



Distinção de ponto de articulação no Português de Belo Horizonte: exemplos em plosivas e fricativas

Distinction of Place of Articulation in Brazilian Portuguese: Examples in Plosives and Fricatives

Rui Rothe-Neves

Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais / Brasil
rothe-neves@ufmg.br

Fabiana Andrade Penido

Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais / Brasil
fabianapenido@gmail.com

Resumo: Este artigo trata das pistas auditivas utilizadas para a distinção de ponto de articulação no português brasileiro (PB). Na primeira parte, descrevemos as propriedades acústicas das fricativas [ʃ] e [s] e das plosivas [b] e [d] seguidas da vogal [a]. Foram analisadas amostras de fala de sete falantes nativos de PB registrados em Belo Horizonte. Foram obtidos valores de duração semelhantes para [ʃ] e [s], bem como para a fase de soltura da oclusão oral nas plosivas. A fase de pré-sonorização foi maior para [b] do que para [d]. Picos de amplitude mais proeminentes e concentração de energia em regiões de maior frequência foram obtidos em [s] quando comparado a [ʃ]. Registramos picos de energia em regiões de maior frequência para [d] em comparação com [b]. Finalmente, F2 foi maior no início da transição de formantes após [ʃ] e [d] quando comparado a [s] e [b]. Na segunda parte, oito falantes nativos de Belo Horizonte realizaram uma tarefa de classificação. Tanto a transição de F2 e F3 quanto o centro de gravidade foram usados para distinção de /japa/-/sapa/. O uso do ruído fricativo foi diferente em razão da transição de formantes, o que é compatível com a descrição acústica em que a

transição foi praticamente nula para [s]. Os participantes não utilizaram a pista de *burst* para a classificação de /bata/-/data/, apenas a transição de formantes. Os resultados confirmam que diferenças linguísticas afetam a percepção da distinção de ponto de articulação e mostram quais características da fala são utilizadas por falantes do PB.

Palavras-chave: percepção da fala; análise da fala; características acústicas; segmentos fonéticos; fonética; linguística.

Abstract: This article deals with the auditory cues used for the distinction of the place of articulation in Brazilian Portuguese (PB). In the first part, we describe the acoustic properties of the fricatives [ʃ] and [s] and the stops [b] and [d] followed by the vowel [a] of Brazilian Portuguese (BP). We analyzed speech samples of seven native BP speakers recorded in Belo Horizonte. Similar duration values were obtained for [ʃ] and [s], as well as for the release phase of oral occlusion in the stops. The prevoicing phase was longer for [b] than for [d]. More prominent amplitude peaks and energy concentration in higher frequency regions were obtained in [s] when compared to [ʃ]. We registered energy peaks in higher frequency regions for [d] as compared to [b]. Finally, F2 was higher at the beginning of formant transition for [a] following [ʃ] and [d] when compared to [s] and [b]. In the second part, eight native speakers of Belo Horizonte performed a classification task. Both the F2 and F3 transition and the center of gravity aided in the distinction of /ʃapa/-/sapa/. The use of fricative noise was different as a function of formant transition, what is compatible with the acoustic description in which the transition was almost null for [s]. For /bata/-/data/, the participants did not use the burst cue for the classification of [b] and [d], only the formant transition. The results confirm that linguistic differences affect the perception of the place of articulation and show which characteristics are used by BP speakers.

Keywords: speech perception; speech analysis; acoustical properties; phonetic segments; phonetics; linguistics.

Recebido em 4 de maio de 2017
Aceito em 31 de outubro de 2017

1 Introdução

Este estudo investiga, por meio de análise acústica e de uma tarefa de classificação, as pistas auditivas para diferenciar as palavras /*japa/-/sapa/* e /*bata/-/data/*. O termo “pistas auditivas” refere-se à informação do sinal acústico que permite ao ouvinte perceber a existência de um contraste fonológico (WRIGHT, 2004). Por exemplo, como a principal propriedade acústica que diferencia as fricativas [ʃ] e [s] é a altura da frequência do ruído fricativo, essa pista pode ser utilizada pelos ouvintes para distinguir esses sons (HARRIS, 1958). No caso das plosivas [b] e [d], a pista que melhor auxilia na distinção desses sons é a diferença entre as amplitudes do ruído de explosão (do inglês, *burst*) que se registra na fase de soltura da oclusão. A transição dos formantes vocálicos também auxilia os ouvintes na distinção das fricativas e das plosivas. No português brasileiro (PB), há alguns estudos acústicos sobre fricativas e plosivas com grupos de falantes, porém poucos tratando especificamente das propriedades acústicas relevantes para investigar a percepção desses sons (SANTOS, 1987; RUSSO; BEHLAU, 1993; BARBOSA, 1999; HAUPT, 2007), tais como os valores de amplitudes do ruído fricativo e os valores das amplitudes do *burst* no caso das plosivas. Além disso, não nos foi possível encontrar estudos do PB que tenham realizado uma investigação da pista de transição dos formantes vocálicos seguintes às fricativas e às plosivas, na região de transição e na região de estabilização dos formantes. Finalmente, tanto quanto nos seja dado conhecer, nenhum estudo procedeu à investigação da percepção da fala manipulando essas características acústicas. Ao possibilitar a comparação de segmentos produzidos em diferentes pontos de articulação, por exemplo, os estudos sobre a acústica desses sons podem revelar características da produção do som. Entretanto, não podem responder à pergunta: em que ponto os falantes da língua deixam de perceber um som e passam a perceber o outro?

