferentes estudos não podem ser comparados diretamente entre si, pois encontram-se em unidades de medidas distintas e a influência da entoação sobre os parâmetros acústicos do acento foi avaliada de forma diferente entre os estudos (posição na sentença ou foco). Além disso, como nos dados deste estudo não foram diferenciadas as pretônicas de oxítonas, paroxítonas e proparoxítonas, os mesmos valores médios foram usados nos três padrões acentuais.

FIGURA 25 – Propriedades acústicas do acento obtidas experimentalmente



Tipo de proeminência		Localização da proeminência	
T	acentuada	1	oxítona
O	postônica	2	paroxítona
E	pretônica	3	proparoxítona
		4	postônica final única
		5	postônica final antecedida por postônica medial
		6	postônica medial
		7	pretônica adjacente à acentuada
		8	pretônica não adjacente à acentuada

Optamos por unir em um mesmo gráfico os valores correspondentes às sentenças neutras e com foco, ao contrário de Moraes, que separou em gráficos diferentes os padrões acentuais em início (FIG. 6) e no fim da sentença (FIG. 7). Na FIG. 25, mantivemos os rótulos que viemos adotando para sílabas pretônicas (E), acentuada (T) e postônicas (O):

- a) pretônicas: adjacente à acentuada (E7), antes da adjacente (E8)
- b) acentuadas: em oxítônica (T1), em paroxítônica (T2), em proparoxítônica (T3)
- c) postônicas: em paroxítônica (O4), final em proparoxítônica (E5), medial em proparoxítônica (E6)

Moraes, em seus gráficos, optou por numerar as três últimas sílabas de cada padrão acentual, em ordem crescente da esquerda para direita, como s1, s2, s3. Com isso, mostra duas pretônicas nas oxítonas, uma nas paroxítonas e nenhuma nas proparoxítonas (exceto no gráfico de F0, que inclui uma sílaba extra, anterior à acentuada, nomeada s0). Ainda que não haja uma correspondência perfeita com os rótulos de Moraes, as diferenças não impedem a comparação global de seus resultados com os obtidos no presente estudo.

Ao comparar as FIG. 6 e 7, retomadas acima, com a FIG. 25, vemos que há aspectos semelhantes e divergentes entre os resultados. Considerando primeiramente as semelhanças, os dois resultados indicam que, na posição sem proeminência entoacional (início de sentença/neutro):

- a) Duração: postônicas são em média mais curtas que as pretônicas, e ambas são mais curtas que as acentuadas; nas proparoxítonas, postônicas mediais são mais curtas que as finais
- b) Intensidade: postônicas são reduzidas em intensidade com relação às acentuadas e pretônicas e, nas proparoxítonas, a postônica medial tem maior intensidade que a final
- c) F0: há uma queda em F0 antes da sílaba acentuada

Na posição com proeminência entoacional (fim de sentença/foco), ambos os estudos indicam um aumento significativo dos contrastes entre as sílabas acentuadas e não acentuadas, em decorrência da sobreposição da proeminência entoacional sobre a acentual.

Quanto às diferenças entre os resultados, podemos mencionar, em nossos dados, uma diferença significativa, nas oxítonas, entre as duas sílabas pretônicas e a acentuada, tanto em termos de duração, quanto em termos de F0. Quanto a F0, observa-se em nossos dados, nas proparoxítonas, um crescimento da curva de F0 ao longo das sílabas postônicas, ou seja, da tônica em direção ao final da palavra. Esses dois aspectos divergentes podem ser decorrentes de diferenças dialetais, uma vez que os dados de

Moraes são da variedade carioca e os nossos, da variedade mineira, que tem como uma das características marcantes de sua fonologia a alta incidência de fenômenos redutivos em sílabas átonas. Os fenômenos redutivos serão melhor explorados mais à frente, quando discutirmos resultados dos experimentos de percepção.

Por fim, observamos em nossos dados que, nas sentenças neutras, há uma curva de F0 que decresce da pretônica para a sílaba acentuada e torna a subir em direção à postônica. Já nas sentenças com foco, a curva tem o mesmo formato, mas sua mudança de trajetória é antecipada, começando a subir em direção à postônica ainda na sílaba acentuada. Os dados de Moraes não mostram o mesmo padrão, talvez porque avaliamos as duas pretônicas mais próximas à sílaba acentuada em todos os padrões acentuais.

Medidas articulatórias

Passaremos, então, a considerar as medidas articulatórias. Retomando as TAB. 3 a 9, que mostram os resultados de testes de comparações múltiplas, observamos que há uma menor relação entre as medidas do EGG e o acento, em comparação com as medidas acústicas. SI (índice de velocidade) não foi significativo em nenhuma circunstância. CQ (quociente de contato) se mostrou significativo, nas sentenças neutras, para diferenciar acentuadas vs. postônicas e acentuadas vs. pretônicas. Cp (fase de fechamento) e Op (fase de abertura) foram significativos nas sentenças neutras, no contraste entre acentuadas vs. postônicas: Cp é menor nas sílabas acentuadas e pretônicas que nas postônicas, enquanto Op é maior no mesmo contexto. Uma vez que SI é uma medida obtida a partir de uma relação entre Cp e Op⁸, a diminuição de um, acompanhada pelo aumento do outro, pode ter sido um dos fatores que levaram à não significância de SI.

Um de nossos objetivos neste experimento era relacionar propriedades acústicas relevantes na produção do acento a propriedades articulatórias em nível glótico. As propriedades acústicas que identificamos como principais candidatos a uma possível correlação foram intensidade e F0, uma vez que podem ser controladas por ajustes na glote. Contudo, não pudemos observar em nossos resultados uma relação sistemática entre nenhuma das variáveis articulatórias e as acústicas. CQ é significativo quando intensidade não é (T ~ E) e vice-versa (O ~ E). Sempre que CQ é significativo, F0 também o é, mas o contrário não é verdadeiro. O mesmo acontece com Cp e Op.

O fato de que as medidas articulatórias obtidas a partir do EGG não acompanhem as medidas acústicas talvez possa ser explicado como decorrente dos mecanismos articulatórios e aerodinâmicos empregados para alteração de F0 e intensidade: se essas propriedades acústicas foram modificadas pelos participantes assintematicamente, tanto

⁸ Retomando: $SI = (Cp - Op) / (Cp + Op)$

pelo recurso à pressão subglótica, quanto por ajustes nas cordas vocais, podem ter sido mascaradas as consequências de cada mecanismo no sinal do EGG. Infelizmente, não é possível testar essa hipótese com os dados de que dispomos no momento, uma vez que não foram realizadas medidas de pressão subglótica.

Iremos finalizar a discussão dos resultados deste experimento avaliando as medidas articulatórias CQ, SI, Cp e Op que obtivemos com os valores de referência adotados (cf. seção anterior). As TAB. 10 a 13, a seguir, apresentam valores médios de cada uma das medidas do EGG, por participante e por sílaba, apenas para as sentenças neutras, que são compatíveis com a circunstância de produção dos valores de referência adotados.

TABELA 10 – Valores médios de CQ (%)

Sílaba	Sujeito	T	E	O
GA	E	57,01	51,60	50,88
	G	53,35	52,32	52,72
	J	56,75	51,04	53,76
	R	59,99	54,97	51,10
	V	54,17	52,16	50,14
KU	E	56,59	56,92	55,13
	G	59,37	61,13	60,00
	J	55,95	55,45	51,66
	R	57,56	51,88	49,68
	V	58,54	55,23	57,31
PA	E	57,36	55,90	55,78
	G	51,17	46,64	52,79
	J	56,28	54,88	54,97
	R	59,22	56,97	53,00
	V	54,30	55,49	58,74
PI	E	56,77	55,79	57,57
	G	55,70	51,06	51,73
	J	53,83	54,37	53,85
	R	58,87	52,56	50,93

Como podemos ver na TAB. 10, os valores médios de CQ estão de acordo com o valor de referência ($\approx 50\%$), indo de pouco mais que 45% a pouco mais que 60%. Interessante notar que os valores médios mais altos e mais baixos pertencem ao mesmo participante (G). De fato, alguns participantes apresentam maior variabilidade nos valores médios de CQ.

TABELA 11 – Valores médios de SI (u.a.)

Sílaba	Sujeito	T	E	O
GA	E	-0,65	-0,65	-0,59
	G	-0,43	-0,36	-0,23
	J	-0,24	-0,28	-0,24
	R	-0,33	-0,30	-0,20
	V	-0,32	-0,37	-0,31
KU	E	-0,63	-0,61	-0,64
	G	-0,27	-0,51	-0,28
	J	-0,35	-0,29	-0,38
	R	-0,27	-0,30	-0,23
	V	-0,41	-0,24	-0,22
PA	E	-0,66	-0,62	-0,71
	G	-0,40	-0,27	-0,25
	J	-0,32	-0,31	-0,31
	R	-0,32	-0,26	-0,14
	V	-0,36	-0,33	-0,40
PI	E	-0,62	-0,68	-0,68
	G	-0,28	-0,12	-0,20
	J	-0,30	-0,33	-0,26
	R	-0,27	-0,35	-0,38
	V	-0,35	-0,45	-0,43

Quanto a SI, podemos ver, na TAB. 11, que os valores médios de quatro dos participantes se distanciam do valor de referência ($\approx -0,65$), e giram em torno de $-0,31$, sempre negativos. O quinto participante (E) apresentou valores que contrastam com os valores dos demais participantes e são bem próximos ao valor de referência.

TABELA 12 – Valores médios de Cp (%)

Sílaba	Sujeito	T	E	O
GA	E	8,55	8,12	9,11
	G	16,66	18,66	22,98
	J	21,71	19,10	21,85
	R	19,67	20,57	21,88
	V	19,51	15,59	18,19
KU	E	8,64	9,82	8,00
	G	23,46	22,38	24,15
	J	17,09	19,62	17,99
	R	20,66	18,02	19,80
	V	17,24	21,96	23,80

Sílaba	Sujeito	T	E	O
PA	E	7,97	9,03	6,59
	G	16,71	19,60	23,30
	J	19,07	18,71	20,10
	R	19,74	20,76	24,74
	V	18,41	19,46	17,18
PI	E	9,55	7,55	7,67
	G	21,79	26,01	27,56
	J	19,53	17,34	20,67
	R	20,12	17,82	18,24
	V	20,80	18,45	21,27

A TAB. 12 aponta que, à semelhança do que observamos para SI, os valores de Cp também se mostraram participantes à variação individual: os valores médios dos mesmos quatro participantes se distanciam do valor de referência (5 a 10%), girando em torno de 21%. O mesmo participante (E) apresentou valores que contrastam com os valores dos demais participantes e são mais próximos ao valor de referência.

TABELA 13 – Valores médios de Op (%)

Sílaba	Sujeito	T	E	O
GA	E	40,19	35,46	34,17
	G	39,59	35,65	34,91
	J	34,85	32,04	33,93
	R	38,05	37,38	32,27
	V	34,32	34,32	33,07
KU	E	36,86	38,18	37,66
	G	38,75	34,29	33,88
	J	34,71	35,67	32,26
	R	33,72	31,66	30,62
	V	36,84	35,21	35,61
PA	E	37,41	36,24	37,26
	G	35,02	32,85	36,27
	J	36,59	35,50	36,88
	R	36,88	34,92	31,50
	V	36,41	37,63	40,07
PI	E	39,59	39,50	40,16
	G	37,29	32,32	32,51
	J	34,77	34,48	35,07
	R	34,81	36,24	32,70
	V	41,61	37,08	38,77

Na análise da TAB. 13, verificamos que os valores médios de Op giram em torno de 35%, conforme o valor de referência (33 a 39%). Nesse caso, se existem tendências individuais, não são evidentes. Contudo, os valores bem distribuídos de Op podem ajudar a explicar tendências individuais que observamos em SI, na TAB. 11: uma vez que SI é calculado a partir da relação entre Cp e Op, pode-se concluir que o contraste dos valores de SI do participante E com os demais participantes decorre na verdade de uma diferença em Cp, já que os valores de Op são relativamente homogêneos para todos os participantes entre si.

Os valores médios de CQ, SI, Cp e Op indicam tendências gerais comuns a todos os participantes, em termos da relação dessas variáveis com o acento. Verifica-se também a ocorrência de variação individual e, especialmente, um comportamento divergente de um participante em algumas medidas. Tal comportamento divergente pode ser reflexo de estratégias fonatórias diferentes empregadas na realização de uma mesma propriedade acústica, e pode ter contribuído para que as medidas fonatórias obtidas pelo EGG não se mostrassem tão relacionadas às medidas acústicas como esperado inicialmente.

6.2.8 Conclusões

Este experimento teve como objetivo avaliar as propriedades acústicas e fonatórias envolvidas na realização do acento. Procuramos testar a hipótese de que parâmetros acústicos e fonatórios produzidos nas sílabas acentuadas são diferentes das sílabas pretônicas e postônicas. Foram avaliadas três propriedades acústicas (duração; F0; intensidade) e quatro propriedades fonatórias obtidas por meio de eletroglotografia (quociente de contato, CQ; índice de velocidade, SI; fase de fechamento, Cp; fase de abertura, Op). Foram analisadas sílabas acentuadas, postônicas e pretônicas, com e sem ênfase prosódica (foco).

Especificamente, esperava-se que, em comparação com as sílabas átonas, as sílabas acentuadas apresentassem maior duração, maior intensidade e F0 menor nas sentenças neutras e F0 maior nas sentenças com foco. Quanto às medidas específicas do EGG, esperava-se, nas sílabas acentuadas, um maior CQ, menor SI, menor Cp e maior Op, com relação às átonas.

Destacaremos a seguir os aspectos mais importantes dos resultados obtidos neste experimento. Os dados avaliados indicam que a propriedade acústica mais significativa na produção do acento no PB é a duração, como apontado por estudos anteriores sobre a acentuação no PB, como Fernandes (1977), Ferreira Neto (2007), Major (1985), Moraes (1998). Em nossos resultados, duração foi a única propriedade capaz de distinguir entre os três tipos de padrão acentual, tanto em

sentenças neutras como em sentenças com foco: sílabas acentuadas são mais longas que as pretônicas e que as postônicas, enquanto as pretônicas são mais longas que as postônicas. Adicionalmente, a duração é capaz de distinguir as sílabas postônicas de proparoxítonas, sendo que as sílabas postônicas finais são mais longas que as mediais.

F0 também se mostrou relevante para a distinção dos três padrões acentuais, mas somente em sentenças neutras. Quando era alterada a estrutura prosódica associada à sentença (i.e., nas sentenças com foco), F0 já não mais distinguia os três padrões acentuais.

Mostrou-se significativa também a intensidade, mas com menor abrangência: sílabas acentuadas e pretônicas são mais intensas que postônicas. Ao lado da duração, intensidade também diferencia sílabas postônicas de uma proparoxítona: postônicas finais são mais intensas que as mediais.

Comparamos os resultados obtidos com os de Moraes (1998) e pudemos observar vários aspectos semelhantes. Na condição sem ênfase prosódica, ambos os estudos indicam que as sílabas acentuadas são mais longas que as pretônicas e estas, mais longas que as postônicas. Entre as postônicas de uma proparoxítona, as finais são mais longas que as mediais. Sílabas acentuadas e pretônicas apresentam maiores valores de intensidade que as postônicas. Há uma queda em F0 antes da sílaba acentuada. Na posição com ênfase prosódica, ambos os estudos indicam um aumento significativo dos contrastes entre as sílabas acentuadas e não acentuadas, em decorrência da sobreposição da proeminência entoacional sobre a acentual. Curiosamente, nas sentenças com foco, verificamos que algumas propriedades acústicas mudaram o papel contrastivo que desempenhavam nas sentenças neutras.

A principal diferença que observamos em nossos resultados com relação aos de Moraes foi a ocorrência de uma curva de F0 decrescente da pretônica até a sílaba acentuada e crescente da acentuada até a postônica, nas sentenças neutras. Nas sentenças com foco, a curva tinha o mesmo formato, mas sua mudança de trajetória era antecipada, começando a subir ainda na sílaba acentuada. Os resultados de Moraes não mostram o mesmo padrão, talvez porque seus dados não contemplam as duas pretônicas mais próximas à sílaba acentuada em todos os padrões acentuais, como avaliamos em nossos dados.

Ao contrário do que era esperado inicialmente, não observamos em nossos resultados uma relação sistemática entre nenhuma das variáveis fonatórias e as propriedades acústicas. Identificamos menos relações entre as variáveis fonatórias e o acento do que entre as variáveis acústicas e o acento.

Constatou-se certa variabilidade individual e valores divergentes de alguns participantes nas medidas fonatórias obtidas a partir do EGG. Os valores divergentes foram interpretados como decorrentes de diferentes estratégias fonatórias empregadas para realizar uma mesma propriedade acústica. Sugerimos que essa explicação poderia justificar também o fato de não termos verificado as correspondências diretas esperadas entre propriedades fonatórias e acústicas.

A principal implicação teórica deste experimento para o presente estudo diz respeito ao papel da duração como propriedade acústica mais relevante na produção do acento, papel compartilhado por outras propriedades acústicas, como F0 e intensidade. Segundo Repp (1982), estabelece-se uma relação de troca (*trading relationship*) entre variáveis que desempenham a mesma função contrastiva, de forma que alterações em uma das variáveis podem ser compensadas pelas outras de mesmo papel. Por exemplo, a sensação térmica, que depende tanto da temperatura quanto da umidade relativa do ar, envolveria uma relação de troca entre essas duas variáveis: ao aumentar a temperatura, pode-se abaixar a umidade de forma tal a obter uma mesma sensação térmica. Argumentamos que o acento no PB envolve uma relação de troca entre duração, F0 e intensidade, em um mapeamento de muitos para um.

Da interação entre acento e foco, pudemos verificar que as propriedades acústicas envolvidas na produção do acento estabelecem uma relação complexa em termos da sinalização de proeminência no nível da palavra (acentuado) e em nível superior a ela (foco). A proeminência prosódica afeta as propriedades acústicas envolvidas no acento não só aumentando os contrastes já existentes entre os padrões acentuais, mas também alterando o tipo de contraste exercido por cada propriedade acústica. Com isso, uma propriedade acústica (e.g. F0) que em determinada condição prosódica é determinante na diferenciação dois padrões (e.g. acentuados vs. pretônicos) pode deixar de realizar esse contraste em outra condição prosódica (e.g. foco). F0 oferece um exemplo claro de influência da prosódia na proeminência acentual, uma vez que essa propriedade serve para estabelecer mais de um contraste sonoro no PB (mapeamento um para muitos). Em nossa abordagem, os mapeamentos um para muitos e muitos para um que envolvem as propriedades acústicas duração, intensidade e F0 decorrem da auto-organização do sistema e são viabilizados pela relação de troca estabelecida entre essas propriedades.

De fato, uma das consequências do estabelecimento de uma relação de troca, em um modelo multirrepresentacional, é a possibilidade de que uma propriedade passe a exercer mais de um papel no sistema, em vista da redundância das pistas

sonoras. Argumentaremos, a seguir, que a duração, no PB, passou a exercer outros tipos de contrastes, por estar em relação de troca com intensidade e F0.

No capítulo 5, apresentamos uma proposta de abordagem multirrepresentacional do acento no português que partia do sistema latino. Nessa proposta, a acentuação do português teria emergido com a auto-organização de variáveis do sistema acentual latino em decorrência de uma série de mudanças fonológicas desencadeadas pela perda da quantidade vocálica. Defendemos que a presença de mais de uma pista acústica para a sílaba acentuada (relação de troca) possibilitou que duração passasse a exercer um papel relevante no contraste entre formas reduzidas por fenômenos de cancelamento final e formas não reduzidas, como apontaremos a seguir.

O latim não tinha oxítonas, que emergiram no português por meio de processos fonológicos de cancelamento final e, especialmente no português brasileiro, por meio do acréscimo lexical de palavras provenientes de línguas indígenas. No PB, fenômenos fonológicos de cancelamento de sílabas átonas continuam em plena atividade, sendo favorecidos por alguns contextos sonoros (cf. os estudos de Souza (2012) e Leite (2006) sobre o cancelamento de vogais pretônicas e postônicas seguidas de sibilantes). Por exemplo, palavras paroxítonas, como *passe*, [‘pasi], são transformadas em oxítonas, com a perda da vogal postônica: [‘pas]. Surgem potenciais homófonos, como *paz* [‘pas] e *passe* [‘pas]. Contudo, essa homofonia é apenas aparente. Sabemos de estudos como Souza (2012) e Leite (2006) que a duração opera como detalhe fonético fino no fenômeno de redução vocálica, mantendo a distinção entre as palavras: *paz* [‘pas] apresenta vogal com menor duração que a vogal de *passe* [‘pa:s] reduzido. A alternância deixa de ser entre presença e ausência de vogal final ou entre oxítona e paroxítona, passando a ser determinada pela duração – de certa forma retomando a propriedade de quantidade vocálica existente anteriormente no latim.

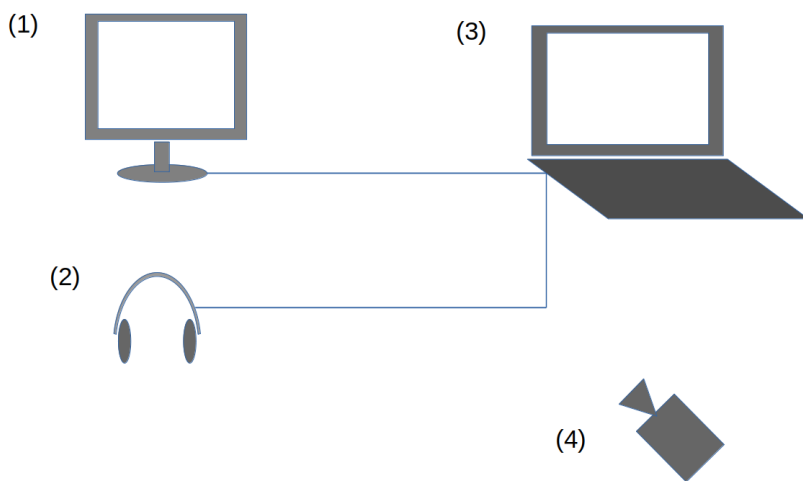
Há consequências da relação de troca também para a percepção, que iremos avaliar nos experimentos específicos descritos na seção seguinte. Os resultados obtidos neste experimento oferecem contribuições para a compreensão dos correlatos acústicos e fonatórios do acento no PB e sua interação com aspectos entoacionais e outros fenômenos sonoros da língua.

6.3 EXPERIMENTOS DE PERCEPÇÃO

Por meio de uma série de três experimentos de percepção, este estudo investigou os mecanismos de processamento e categorização do acento lexical no português brasileiro.

Mais especificamente, buscou-se avaliar possíveis influências da estrutura linguística na percepção de proeminência em padrões sonoros não linguísticos (experimento 1) e linguísticos (experimentos 2 e 3). Os dois primeiros experimentos se baseiam na desambiguação de sequências sonoras ambíguas, enquanto o terceiro, na identificação de um intervalo de estímulos progressivamente distintos. Os três experimentos utilizaram um mesmo projeto experimental, representado pela FIG. 26, a seguir.

FIGURA 26 – Projeto experimental utilizado nos três experimentos de percepção



Os participantes se posicionavam em frente ao monitor (1), que apresentava os estímulos visuais, portando os fones de ouvido (2), pelos quais escutavam os estímulos auditivos. Os experimentos eram controlados pelo computador (3), com extensão de tela no monitor (1), e gravados em vídeo, por meio de câmera digital (4), para documentação. Uma pequena alteração na montagem foi introduzida no terceiro experimento, que continha tarefa auto-monitorada (*self-pacing*), como veremos adiante. Neste caso, foi retirado o monitor (1) em tela estendida; os participantes receberam os estímulos visuais e autocontrolaram o experimento diretamente do computador (3).

6.3.1 Experimento 1: desambiguação de padrões não linguísticos

O primeiro experimento de percepção foi projetado de forma a oferecer resultados comparáveis aos de Bell (1977), que estudou possíveis influências da estrutura linguística na interpretação prosódica de estímulos não linguísticos. O autor testou a percepção de combinações simples de tons puros que simulavam a proeminência acentual em grupos de falantes de cinco línguas diferentes. Como o português não se encontrava entre as sete línguas testadas pelo autor, um de nossos objetivos com o presente experimento é avaliar o comportamento dessa língua entre os resultados do estudo original.

Estruturas rítmicas são formadas pela repetição de sequências de sons em que alternam elementos semelhantes e diferentes. Sabemos de estudos em psicoacústica que sequências de sons tendem a ser quebradas (ou *parsed*) em agrupamentos menores, em que um dos elementos geralmente é mais saliente (HANDEL, 1993). Como relata o autor, os agrupamentos tendem a ser compostos por dois, três ou quatro elementos. Um agrupamento binário é formado de dois elementos diferentes, geralmente sendo a um deles conferida maior saliência cognitiva. O elemento mais saliente pode se posicionar no início (i.e. sw⁹) ou no fim da sequência (i.e. ws). Por sua vez, um agrupamento ternário é composto de dois elementos iguais e um diferente, podendo este último, o mais saliente, se localizar no início (i.e. sww), meio (i.e. wsw) ou fim da sequência (i.e. wws). Esses dois tipos de agrupamentos, binário e ternário, são os de maior interesse para o presente estudo, uma vez que guardam correspondência com os padrões acentuais mais frequentes no PB.