Com este artigo, espera-se poder fazer afirmações sobre as propriedades acústicas que os falantes usam para diferenciar as fricativas [ʃ] e [s] e as plosivas [b] e [d] associadas a vogal [a] do PB, em posição inicial de palavra, de modo a enriquecer a caracterização desses segmentos. Para isso, na primeira parte deste texto, analisamos e descrevemos propriedades acústicas dessas consoantes seguidas da vogal [a] em posição inicial de palavra (estudo 1). Na segunda parte, realizamos um experimento de classificação com os estímulos sintetizados, em que a

manipulação das propriedades acústicas descritas no estudo 1 possibilitou avaliar a distinção entre as fricativas [ʃ] e [s] e as plosivas [b] e [d] por falantes da língua (estudo 2). Apresentam-se a seguir as propriedades acústicas sabidamente utilizadas para a distinção de ponto de articulação, com base em estudos sobre o PB e outras línguas. Em seguida, relatam-se o estudo de fonética acústica e, então, o experimento de classificação.

2 Estudos precedentes

Diversos autores contribuíram para a compreensão das propriedades acústicas dos sons fricativos (FANT, 1960; HEINZ; STEVENS, 1961; HIXON, 1966; LADEFOGED; MADDIESON, 1996), cujas características espectrais dependem sobretudo do lugar da constrição no trato vocal, do formato do orifício na constrição e da queda de pressão nessa região. Durante a produção do [s], temos a formação de um canal mais estreito. O canal mais largo para o [ʃ] faz o ar ter menos velocidade. Devido à diminuição da velocidade da corrente de ar, ao arredondamento dos lábios (característica coarticulatória desse som em diversas línguas) e à extensão maior do trato vocal depois da constrição, [ʃ] tem maior concentração de energia em regiões mais baixas do espectro quando comparado a [s].¹

Embora as características articulatórias sejam semelhantes, há diferenças acústicas entre as línguas. No inglês, o espectro de [ʃ] aparece com a principal concentração de energia na faixa de frequência de 2-3 kHz. Para [s], o espectro mostra um aumento de energia que começa em torno de 3,56-4,4 kHz (STEVENS; KLATT, 1968). No italiano, observou-se a frequência central de [s] em torno dos 4 kHz e de [ʃ], em torno dos 2 kHz (SHINDLER, 1974). No espanhol argentino, [s] apresenta picos espectrais ao redor de 5-8 kHz, e [ʃ], por volta de 2,5-5 kHz (BORZONE DE MANRIQUE; MASSONE, 1981). No português europeu, Lacerda (1982) concluiu, por meio de testes perceptivos, que [s] é mais bem percebido quando o estímulo tem altos níveis de intensidade e picos espectrais na região de 5 kHz, e [ʃ] é normalmente associado a altos níveis de intensidade juntamente com

¹ Esta apresentação é restrita por razões de espaço. Para detalhes sobre a produção de sons da fala em geral, ver Marchal e Reis (2012) e, para uma introdução à análise acústica dos sons da fala, Barbosa e Madureira (2015).

picos espectrais na região dos 3 kHz. No português brasileiro, [s] aparece como um som mais agudo, com faixa de frequência entre 4,5-8 kHz, enquanto a alveolopalatal [ʃ] também apresenta uma faixa de frequência ampla, semelhante a [s], porém mais grave, entre 2,5-6 kHz (RUSSO; BEHLAU, 1993; SANTOS, 1987; HAUPT, 2007). Gordon, Barthmaier e Sands (2002) compararam fricativas não-vozeadas em sete línguas da América do Norte, Escócia e Ásia. Um dos parâmetros analisados foi o centro de gravidade (do inglês, *center of gravity*, CG), que pode ser definido como a frequência abaixo (ou acima) da qual estão concentradas 50% da energia de determinado som. Indica, portanto, a faixa de frequências mais intensas no ruído fricativo. Os autores concluíram que a fricativa [s] também apresentou CG maior do que a fricativa [ʃ] em seis línguas. O valor do CG da fricativa [ʃ] foi maior somente em tuda, uma língua dravidiana falada na Índia.