Bell (1977) identificou em seus resultados uma preferência geral para a acentuação inicial dos grupos na sequência controle, em que nada variava, assim como nas sequências em que intensidade e frequência eram alteradas. Por outro lado, houve uma tendência generalizada à acentuação medial dos grupos na sequência em que a duração variava. Analisando os resultados obtidos, o autor afirma que “the tendencies that appear for each language do not follow the pattern of the language’s accent placement” e concluiu que “the effect of accentual structure upon rhythmic perception of non-speech sounds, if it exists, is unlikely to be of the same magnitude as the general psychophysical effects discussed earlier.” (BELL, 1977, p. 7). Por outro lado, Jakobson, Fant e Halle (1952), mencionados por Bell (1977), defendem que o sistema acentual da língua de um falante exerce influências sobre a percepção de padrões rítmicos em geral, mesmo os não linguísticos. Considerando os modelos teóricos adotados neste estudo, para os quais as representações são multirrepresentacionais e a linguagem se baseia em processos cognitivos gerais, espera-se que a percepção não linguística esteja sujeita a influências da estrutura linguística.

Este primeiro experimento de percepção pretende testar as seguintes hipóteses: (1) a percepção de padrões sonoros é afetada diferentemente por alterações nas propriedades acústicas; (2) em padrões não linguísticos contínuos do tipo ternário, há uma maior probabilidade de percepção do padrão com proeminência medial (ws_w, i.e. *weak-strong-weak*), correspondente ao padrão paroxítono, que apresenta maior frequência de tipo no PB. Esta segunda hipótese equivaleria a testar se padrões rítmicos da língua influenciam a percepção rítmica (linguística e não linguística), como defendido por Jakobson, Fant e Halle (1952). No contexto dos modelos teóricos adotados neste estudo – em especial, os Modelos de Uso (BARLOW; KEMMER,

⁹ Nos padrões sonoros, a relação de saliência cognitiva será marcada por meio da notação “s”, para saliente (“*strong*”) e “w”, para não (ou menos) saliente (“*weak*”), ainda que não esteja em jogo necessariamente uma relação de força.

2000; BYBEE, 2001), para os quais a frequência de uso desempenha um papel crucial na construção da gramática do falante –, o presente experimento tem potencialmente como interesse adicional avaliar efeitos de frequência de tipo sem influência da frequência de ocorrência, uma vez que os estímulos não são linguísticos.

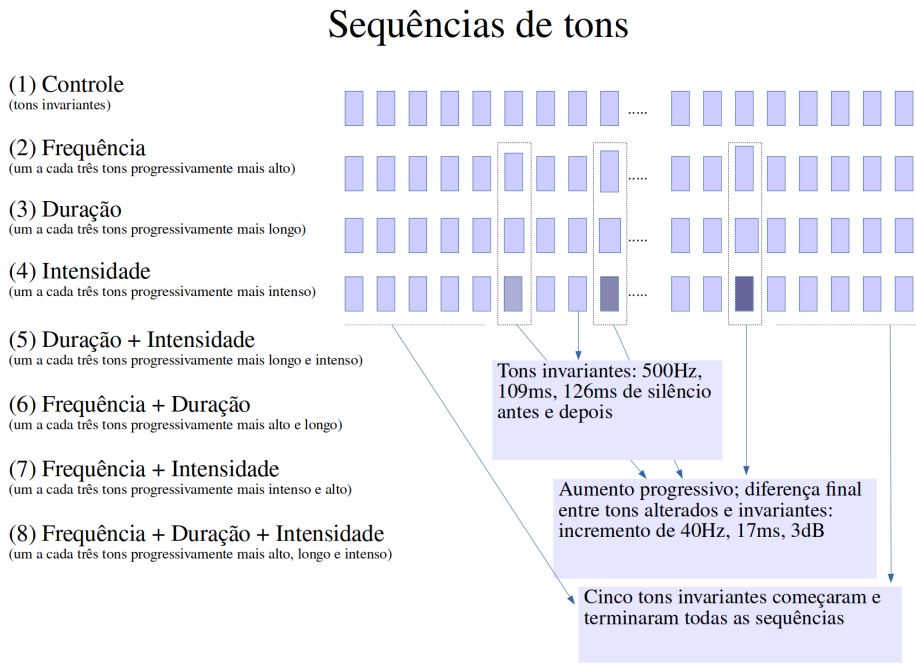
6.3.1.1 Participantes

Foram avaliados 12 participantes, sendo 6 homens e 6 mulheres, todos com nível universitário, na faixa etária entre 20 e 30 anos. Todos são falantes nativos do português e não relataram histórico de problemas de audição ou fala. Além disso, todos os participantes são leigos em teoria e instrumentos musicais. Este fator foi controlado, em vista de possíveis diferenças no desempenho de músicos em comparação a não músicos em tarefas envolvendo estímulos não linguísticos. Os mesmos voluntários participaram deste experimento e do experimento 2, em uma mesma sessão de coleta.

6.3.1.2 Estímulos

Os estímulos consistiram em oito sequências de tons puros, sendo uma sequência controle (em que nada variava entre os elementos do grupo) e sete sequências testes, nas quais um tom a cada três variava, simultaneamente ou não, em duração, frequência e/ou intensidade. A FIG. 27, abaixo, ilustra esquematicamente cada sequência de tons construída.

FIGURA 27 – Sequências de tons utilizadas como estímulos no experimento de percepção 1

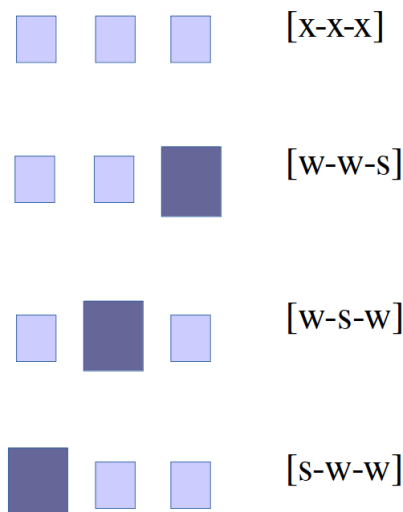


6.3.1.3 Procedimentos

O teste realizado no presente experimento segue a metodologia de Bell, com algumas adaptações. A tarefa consistiu em escutar uma sequência de tons ambígua em termos de *parsing* da proeminência (e.g. ...wwswswswsw...→...[wws][wws][wws]ww... ou ...w[ws w][ws w][ws w]w... ou ...ww[sww][sww][sww]...) e em seguida realizar a decisão forçada por um entre quatro padrões apresentados, nos quais variava o elemento mais proeminente. Neste experimento, os participantes levaram cerca de 15 minutos para completar a tarefa solicitada.

Os participantes escutavam, por meio de fones de ouvido, sete tipos de sequências de três tons – como mostra a FIG. 27, acima –, que variavam sistematicamente em F0, intensidade e/ou duração. Cada sequência foi formada por dois tons invariáveis e um terceiro tom, variável, alterado em um ou mais parâmetros acústicos. Entre cada tom havia um intervalo de silêncio de 126ms nos tons invariantes, sendo reduzido apenas nos tons em que foi alterada a duração, de forma a manter constante, em todas as sequências, a soma da duração dos tons com os intervalos entre eles, como fez Bell (1977). Quando tocadas sucessivamente, as sequências apresentam ambiguidade em sua segmentação com relação à localização do tom variável perante os dois invariantes. Após escutar cada sequência de tons, os participantes deviam categorizar cada sequência ouvida, apontando para um dos quatro diagramas apresentados simultaneamente na tela de um computador. Tais diagramas correspondiam à estrutura xxx; wws; wsw; sww (sendo x = tom invariável; s = tom saliente; w = tom não saliente), como mostra a FIG. 28, na sequência.

FIGURA 28– Estímulos visuais utilizados no experimento de percepção 1



6.3.1.4 Resultados

Neste experimento, a variável dependente é a resposta ao padrão sonoro como neutro (XXX), proeminência final (WWS), proeminência medial (WSW), ou proeminência inicial (SWW). A variável independente é a natureza acústica da proeminência do padrão sonoro, que poderia ser realizada apenas por frequência (F), duração (D) ou intensidade (I), isoladamente, ou por meio das interações entre os três níveis: FD, FI, DI, FDI. A variável “natureza acústica” contém ainda um último nível, referente ao padrão sonoro controle (C), em que não havia relação de proeminência.

Os dados foram ajustados por um modelo generalizado linear com três fatores fixos¹⁰, representado em (i). Como se pode ver, os níveis da variável independente “natureza acústica” foram inseridos individualmente no modelo, transformados em fatores fixos.

(i) resposta ~ frequência + duração + intensidade

Em decorrência de restrições do pacote de funções utilizado para análise dos dados, que não permite a incorporação de fatores aleatórios, o modelo contém apenas os três fatores fixos frequência, duração e intensidade. Ou seja, os informantes não puderam ser considerados como fator aleatório no modelo, o que poderia ter prevenido uma possível inflação dos graus de liberdade dos resíduos.

Adotou-se um modelo generalizado porque a variável dependente são proporções de escolha de cada padrão sonoro, com distribuição de erros multinomial: uma vez que as respostas poderiam ser de quatro tipos diferentes, a variável dependente é um vetor de proporções das três probabilidades de resposta referentes aos padrões acentuais: P(WWS), P(WSW) e P(SWW). Obtém-se a probabilidade de resposta referente ao padrão neutro a partir das outras três probabilidades: $P(XXX) = 1 - P(WWS) - P(WSW) - P(SWW)$.

Antes, contudo, de ajustar os dados ao modelo linear, foi necessário transformar os valores de probabilidades, que se restringem ao intervalo [0, 1], a um intervalo $[-\infty, \infty]$. Para tal, foi realizada a seguinte transformação logarítmica:

Após a transformação dos dados, o modelo linear foi ajustado, criando-se três vetores, um para F, um para D, outro para I, assim como um vetor intercepto, que corresponde a (C). Um dos benefícios do modelo (i) adotado é possibilitar a previsão do resultado dos padrões sonoros formados pela interação de efeitos a partir da soma dos efeitos individuais de F, D e I. É possível obter, por exemplo, o efeito de FD (frequência-duração) somando os efeitos de F e D. Por requerer um número menor de coeficientes, é preferível ao modelo (ii), que realiza o ajuste para cada um dos coeficientes das interações.

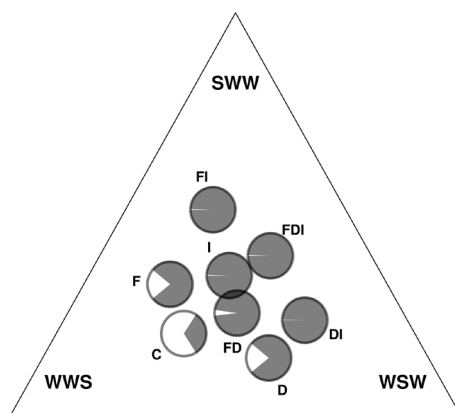
¹⁰ R: função *vglm*, pacote *VGAM*

(ii) resposta \sim frequência * duração * intensidade

Os modelos (i) e (ii) foram comparados por meio do teste de razão de verossimilhança¹¹, que respeita a distribuição de qui-quadrado. O valor p obtido é $> 0,813$, o que indica que os modelos não são significativamente diferentes entre si e justifica, portanto, a utilização do modelo mais econômico (i).

A FIG. 29, a seguir, mostra os resultados do ajuste ao modelo, na forma de um gráfico triangular, que representa a distribuição das proporções previstas para os efeitos do modelo e suas combinações.

FIGURA 29 – Distribuição das probabilidades de resposta de acordo com o padrão sonoro dos estímulos



F = frequência; D = duração; I = intensidade

Equivalências com padrões acentuais: WWS = oxítônica;

WSW = paroxítônica; SWW = proparoxítônica

Os vértices do gráfico correspondem às três possibilidades de resposta (SSW, WSW ou WWS) referentes à variável dependente. Há sete pontos dentro do triângulo que correspondem aos níveis da variável independente, referente à natureza da proeminência dos padrões sonoros (F, D, I, FD, FI, DI, FDI). O oitavo ponto corresponde ao intercepto (C), referente ao estímulo controle, em que não havia variação. Os pontos são posicionados no interior do triângulo de acordo com a probabilidade de escolha das respostas: a distância entre um ponto e o lado oposto

¹¹ R: função *pchisq*, pacote *stats*

a um vértice é proporcional a $1 - P$, sendo P a proporção de escolha de tal vértice. Ou seja, a probabilidade de que um estímulo seja reconhecido como um dos padrões é proporcional a sua proximidade ao vértice desse padrão. Dessa forma, é possível avaliar as diferenças entre as variáveis em termos de sua distância com relação ao ponto C e em termos da direção do vetor formado entre elas e C em relação aos vértices. Por fim, a cada ponto é sobreposto um “subgráfico” de pizza, cuja porção branca representa a probabilidade de escolha da resposta neutra (XXX).

Algumas tendências podem ser observadas na FIG. 29. Exploraremos as características dos vetores formados por cada variável e o intercepto (C). O vetor $C \rightarrow D$ se direciona para o vértice WSW. Isto indica que a mudança da condição Controle © à condição Duração (D) favorece o padrão WSW, com proeminência medial, correspondente às paroxítonas. Já o vetor $C \rightarrow F$ tem como direção $WSW \rightarrow SWW$. Ou seja, com relação à condição Controle (C), a condição Frequência (F) favorece as respostas SWW e desfavorece as respostas WSW. O vetor $C \rightarrow I$, por sua vez, aponta em direção oposta ao vértice WWS. Isto significa que a condição Intensidade (I) favorece as respostas WWS.

O modelo ajustado não contém interações entre as variáveis básicas (F, D, I), de forma que os padrões conjugados (FI, FD, DI e FDI) podem ser previstos em função dos três vetores de base: $C \rightarrow F$, $C \rightarrow D$ e $C \rightarrow I$.

Uma última observação a respeito da distribuição dos dados pode ser feita. Considerando as respostas ao padrão controle (C) diferentes de XXX, vemos que o mesmo se coloca mais próximo de WWS que das demais respostas.

6.3.1.5 Discussão dos resultados

Neste experimento, procurou-se testar as hipóteses de que (1) a percepção de padrões sonoros é afetada diferentemente por alterações nas propriedades acústicas; (2) padrões rítmicos da língua influenciam a percepção rítmica (linguística e não linguística). Enquanto a primeira hipótese pode ser testada diretamente, a segunda deve levar em consideração o sistema acentual do PB, cujas propriedades acústicas e fonatórias foram abordadas neste estudo por meio de experimento de produção. Sabemos da coleta de dados realizada junto ao ASPA (cf. sessão 5.2) que o padrão acentual paroxítono é o mais frequente no PB, representando 71,4% dos verbos e 67,1% dos não verbos. Portanto, a segunda hipótese equivale encontrar uma maior probabilidade de percepção do padrão com proeminência medial (WSW), correspondente ao padrão paroxítono. Essa hipótese vai ao encontro do que defendem Jakobson, Fant e Halle (1952) e contraria os resultados apresentados por Bell (1977).

Os resultados obtidos neste experimento indicam que as propriedades acústicas exercem influência na percepção da proeminência em padrões sonoros, em termos de tendências gerais. Observamos que, isoladamente, a frequência desfavorece a percepção do estímulo como proeminência medial, tendendo a favorecer a proeminência final. A duração favorece a percepção como proeminência medial. A intensidade desfavorece a percepção como proeminência final. O padrão controle, uma sequência sonora de sons idênticos, favorece a percepção de ausência de proeminência em mais de 60% das respostas, como esperado, mas, nos demais casos, favorece a percepção da proeminência final. O QUADRO 7, a seguir, resume as interpretações das relações entre estímulo e proeminência percebida apresentadas acima e as previsões para interações entre fatores.

QUADRO 7 – Tendências de favorecimento e desfavorecimento à percepção padrões

	SWW	WSW	WWS
F	favorece	desfavorece	
D		favorece	
I			desfavorece
FI	favorece		
FD			desfavorece
DI		favorece	desfavorece
FDI			desfavorece

F = frequência; D = duração; I = intensidade;

FI = frequência+intensidade; FD = frequência+duração;

DI = duração+intensidade; FDI = frequência+duração+intensidade

O QUADRO 7 indica que D e DI favorecem WSW, ao passo que F e FI favorecem SWW. Além disso, I, FD, DI, FDI desfavorecem WWS. Podemos apontar uma tendência de que, quando está em jogo a duração, os participantes tendem a perceber a proeminência medial, que corresponde ao padrão paroxítono, em detrimento da proeminência final, que corresponde à oxítônica. Por outro lado, quando estava em jogo a frequência, os participantes tendiam a perceber a proeminência inicial, correspondente à proparoxítona.

Retomando o estudo de Bell (1977), cuja metodologia foi adotada no presente experimento, há algumas diferenças entre nossos resultados e os obtidos pelo autor. Primeiramente, cabe lembrar que o autor não avaliou estímulos em que mais de uma propriedade variava, como fizemos. Bell identificou em seus resultados uma preferência geral de associar C à proeminência inicial, o que não aconteceu em nosso experimento – visto que na FIG. 29 o ponto correspondente

a Controle encontra-se mais próximo do vértice WWS, havendo, portanto, maior probabilidade de ser identificado como proeminência final. O autor verificou uma preferência pela proeminência inicial também para F e I, o que curiosamente não aconteceu em nossos estímulos equivalentes, com essas propriedades isoladas, mas aconteceu quando elas eram conjugadas, em FI. Por outro lado, Bell observou uma tendência generalizada à proeminência medial quando a duração variava, o que pode ser observado em nossos dados, estendendo-se ainda ao par conjugado DI.

Em uma análise geral, nossos resultados não evidenciam uma correspondência clara e direta entre os padrões de proeminência preferentemente percebidos em cada sequência sonora e os padrões acentuais do português, mas aponta tendências gerais condizentes com o sistema acentual dessa língua: as paroxítonas são o padrão mais frequente de acento no domínio da palavra, que tem como principal propriedade acústica a duração. Ressalte-se que, em tons puros, a percepção conta prioritariamente com propriedades físicas, ao passo que, nos eventos linguísticos, o significado é um fator determinante para a identificação dos padrões sonoros, também influenciada pelo uso linguístico. Portanto, no caso de experimentos de percepção envolvendo estímulos não linguísticos, mesmo que haja influência de estruturas específicas da língua, não se deve esperar relações diretas. É necessário considerar a força antagônica exercida pela inseparabilidade do elo arbitrário entre padrões sonoros e o significado, que tem consequências para a percepção: na ausência de significado linguístico, há uma tendência de que os padrões sonoros sejam interpretados como eventos não linguísticos – portanto, independentes da estrutura linguística. Sugerimos que a interação entre a força exercida pela estrutura linguística e a força exercida pela interpretação não linguística tem como consequência a emergência de padrões complexos de resposta. Ao lado de questões metodológicas, como o número reduzido de participantes, esse que pode ser um dos fatores responsáveis pelo resultado que Bell (1977) interpretou como evidência contrária à influência da linguagem na percepção de padrões rítmicos não linguísticos.

Ao analisar caso a caso, podemos identificar correspondências entre as tendências na percepção de padrões e a estrutura linguística do português: como sabemos do experimento de produção, apresentado na seção anterior, a duração é a propriedade acústica mais relevante na produção da proeminência acentual no português, que tem como padrão acentual mais frequente as paroxítonas. No presente experimento, houve uma tendência à percepção da proeminência medial, equivalente a paroxítonas, quando avaliamos a duração isoladamente ou conjugada com intensidade. Como menciona Cutler (2005), um resultado aproximado foi obtido por Turk e Sawusch (1996), que apontam para a duração ou a associação de amplitude e duração como propriedades relevantes em julgamentos de proeminência.

Do mesmo modo, F0, que exerceria funções variadas no sistema sonoro do PB, parece alternar ora entre o favor das oxítonas, ora em favor das proparoxítonas.

6.3.2 Experimento 2: desambiguação de padrões linguísticos

O segundo experimento de percepção teve como propósito avaliar a percepção da proeminência em sequências sonoras linguísticas. Este estudo foi publicado isoladamente na Revista Virtual de Estudos da Linguagem – ReVEL (CANTONI, 2017). O paradigma empregado é semelhante ao do primeiro experimento, mas trocando os tons puros por estímulos linguísticos. Com isso, além da frequência de tipo, torna-se possível acessar diretamente também possíveis influências da frequência de ocorrência na percepção de padrões sonoros. Pretende-se testar a seguinte hipótese: nos padrões linguísticos acentuais, a estrutura percebida sofre influência da frequência de ocorrência e de tipo das palavras testadas.

6.3.2.1 Participantes

Foram avaliados os mesmos participantes do primeiro experimento de percepção, na mesma sessão de coleta de dados: 12 participantes, sendo 6 homens e 6 mulheres, todos com nível universitário, na faixa etária entre 20 e 30 anos. Todos são falantes nativos do português e não relataram histórico de problemas de audição ou fala.

6.3.2.2 Estímulos

Os estímulos utilizados neste experimento seguiram o mesmo princípio do experimento anterior, que avaliou a percepção de sequências de tons puros ambíguas em termos de *parsing* da proeminência, mas, no presente caso, eram sequências de sons linguísticos, interpretáveis como palavras (e.g. ...pɔsipɔsipɔsipɔs... → ...pɔ[sipɔ][sipɔ][sipɔ][sipɔ]s... “cipó” ou ...[pɔsi][pɔsi][pɔsi][pɔsi]pɔs... “posse”). Foram selecionados como estímulos dez sequências de sílabas fracas e fortes com segmentação ambígua, ou seja, que pudessem ser interpretadas como dois padrões acentuais diferentes, e, consequentemente, como duas palavras distintas, a depender do *parsing* realizado no momento da percepção (e.g. ...sipɔsipɔsipɔsi... → “cipó”, oxítona, ou “posse”, paroxítona). Portanto, a cada estímulo, estavam associadas duas palavras quanto à possibilidade de identificação: uma oxítona e outra paroxítona. Considerando que a frequência de ocorrência poderia influenciar a identificação entre uma das duas palavras, foram selecionados pares de palavras com valores de ocorrência tanto próximos quanto muito distantes,

de modo a avaliar se haveria influência desse fator na identificação das palavras pelos informantes. A frequência de ocorrência das palavras, obtida a partir do banco de dado do Projeto Aspa, já mencionado anteriormente (cf. capítulo 4), é mostrada no QUADRO 8, a seguir, para cada uma das palavras associadas à cada estímulo. Como se vê na coluna “Estímulo”, foi adotado, para cada estímulo, um código de identificação que corresponde às três primeiras letras da palavra oxítônica associada a tal estímulo.

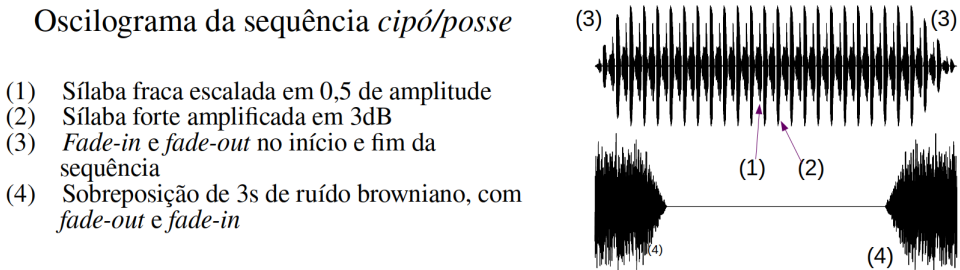
QUADRO 8 – Pares de estímulos utilizados no experimento de percepção 2 e seus valores de ocorrência

Estímulo	Oxítônica	Frequência de Ocorrência	Paroxítônica	Frequência de Ocorrência
man	manoele	7.958	noelma	16
dav	davi	1.373	vida	112.365
caj	caju	873	juca	1.922
tum	tumor	2.212	morto	20.475
dal	dali	2.161	lida	1.775
tup	tupi	1.420	pito	206
bab	babá	1.283	baba	432
mar	marrom	1.111	roma	9.689
mac	macau	878	calma	3.501
jab	jabuti	422	botija	43
fub	fubá	357	bafo	205
tut	tutor	312	torto	783
cip	cipó	271	posse	15.246
gui	guichê	227	chegue	2.470
cas	cassis	125	cisca	10
sum	sumô	127	moço	1.639
caq	caqui	107	quica	22
mas	maçom	20	soma	7.913
cja	cajá	59	jaca	127

Para construir cada estímulo, as sílabas que compunham as palavras ambíguas foram gravadas isoladamente e, em seguida, concatenadas. A diferença entre a sílaba forte e a sílaba fraca foi posteriormente reconstruída por meio de manipulação do sinal, utilizando como referência valores de intensidade e duração, respectivamente, das sílabas forte e fraca de uma palavra gravada separadamente. A locutora da gravação foi a própria pesquisadora – 27 anos, falante nativa da variedade de Belo Horizonte. Optou-se por gravar as sílabas isoladamente, e não as próprias palavras, a fim de evitar que possíveis detalhes fonéticos finos característicos de cada palavra

e decorrentes da ordenação das sílabas (e.g. transições de formantes) pudessem influenciar na tarefa de identificação dos estímulos. Uma vez gravadas as sílabas, foram realizados os seguintes procedimentos de manipulação, de maneira uniforme para todos os estímulos, por meio do *software* Audacity (AUDACITY Team, 2008), com objetivo de simular a relação de proeminência para cada estímulo, a sílaba forte da sequência foi amplificada em 3dB e a sílaba fraca teve, então, sua amplitude escalonada em 0,5 com relação à forte. Além disso, a duração das sílabas foi alterada por meio da função “Alterar tempo sem alterar tom”: as sílabas fortes passaram a ter 0,4 s e as fracas, 0,3 s. As sílabas forte e fraca foram concatenadas e repetidas 25 vezes cada uma. Por fim, duas medidas foram tomadas para mascarar o início e o fim das sequências, de forma a evitar que os participantes fossem influenciados pela ordem inicial ou final para identificação do estímulo, como relatado no experimento de Bell (1977). Primeiro, o início e o fim das sequências foram submetidos respectivamente a efeitos de *fade-in* (atenuação regressiva do sinal) e *fade-out* (atenuação progressiva do sinal). Segundo, a sequência foi transformada de mono para stereo e, no canal adicional, foram introduzidos 3 s de ruído browniano alinhados com o início e com o fim das sequências, tendo, respectivamente, seu final submetido a *fade-out* e início submetido a *fade-in*. A FIG. 30, a seguir, apresenta o sinal construído para um dos estímulos utilizados no experimento 2.