Outro parâmetro que parece importante para distinção nos sons fricativos é a duração. Diversos estudos evidenciaram que os sons fricativos alveolares [s, z] são mais longos que os alveolopalatais [ʃ, ʒ], e os sons não-vozeados são, por sua vez, mais longos do que os vozeados (STEVENS; KLATT, 1968; BORZONE DE MANRIQUE; MASSONE, 1981; SANTOS, 1987; HAUPT, 2007).

Quanto às consoantes plosivas, são segmentos produzidos com o bloqueio total da corrente de ar em algum ponto do trato vocal, com posterior soltura desse bloqueio. A plosiva bilabial vozeada [b] é produzida por meio da constrição dos lábios com vibração concomitante das pregas vocais, enquanto a plosiva [d] é produzida com a constrição da ponta da língua contra a face interna dos dentes incisivos superiores ou dos alvéolos, também com vibração das pregas vocais (LADEFOGED; MADDIESON, 1996).

Do ponto de vista da dinâmica articulatória, a produção das plosivas é identificada por duas fases distintas: a fase de oclusão e a fase de soltura (ISTRE, 1983). A fase de oclusão corresponde ao intervalo em que os articuladores interrompem completamente a passagem de ar. Essa fase é caracterizada somente pelo período de silêncio, no caso das plosivas não-vozeadas, ou pelo período de silêncio associado à sonorização, no caso das plosivas vozeadas. A fase de soltura, que corresponde à liberação da corrente de ar previamente bloqueada em algum ponto do trato vocal, é caracterizada pela espícula de plosão ou *burst*, seguida da região de transição dos formantes para a vogal seguinte. Em geral, observa-se uma

espícula mais intensa para as plosivas não vozeadas [p, t, k] do que para as plosivas vozeadas [b, d, g]. O *burst* se mostrou uma pista importante para a distinção de consoantes plosivas (STEVENS; BLUMSTEIN, 1978) – o *burst* das plosivas bilabiais tem energia concentrada em regiões de frequência mais baixas quando comparado ao das plosivas alveolares. Em português brasileiro, os segmentos plosivos [p, b] são os mais graves e têm energia em torno de 0,5-1,5 kHz. Os segmentos [t, d] são considerados os mais agudos, com energia concentrada por volta de 4 kHz, com um pico secundário fraco, em torno de 0,5 Hz (RUSSO; BEHLAU, 1993).

Por fim, também a transição de formantes ao início ou ao fim da consoante são pistas auditivas importantes para o ponto de articulação. “As cavidades dentro do trato vocal agem como um filtro multirressoante sobre o ar transmitido e nele imprimem uma estrutura correspondente de formantes superposta à fina estrutura harmônica” (FANT, 1973, p. 5). Os três primeiros formantes (F1, F2 e F3) são os mais importantes e considerados suficientes para diferenciar as vogais. A transição, isto é, a mudança nos valores dos formantes nessa região é considerada uma pista auditiva importante para distinções linguisticamente relevantes. Por exemplo, espera-se observar diferenças nos valores do segundo formante vocálico (F2) da vogal [a] que segue as fricativas [s] e [ʃ], com uma frequência mais alta de F2 na vogal que segue a fricativa [ʃ], quando comparada àquela que segue [s].² Assim como nas fricativas, a diferença entre [ba] e [da] se concentra principalmente na região de transição de F2. Dessa forma, espera-se verificar valores do F2 para [da] maiores quando comparado a [ba], na região de transição dos formantes.

3 Análise acústica

Nesta parte do artigo, analisamos e descrevemos as propriedades acústicas das fricativas [ʃ] e [s] e das plosivas [b] e [d] associadas à vogal [a], em posição inicial de palavra. A análise visa a uma descrição acústica desses sons considerando as características apresentadas na literatura, incluindo aquelas ainda não descritas para o PB.

² A explicação físico-acústica dos movimentos dos formantes em função das consoantes vizinhas foi oferecida pela Teoria da Perturbação; ver apresentação resumida em Barbosa e Madureira (2015, p. 106 *et seq.*)

3.1 Métodos

Os participantes deste estudo foram sete adultos do sexo masculino, falantes nativos do português brasileiro de Belo Horizonte (MG), com faixa etária entre 20 e 45 anos de idade. Os participantes assinaram o Termo de Consentimento Livre e Esclarecido, após serem informados sobre os aspectos gerais da pesquisa. O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais (CAAE – 62617616.0.0000.5149).