FIGURA 30 – Exemplo de estímulo utilizado no experimento de percepção 2



6.3.2.3 Procedimentos

Este experimento consistiu em um teste de identificação, em que os participantes escutavam estímulos auditivos e deveriam relatar que palavra escutaram, em tarefa de livre seleção. Os participantes escutavam, por meio de fones de ouvido, cada uma das sequências sonoras ambíguas. Em seguida, deviam identificar e relatar a palavra ouvida, e a resposta era anotada pela pesquisadora. Neste experimento, assim como no experimento de percepção 1, os participantes levaram cerca de 15 minutos para completar a tarefa solicitada.

Como a tarefa proposta era de livre seleção, foram obtidas respostas que não corresponderam a nenhuma das duas palavras associadas ao estímulo: um total de 34 casos de não correspondência, que não foram consideradas na análise dos dados. Para avaliar o desempenho da tarefa, foi calculada uma taxa de identificação por participante e por palavra. A taxa de identificação foi determinada como a porcentagem de respostas do participante ou da palavra que correspondessem a uma das duas palavras esperadas. Dois participantes apresentaram taxa de identificação próxima de 50% – o que significa que quase a metade de suas respostas não foram aproveitadas. Juntos, responderam por 17 do total de 34 palavras sem correspondência. Possivelmente seu desempenho foi comprometido por algum fator, por exemplo, por não terem compreendido bem a tarefa. Contudo, durante a análise dos dados, verificamos que os resultados não eram alterados significativamente com a retirada desses participantes, portanto, optamos por mantê-los na análise dos dados.

Com relação aos estímulos, observou-se que alguns apresentaram taxa de identificação maior que os demais. Os que apresentaram os maiores valores foram *dali-lida* (dal), *guichê-chegue* (gui) e *cassis-cisca* (cas). Durante a análise dos dados, verificamos que a exclusão desses estímulos também não alterou significativamente os resultados, motivo pelo qual optamos por mantê-los na análise.

6.3.2.4 Resultados

A variável dependente neste experimento é a palavra identificada para cada estímulo: uma oxítônica ou uma paroxítônica – “resposta paroxítônica”. A variável independente é a frequência de ocorrência das palavras associadas aos estímulos, que foi avaliada como a “frequência relativa da palavra paroxítônica”:

frequência absoluta da paroxítônica

(frequência absoluta da paroxítônica + frequência absoluta da oxítônica)

Os dados obtidos foram ajustados com base em um modelo generalizado linear com efeitos mistos¹², como mostrado em (i).

(i) resposta \sim frequência + (1 | sujeito)

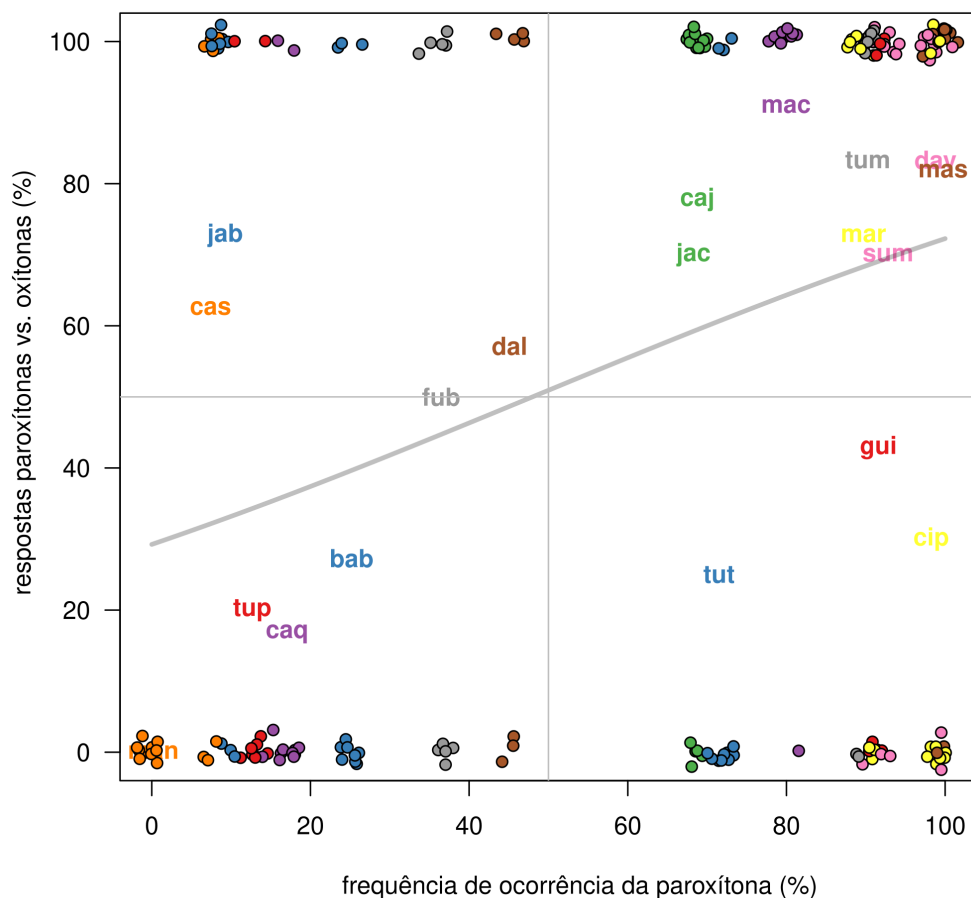
Foi necessário adotar um modelo generalizado linear visto que a variável dependente “resposta da paroxítônica” é percentual, com uma distribuição de erros binomial. A variável independente “frequência relativa da paroxítônica” de cada estímulo foi considerada como um efeito fixo e contínuo no modelo. Os participantes foram considerados como um efeito aleatório no modelo, o que equivale

¹² R: função *lmer*, pacote *lme4*

a assumir que cada participante apresenta uma tendência particular quanto à proporção de escolha da forma paroxítona.

A seguir, a FIG. 31 mostra a distribuição dos dados obtidos e os dados ajustados ao modelo linear.

FIGURA 31 – Percentual de respostas paroxítonas em função da frequência de ocorrência relativa das paroxítonas



man	(manoe-noelma)	mar	(marrom-roma)	cas	(cassis-cisca)
dav	(davi-vida)	mac	(macau-calma)	sum	(sumô-moço)
caj	(caju-juca)	jab	(jabuti-botija)	caq	(caqui-quica)
tum	(tumor-morto)	fub	(fubá-bafo)	mas	(maçom-soma)
dal	(dali-lida)	tut	(tutor-torto)	cja	(cajá-jaca)
tup	(tupi-pito)	cip	(cipó-posse)		
bab	(babá-baba)	gui	(guichê-chegue)		

FONTE: CANTONI, 2017

Na FIG. 31, o eixo horizontal indica a frequência de ocorrência relativa da palavra paroxítona associada aos estímulos. O eixo vertical indica a porcentagem de respostas paroxítonas. As respostas dos participantes para cada estímulo correspondem aos pontos plotados em 0% (resposta = palavra oxítona) ou em 100% (resposta = palavra paroxítona) de resposta à paroxítona. Note-se que as cores dos estímulos se repetem. Os estímulos foram representados em cores diferentes e os pontos foram ligeiramente espaçados, para melhor visibilidade. Os códigos de três letras (as três primeiras letras da palavra oxítona do estímulo) representam a resposta média de cada estímulo. A reta em cinza é a curva de regressão obtida no ajuste ao modelo.

A FIG. 31 ilustra que a curva de regressão apresenta inclinação positiva e é significativamente diferente de zero ($z = 4.26$, $p < 0.0001$). Contudo, ao contrário do que havia sido proposto na hipótese inicial, não pode ser identificada qualquer tendência de preferência por paroxítonas, uma vez que, quando a frequência relativa da paroxítona é igual a 50%, a resposta da paroxítona é 51%, que não é significativamente diferente de 50% ($z = 0.24$, $p > 0.81$). Se houvesse uma tendência a favor das paroxítonas, seria esperada uma resposta à paroxítona significativamente maior que 50% quando a frequência relativa da paroxítona fosse 50% – ou seja, seria esperado que a reta de correlação fosse deslocada do centro para a parte superior do gráfico. Na avaliação do efeito aleatório do modelo¹³, “tendência por informante”, foram obtidos valores nulos, o que indica que tal efeito não é significativo. Pode-se observar uma grande dispersão dos resultados ao redor da curva de regressão, com destaque para os estímulos *jabuti-botija* (jab), *cassis-cisca* (cas), *tutor-torto* (tut), *guichê-chegue* (gui) e *cipó-posse* (cip), que se encontram muito deslocados da curva.

6.3.2.5 Discussão dos resultados

Neste experimento, avaliamos a percepção de sequências sonoras linguísticas ambíguas. Procuramos testar a hipótese de que, nos padrões linguísticos acentuais, a estrutura percebida sofreria influência da frequência de ocorrência e de tipo das palavras testadas. Os resultados obtidos neste experimento confirmam parcialmente tal hipótese: a percepção do estímulo ambíguo como paroxítona ou oxítona depende da frequência relativa de ocorrência das palavras paroxítona e oxítona associadas ao estímulo. Tal relação pode ser interpretada como uma evidência positiva de que o uso linguístico influencia o acesso lexical e de que, portanto, informações sobre a frequência de ocorrência das palavras devem ser armazenadas no léxico, indo ao encontro de diversos estudos já mencionados no capítulo 3. Por outro lado, os resultados não evidenciam nenhuma influência da frequência de tipo – que equivaleria a uma preferência absoluta pela escolha de paroxítonas –, ao contrário do que previa a hipótese inicial. O experimento, por si só, não descarta, contudo, a existência de tal tendência.

¹³ R: função *ranef*, pacote *lme4*

Os resultados apresentados conseguem apreender um efeito estatístico que é global para a população – a correlação entre a identificação da palavra e a frequência de ocorrência. Observa-se, contudo, uma grande dispersão nos resultados, de forma que alguns estímulos se afastam mais que os outros da curva de correlação (cf. FIG. 31). Esse fato pode ser devido a uma variação maior nas frequências próprias de cada informante para tais estímulos, que diferem de algum modo da frequência indicada pelo *corpus* de referência utilizado. O efeito aleatório participante não ter sido significativo indica que o experimento não foi capaz de identificar a existência de tendências pessoais para os participantes. Não descarta, todavia, a existência de tais tendências.

Revedo todo o paradigma experimental, foi identificada uma circunstância que pode ter contribuído para a não significância do efeito da paroxítona. Na manipulação das propriedades acústicas das sílabas para construção dos estímulos, foi utilizada como referência a proporção entre a sílaba forte e a sílaba fraca de uma palavra gravada separadamente. Esse procedimento, entretanto, ainda que tenha promovido o controle experimental, pode ter contribuído para que não fosse possível evidenciar uma preferência à escolha de paroxítonas: não foi considerada a diferença na duração e intensidade entre a sílabas fracas pretônica (da oxítona) e a postônica (da paroxítona). Como mostrado pelos resultados do experimento de produção, as sílabas pretônicas apresentam duração maior que as postônicas. Ou seja, a sílaba átona de uma oxítona é mais longa e mais intensa que a de uma paroxítona. Na manipulação das sílabas, foram usados valores de referência para duração e intensidade que, em média, se aproximam mais aos de uma pretônica. Como resultado, a sílaba fraca não foi suficientemente reduzida em duração e intensidade. Com isso, os estímulos podem ter se tornado tendenciosos à percepção de oxítonas e mascarado uma possível preferência global pela percepção das paroxítonas, que era esperava pela hipótese inicial.

A fim de solucionar o problema da construção dos estímulos e, ao mesmo tempo, possivelmente revelar uma tendência à percepção da paroxítona, seria necessário repetir o experimento, mas com novos estímulos, mais “ecológicos”.

6.3.3 Experimento 3: percepção categórica

Este último experimento de percepção teve como objetivo investigar a categorização de padrões acentuais. Como vimos no capítulo 4, a percepção categórica é um fenômeno presente em diversas esferas da linguagem humana, dos sons aos significados, sempre que se verifica uma mudança abrupta na percepção de padrões contínuos. A percepção categórica está em jogo, por exemplo, na tarefa de reconhecer palavras, uma vez que se trata de buscar e acessar categorias lexicais. Uma das formas de se testar em que nível uma tarefa perceptual é categórica envolve testes de identificação: participantes são submetidos a estímulos isolados e devem classificá-los ou nomeá-los. A partir do ajuste logístico dos dados obtidos no teste, é traçada, para cada participante, uma curva psicométrica, que modela a probabilidade de resposta como função do parâmetro físico de interpolação dos estímulos. As principais medidas realizadas na curva ajustada são

intercepto e inclinação, que refletem, respectivamente, o ponto no contínuo dos estímulos em que ocorre a transição de percepção entre as categorias e quão abrupta é a transição de categorias (cf. capítulo 4 para maiores detalhes sobre essas medidas).

Retomando a hipótese inicialmente levantada neste estudo de que o acento é lexicalmente especificado, decorre que, se ele é parte do conhecimento que o falante dispõe acerca de uma palavra, é, portanto, determinante para seu reconhecimento. A relação de proeminência estabelecida pelas sílabas acentuadas – ao lado de informações segmentais, semânticas, contextuais e mesmo outras, de natureza não linguística – seria um dos fatores que guiam o acesso lexical. A hipótese da especificação no léxico pode ser, assim, expandida, prevendo que, na ausência de informações sobre a estrutura sonora, o reconhecimento das palavras poderia estar sujeito a outros fatores, como o contexto ou a frequência de ocorrência, como defendem os Modelos de Uso. O mesmo é esperado quando as informações sobre o acento (e segmentais) se tornam maximamente ambíguas (e.g. se for gerado artificialmente um estímulo acústico que mescle em igual medida uma palavra oxítona, como *cará*, e uma palavra paroxítona, como *cara*). Essa hipótese subsidiária pode ser avaliada por um teste de identificação, utilizando como estímulos graus diferentes de mesclagem entre duas palavras que diferem apenas pela localização do acento (como *cará* e *cara*). Espera-se que, se as duas palavras mescladas apresentarem valores de frequência de ocorrência semelhantes, a transição entre as categorias lexicais na curva psicométrica tenda a se localizar ao redor do nível mais intermediário dos estímulos, correspondente à forma acusticamente mais ambígua. Ao contrário, quando os valores de ocorrência das palavras mescladas são significativamente distintos, espera-se que a curva psicométrica reflita tal diferença com um deslocamento do intercepto, havendo um maior número de estímulos correspondentes à categoria mais frequente antes da mudança abrupta. Ou seja, é como se a categoria com maior frequência de ocorrência “empurrasse” a curva psicométrica e se estendesse pelos estímulos mais ambíguos, de forma a conquistar maior espaço perceptual, sendo percebida por mais tempo que a categoria menos frequente. A hipótese específica a ser testada neste experimento é a de que, em um conjunto de estímulos progressivamente ambíguos, o espaço perceptual das palavras associadas a tal conjunto pode ser determinado a partir de sua frequência de ocorrência.

6.3.3.1 Participantes

Foram avaliados 22 participantes, estudantes de nível universitário, voluntários, de 20 a 30 anos, igualmente divididos por sexo. Todos os participantes são falantes nativos do português brasileiro e não relataram qualquer história de problemas de fala ou audição.

6.3.3.2 Estímulos

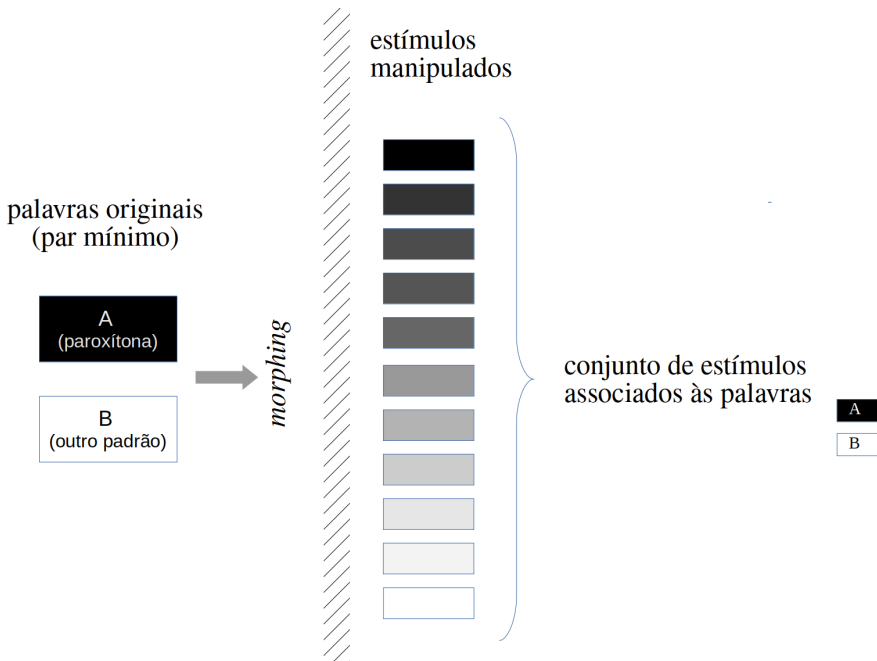
Os estímulos testados consistiram em 10 pares mínimos quanto ao acento lexical (i.e. duas palavras diferentes com sequências sonoras idênticas, exceto pela posição do acento, e.g. “tato”, “tatu”). Os pares mínimos selecionados foram manipulados por

meio do programa de *morphing* acústico Morphala (LABOISSIÈRE et al., 2012), que tem como grande vantagem a naturalidade dos padrões sonoros gerados, ao lado do controle preciso e objetivo das variáveis manipuladas. De forma simplificada, o *morphing* opera mesclando progressivamente, por interpolação, os dois sons originais, gerando “híbridos” distribuídos de forma uniforme entre os estímulos extremos. Também os níveis extremos são manipulados, de forma a homogeneizar os estímulos quanto a possíveis imperfeições decorrentes do processo de filtragem e ressíntese envolvidos na manipulação. Cada “híbrido” H corresponde a um nível da interpolação entre o extremo A e o extremo B , em n níveis – podendo ser descrito como

$$H = \frac{(n - i)A + iB}{n} \quad \text{onde } i = 0, 1, 2, \dots, n$$

O *Morphala* foi utilizado para interpolar as duas palavras de cada par mínimo em onze níveis, resultando em 110 estímulos: 11 níveis x 10 pares de palavras. O diagrama apresentado na FIG. 32 resume a construção e organização dos estímulos experimentais.

FIGURA 32 – Diagrama esquemático do processo de construção dos estímulos do experimento de percepção 3



Como se pode ver na FIG. 32, cada par mínimo selecionado é composto por duas palavras que diferem apenas quanto à posição do acento: palavra A e palavra B. Tais palavras foram submetidas ao *morphing*, de forma a gerar, em níveis progressivos de semelhança, 11 estímulos manipulados, que formam, neste experimento, o conjunto de estímulos associados às palavras A e B.

As FIG. 33a, 33b e 33c, a seguir, mostram exemplos de estímulos gerados pelo *morphing*, sendo os extremos “fábrica” e “fabrica” contrapostos aos níveis dois, quatro, seis e oito de interpolação.

FIGURA 33a – Exemplos de estímulos gerados pelo *morphing* no experimento de percepção 3 conjunto de estímulos *fabrica-fábrica*, níveis 0 (extremo *fábrica*) e 2

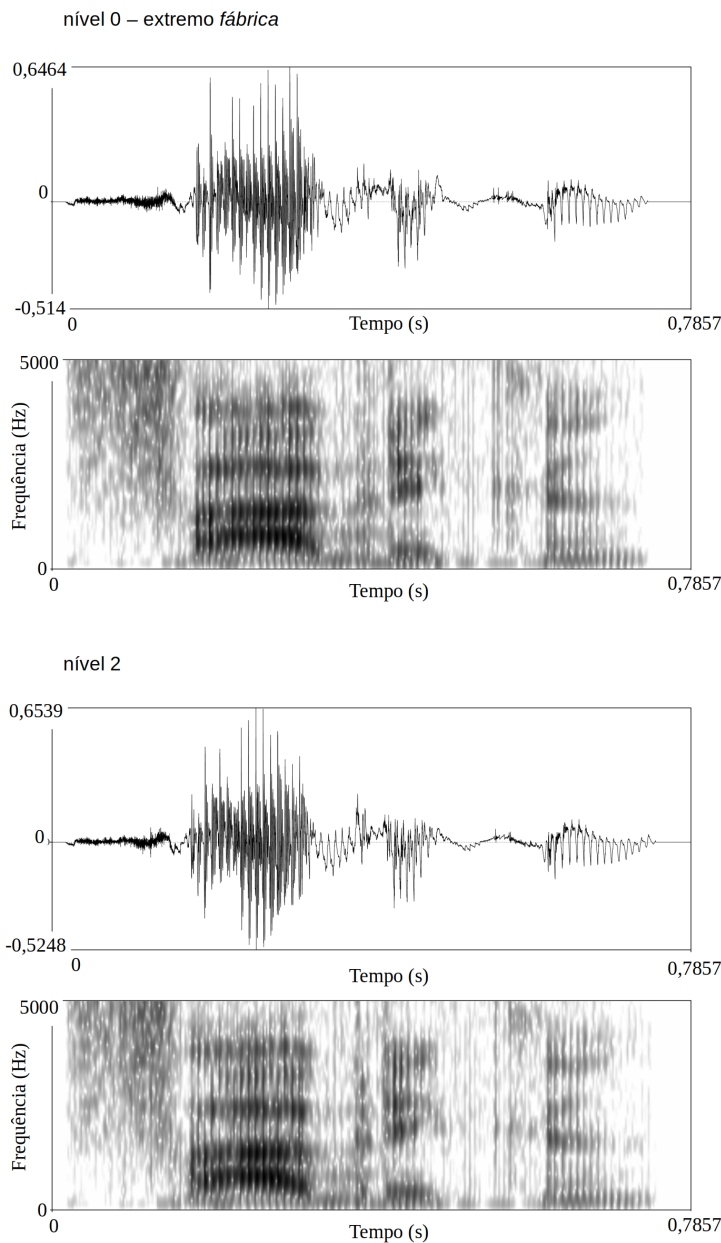


FIGURA 33b – Exemplos de estímulos gerados pelo *morphing* no experimento de percepção 3 conjunto de estímulos *fabrica-fâbrica*, níveis 4 e 6

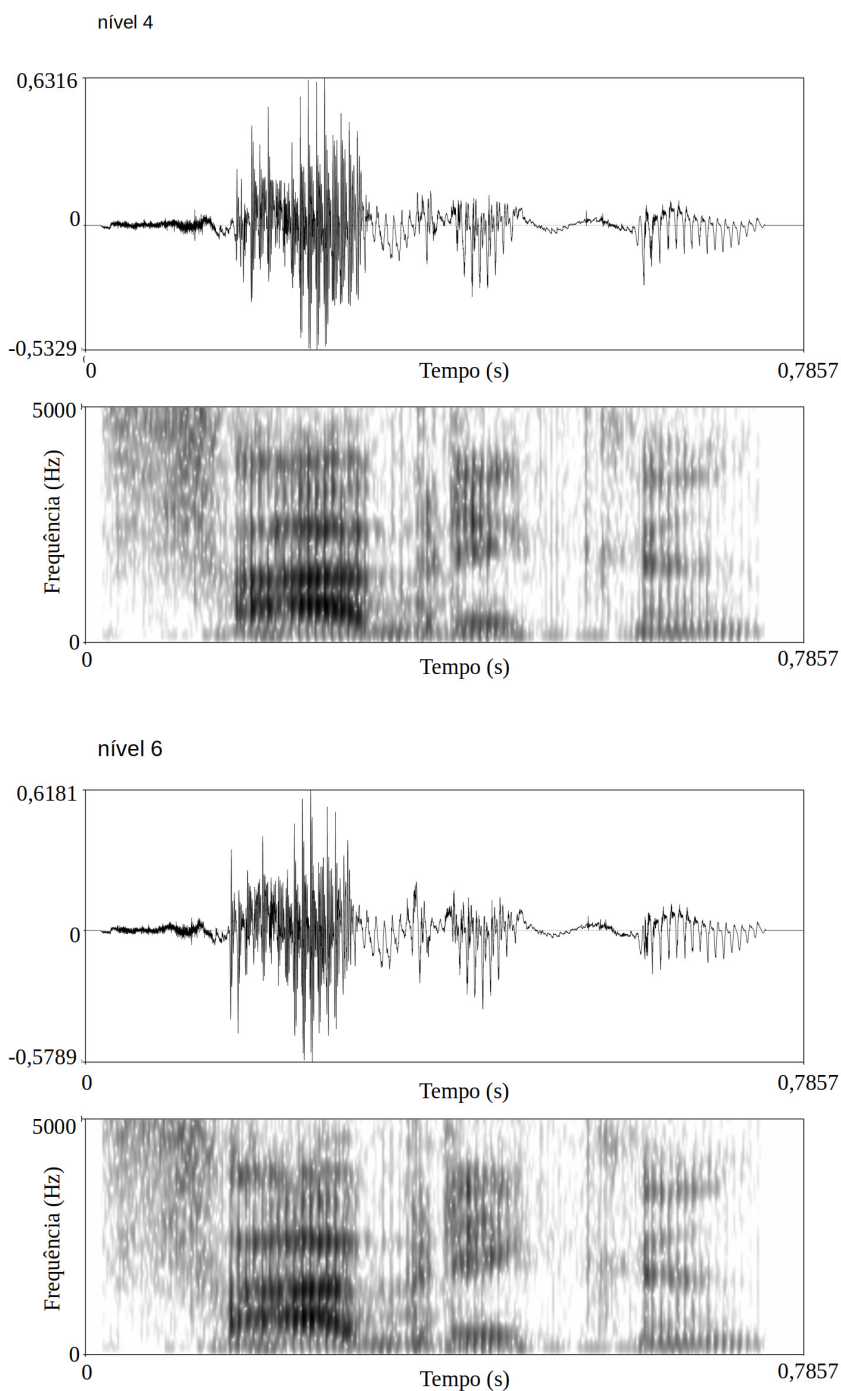
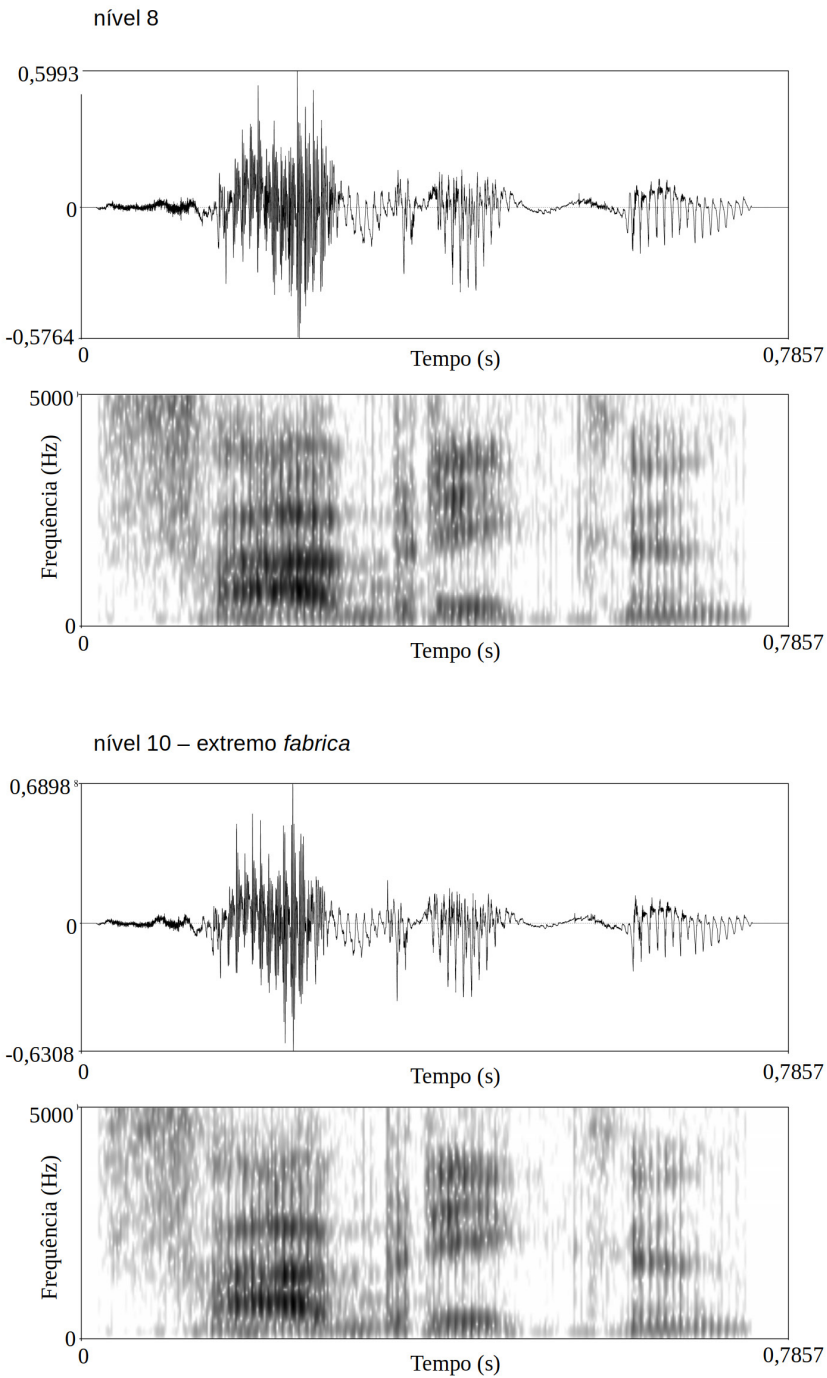


FIGURA 33c – Exemplos de estímulos gerados pelo *morphing* no experimento de percepção 3 conjunto de estímulos *fabrica-fábrica*, níveis 8 e 10 (extremo *fabrica*)



A seguir, o QUADRO 9 apresenta os pares de palavras associadas aos conjuntos de estímulos, assim como seus respectivos valores de frequência de ocorrência, obtidos no Projeto Aspa (cf. cap. 5).

QUADRO 9 – Palavras associadas aos conjuntos de estímulos utilizados no experimento de percepção 3 e seus valores de ocorrência

Palavra 1	Frequência de Ocorrência	Palavra 2	Frequência de Ocorrência
bate	11056	bati	278
pito	206	pitu	29
come	2879	comi	265
tato	441	tatu	388
ficaram	22944	ficarão	3738
secretária	9941	secretaria	43094
publica	3881	pública	52695
hesito	32	êxito	4106
fabrica	1652	fábrica	18455
pratica	1638	prática	27106

Cada conjunto de estímulos é associado a um par de palavras, sendo uma delas sempre paroxítona e a outra podendo ser oxítona ou proparoxítona. Por essa razão, a palavra paroxítona associada ao conjunto foi adotada como referência para o conjunto de estímulos. Nota-se que encontramos dificuldades em determinar qual das palavras do par *secretária-secretaria* seria a paroxítona, uma vez que a sequência de vogais no fim das duas palavras pode ser pronunciada tanto como ditongo quanto como hiato. Optamos provisoriamente por classificar *secretária* como proparoxítona e *secretaria*, como paroxítona (i.e. interpretando a sequência de vogais como um hiato em ambas as palavras).

6.3.3.3 Procedimentos

Os 110 estímulos sonoros originados pelo *Morphala* foram apresentados um a um, de forma aleatória, junto com uma tela contendo as duas palavras associadas ao estímulo, por meio do DmDX¹⁴ (FORSTER; FORSTER, 2003). A cada estímulo ouvido, o participante deveria selecionar uma entre duas palavras apresentadas na tela do computador, em uma tarefa de decisão forçada. A localização de cada

¹⁴ Programa desenvolvido por Jonathan Forster para apresentação de estímulos auditivos e visuais e para medição de tempo de reação em experimentos de psicologia. Disponível em <http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm>.

palavra à esquerda ou à direita da tela também foi determinada aleatoriamente. A seleção das palavras pelos participantes era feita apertando-se a tecla “e” ou a tecla “p” do teclado – sobre as quais foi afixado um quadrado colorido –, sendo “e” correspondente à palavra da esquerda e “p” à palavra da direita. A apresentação dos estímulos foi precedida por um breve treinamento, que consistiu em apresentar a gravação original de palavras de cinco pares mínimos diferentes dos testados, seguindo o mesmo paradigma de apresentação, mas indicando ao participante qual tecla apertar. O treinamento tinha como objetivo habituar o participante à tarefa do experimento: escutar um som, identificar a palavra ouvida entre as duas palavras apresentadas na tela do computador e apertar o botão do teclado correspondente à palavra identificada. Na metade dos estímulos em teste, era realizada uma pausa regulada pelos próprios participantes e que variou de um a dois minutos e meio. Neste experimento, os participantes levaram cerca de 30 minutos para completar a tarefa solicitada.

6.3.3.4 Resultados

A variável dependente no experimento são as respostas como oxítona ou paroxítona. Foi inserida no modelo como a resposta à palavra com maior valor de frequência de ocorrência. Por exemplo, no conjunto de estímulos *pratica-prática*, as respostas foram computadas em relação à palavra *prática*, que é mais frequente que *pratica*. As variáveis independentes são (1) o nível de *morphing* do estímulo, avaliado no modelo como valores de 0 a 1 referentes aos 11 níveis de *morphing* e (2) a frequência de ocorrência das palavras associadas aos estímulos, avaliada no modelo como a frequência relativa da palavra paroxítona (calculada como a frequência absoluta da paroxítona dividida pela frequência absoluta da outra palavra do par, fosse ela oxítona ou proparoxítona).

Os dados obtidos foram ajustados a partir de um modelo linear generalizado com efeitos mistos¹⁵, tendo dois fatores fixos – intercepto e inclinação globais – e dois fatores aleatórios – sujeito e conjunto de estímulos –, como representado em (i).

(i) respostas \sim morphing + (1 + morphing | sujeito) + (1 + morphing | estímulo)

Tal modelo assume que cada participante e cada conjunto de estímulos apresentam intercepto e inclinação próprios. As quatro possibilidades de combinação entre fatores fixos e aleatórios foram verificadas individualmente – comparando-se os modelos reduzidos, contendo as diferentes combinações de fatores, e o

¹⁵ R: função *glmer*, pacote *lme4*

modelo completo, que contém os quatro fatores –, por meio do teste de razão de verossimilhança¹⁶, respeitando a distribuição de qui-quadrado:

(inclinação | sujeito): $\chi^2[2] = 35,4$ $p < 0,0001$

(inclinação | estímulo): $\chi^2[2] = 1,7$ $p = 0,42$

(intercepto | sujeito): $\chi^2[2] = 45,3$ $p < 0,0001$

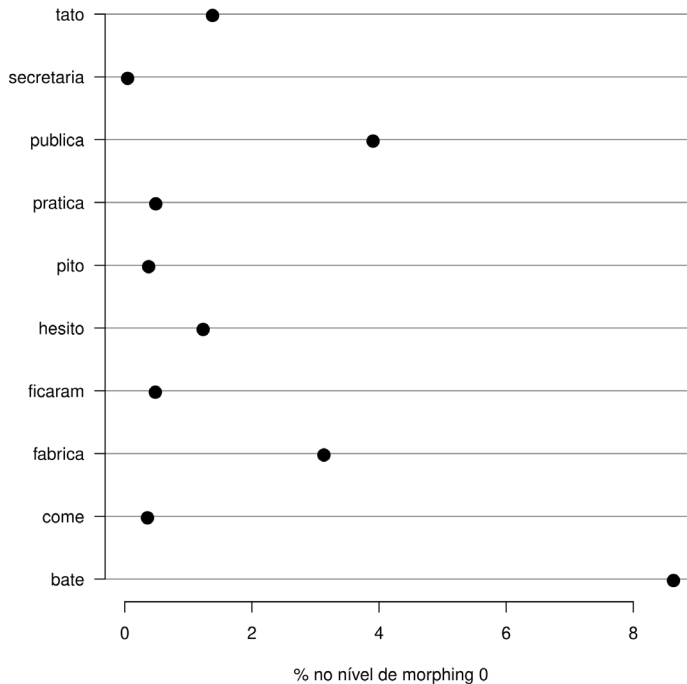
(intercepto | estímulo): $\chi^2[2] = 30,0$ $p < 0,0001$

Como se pode ver, todas as combinações de fatores se mostraram significativas a pelo menos $p < 0,0001$, exceto a que toma a combinação (inclinação | estímulo), que foi retirada do modelo completo, resultando no seguinte modelo final:

(ii) respostas ~ morphing + (1 + morphing | sujeito) + (1 | estímulo)

Ao analisar os dados, verificou-se uma resposta anômala a dois pares de estímulos, bate-bati e secretaria-secretaria, que apresentaram valores muito discrepantes nos extremos, em que se esperam valores próximos de 0% e 100%. As FIG. 34 e 35, na sequência, mostram os valores de respostas de cada par de estímulos nos extremos, respectivamente, 0% e 100%.

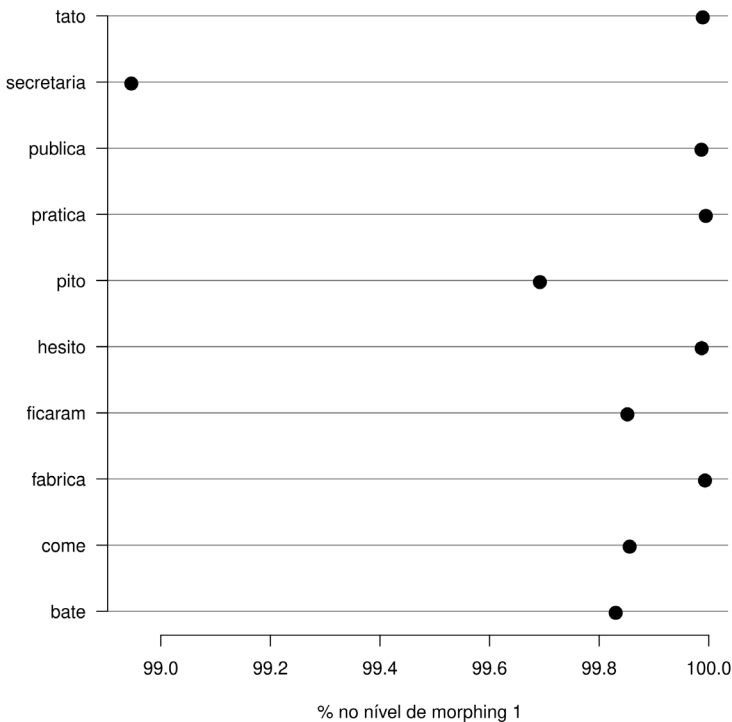
FIGURA 34 – Valores das respostas a cada par de estímulos no extremo 0%



¹⁶ R: função *anova*, pacote *stats*

A FIG. 34 mostra que os pares de palavras testados apresentam valores distintos de resposta no nível extremo de *morphing* a 0%, em que seria esperada uma aglomeração das respostas em torno do 0.

FIGURA 35 – Valores das respostas a cada par de estímulos no extremo 100%



A FIG. 35 mostra que os pares de palavras testados apresentam valores distintos de resposta no nível extremo de *morphing* a 100%, em que seria esperada uma aglomeração das respostas em torno do 100.

Diante da variabilidade nas respostas de diferentes estímulos junto aos níveis extremos de *morphing*, decidiu-se por eliminar quaisquer pares de estímulos que não atingissem 5% ou 95% nos extremos. Apenas o par *bate-bati* não passou no critério e foi eliminado. O modelo foi, então, reajustado sem esse par.

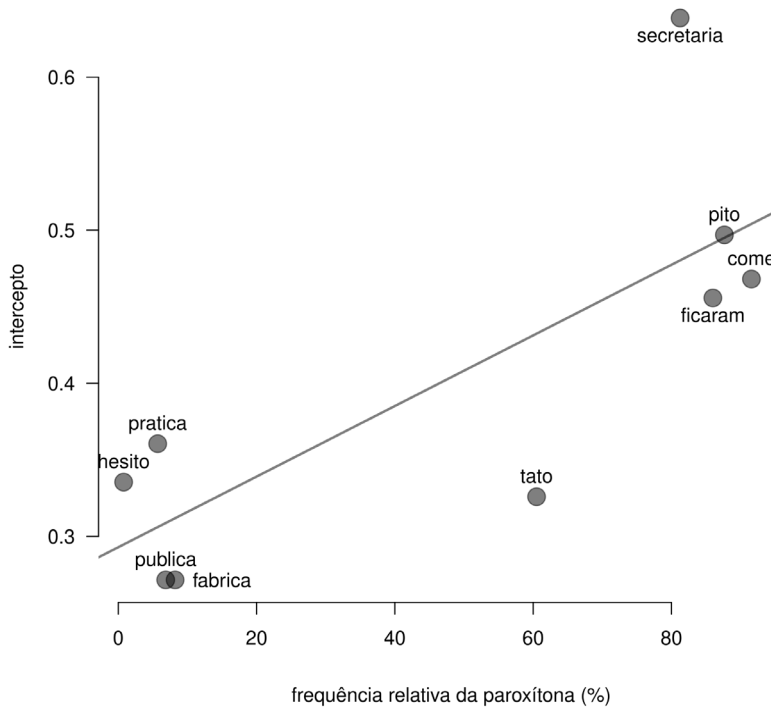
A frequência de ocorrência não foi inserida como fator fixo para evitar a inflação dos graus de liberdade dos resíduos. O tratamento dessa variável independente será realizado na etapa imediatamente seguinte.

Foram então obtidos os coeficientes dos fatores fixos e aleatório¹⁷. A partir do coeficiente do fator aleatório, o intercepto estimado para cada conjunto de

¹⁷ R: funções *fixef* e *ranef*, respectivamente, ambos no pacote *lme4*

estímulos foi confrontado com a frequência de ocorrência relativa da palavra paroxítona associada ao conjunto de estímulos correspondente, por meio de um teste Pearson de correlação¹⁸. Foi obtido um coeficiente de correlação $R = 0,78$, que é significativamente diferente de zero ($t[7] = 3,28$, $p < 0,05$). Pode-se concluir, portanto, que há uma correlação entre as duas variáveis: intercepto específico a cada conjunto de estímulos e frequência de ocorrência relativa da paroxítona associada, como se pode observar na FIG. 36 abaixo.

FIGURA 36 – Correlação entre intercepto dos conjuntos de estímulos e frequência de ocorrência relativa da paroxítona ($R = 0,78$, $t[7] = 3,28$, $p < 0,05$)

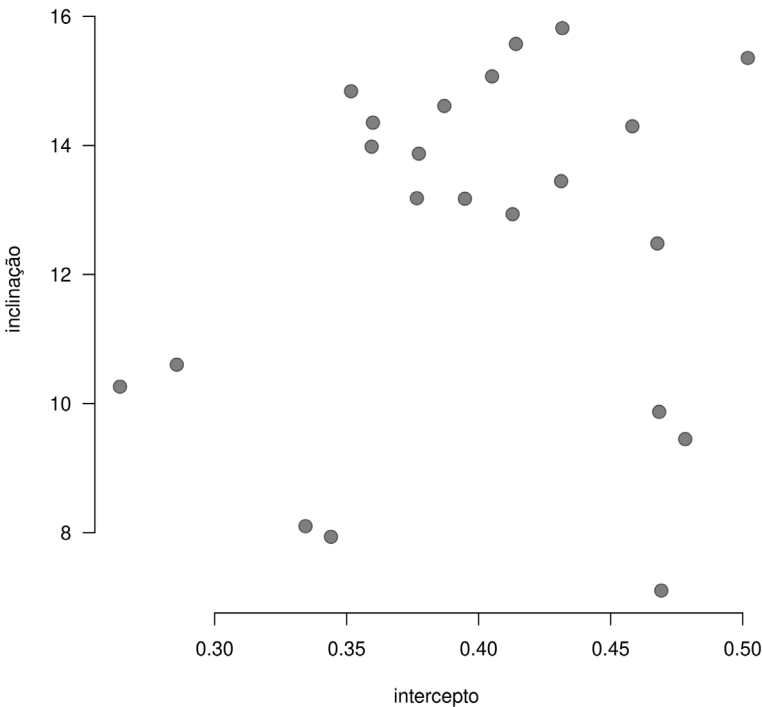


Como ilustrado na FIG. 36, há uma correlação entre a localização do intercepto das respostas para cada par de palavras – i.e., a transição entre categorias – e a frequência de ocorrência relativa dos itens.

No Apêndice C, as FIG. 38 a 46 mostram, para cada conjunto de estímulos, as curvas psicométricas correspondentes às respostas individuais de cada um dos participantes. Por fim, a FIG. 37, a seguir, ilustra o fator aleatório participante.

¹⁸ R: função *cor.test*, pacote *stats*

FIGURA 37 – Inclinação vs. intercepto do fator aleatório participante



A FIG. 37 mostra a distribuição dos participantes considerando a inclinação e o intercepto de suas respostas. Não pode ser observada nenhuma tendência, o que corrobora a adoção dos participantes como efeito aleatório no modelo.

6.3.3.5 Discussão dos resultados

Este experimento avaliou a percepção categórica de padrões acentuais no PB. A hipótese inicial deste terceiro experimento previa uma influência da frequência de ocorrência na identificação de palavras em um conjunto de estímulos progressivamente ambíguos. Os resultados levantados experimentalmente dão suporte a essa hipótese, uma vez que se verificou uma correlação entre o intercepto particular a cada conjunto de estímulos e a frequência de ocorrência das palavras associadas ao conjunto de estímulos. Tal resultado tem duas grandes consequências em termos de implicações teóricas. Em primeiro lugar, pode ser interpretado como evidência empírica da influência do uso linguístico na percepção da fala, como atestado na literatura. Em segundo lugar, serve também como evidência de que o acento deve

ser lexicalmente especificado, uma vez que, na ausência de informações sobre a estrutura sonora, uma informação lexicalmente armazenada – a frequência de uso – foi capaz de influenciar a identificação das palavras.

Um resultado surpreendente foram as tendências de favorecimento de uma das palavras do par de estímulos. A nos basearmos em um possível efeito da frequência de tipo, seria esperado que as palavras pertencentes ao tipo de maior frequência – as paroxítonas – fossem privilegiadas frente à outra palavra do par de estímulos e fossem percebidas antes, contando, assim, com maior espaço perceptual. Entretanto, os resultados indicam que as paroxítonas parecem ser desfavorecidas com relação às oxítonas e proparoxítonas. Retomando a FIG. 31, vemos que os pares de palavras em que a paroxítona é menos frequente apresentam menor intercepto que os pares em que a paroxítona é mais frequente. É claro que, devido ao número reduzido de palavras, não é possível afirmar se tal comportamento é geral. A interpretação de tais resultados não é simples e talvez não seja possível explicar exatamente os mecanismos por trás do desfavorecimento das paroxítonas.

A tentativa de explicação proposta recorre a características fonéticas das sílabas acentuadas e átonas, bem como a processos fonológicos observados na variedade de Minas Gerais, de onde são os participantes do experimento e o locutor dos estímulos.

Os pares de palavra em que se observou um desfavorecimento das paroxítonas são formados por paroxítona-proparoxítona (e.g. *pratica-prática*). Como sabemos do experimento de produção, as sílabas pretônicas são acusticamente mais semelhantes às acentuadas (em termos de valores de duração, intensidade e F0) do que as postônicas. Em uma palavra proparoxítona, temos duas sílabas postônicas, o que, por extensão, tornaria paroxítonas acusticamente mais semelhantes às oxítonas que às proparoxítonas. Nas proparoxítonas, adicionalmente, há dois *loci* de contraste: além do contraste da acentuada com a postônica, as duas postônicas contrastam entre si. Tais características acústicas fariam a proparoxítona mais saliente em comparação com a paroxítona associada ao mesmo par, e poderiam estar por trás do maior espaço perceptual conferido à proparoxítona, o que explicaria parcialmente nossos resultados. A outra parte da explicação servirá para justificar a preferência pelas oxítonas, em decorrência de fenômenos sonoros de mudança e variação.

Como discutimos no capítulo 4, a fonologia do PB se estrutura ao redor das sílabas acentuadas. Na fala corrente, muitas vezes, devido aos inúmeros cancelamentos de vogais átonas e emergência de encontros consonantais, temos um grande índice de sílabas acentuadas na produção da fala (cf. os estudos de Souza

(2012) e Leite (2006) sobre o cancelamento de vogais pretônicas e postônicas na variedade de Belo Horizonte). Em decorrência do fenômeno de cancelamento de vogais átonas finais, que tem ampla ocorrência na variedade mineira, ocorre uma reestruturação dos padrões silábicos: paroxítonas se tornam oxítonas pelo cancelamento de vogais átonas (e.g. *passee* ['pas]); proparoxítonas se tornam paroxítonas ou mesmo oxítonas, seja pelo mesmo fenômeno, seja pela formação de encontros consonantais (e.g. *príncipe* ['prĩsipi] ~ ['prĩspi]); novas proparoxítonas são formadas por epêntese (e.g. *ritmo* ['hitʃimo]).

Esse percurso de reestruturação pode ser interpretado, em uma perspectiva dinâmica, como decorrente da auto-organização do sistema sonoro, em que se mantêm os padrões acentuais, mas mudam as palavras associadas a cada padrão, como em uma “dança das cadeiras”. Mudanças em cadeia são recorrentes nas línguas do mundo e podem ser interpretadas como resultado da coordenação de padrões gramaticais. Uma das consequências dessa cadeia de eventos para a estrutura sonora do PB é uma super-inflação do novo padrão oxítono, que passa a abarcar também as antigas paroxítonas, mais numerosas em tipo.

A modelagem multirrepresentacional permite explicar como, diante de fenômenos de mudança e variação linguística, somos capazes de lidar simultaneamente com formas concorrentes de uma mesma palavra. Se no contínuo da fala os contrastes entre as três classes acentuais se reduzem drasticamente, contudo, as formas isoladas, associadas a uma das três classes, não se perdem, pois continuam lexicalmente armazenadas nas representações dos exemplares. A redundância das informações nos permite acessar, dessa forma, tanto a forma reduzida da palavra na fala contínua quanto a forma completa, em isolamento.