As fricativas [ʃ] e [s] e as plosivas [b] e [d] foram registradas em início de palavra, seguidas da vogal [a], nas palavras “chapa”, “sapa”, “bata” e “data”. A gravação foi realizada em uma sala acusticamente tratada, com a utilização de um *notebook* (*Macbook Air*), de um microfone de cabeça (*Philips*) e do *software* Praat (versão 5.2.35) (BOERSMA; WEENINK, 2011), com taxa de amostragem de 22,05 kHz. Cada palavra foi produzida três vezes por cada um dos sete participantes. As palavras foram inseridas na seguinte frase veículo: “Eu digo _____ para ela”. Após a gravação, os dados de fala foram etiquetados e analisados.

A análise acústica dos dados de fala iniciou-se pela inspeção visual do sinal de fala e do espectrograma. Em seguida, todas as produções dos informantes foram segmentadas e etiquetadas, ou seja, cada som foi identificado em sua porção inicial e final. Após a realização da etiquetagem, os sons foram analisados separadamente. A análise acústica dos segmentos fricativos, plosivos e da vogal [a] compreendeu os seguintes parâmetros:

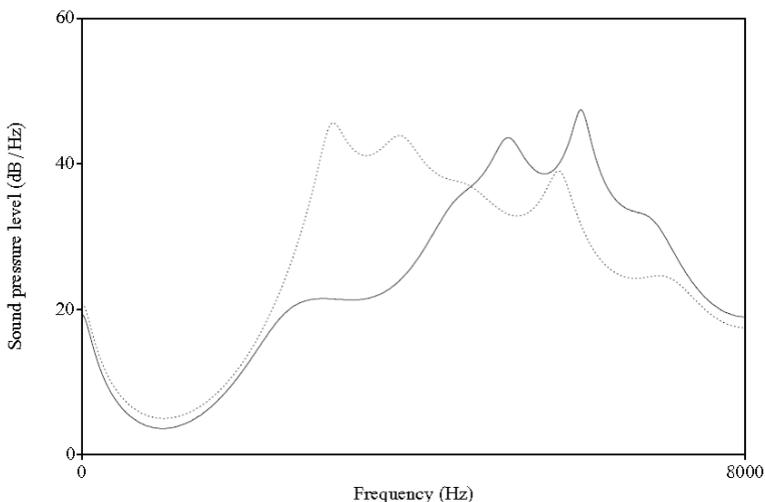
- (a) Duração total dos segmentos fricativos, plosivos e vocálicos;
- (b) Centro de gravidade (CG) do ruído fricativo;
- (c) Picos de amplitude e formantes do ruído fricativo;
- (d) Duração da pré-sonorização do segmento plosivo;
- (e) Duração da soltura da oclusão (do inglês, *burst*), representada pela barra de plosão, nos segmentos plosivos;
- (f) Picos de amplitude e formantes da soltura da oclusão;
- (g) Valores dos três primeiros formantes vocálicos (F1, F2 e F3) nas porções inicial e estável da vogal.

Descrevemos em maior detalhe a maneira como se obtiveram essas medidas para cada classe de segmento analisada.

3.1.1 Propriedades acústicas dos segmentos fricativos

O ruído fricativo foi selecionado excluindo-se a região de transição de formantes da vogal [a] seguinte às fricativas. Selecionado o ruído, os valores dos formantes foram obtidos por LPC com base nos seguintes parâmetros de análise: quantidade máxima de formantes = 6; formante máximo = 8,5 kHz; janela = 0,025. Foram obtidos os valores de frequência dos formantes de toda a porção selecionada do ruído fricativo e calcularam-se os valores médios das seis frequências proeminentes do ruído fricativo (F1 a F6). A obtenção dos valores das amplitudes dos formantes do ruído fricativo foi realizada com base nos valores dos formantes. A Figura 1 a seguir mostra a configuração espectral dos ruídos fricativos de [s] e [ʃ] de um dos informantes do estudo.

FIGURA 1 – Configuração espectral dos ruídos fricativos em [ʃ] (linha pontilhada) e em [s] (linha contínua) obtidos de um dos informantes do estudo



Fonte: Elaborada pelos autores.

O valor de CG foi obtido somente com a seleção de todo o ruído fricativo excluindo a região de transição dos formantes da vogal [a] seguinte às fricativas. Esse parâmetro não foi investigado separadamente, ou seja, porção inicial, medial e final do ruído, devido à constatação de que há uma variação pequena nesses valores. Por fim, a duração total do

ruído fricativo foi obtida selecionando-se o ruído fricativo, sem a porção de transição dos formantes entre o ruído e a vogal subsequente.

3.1.2 Propriedades acústicas dos segmentos plosivos

O parâmetro acústico mais relevante nos segmentos plosivos para a distinção entre [b] e [d] são as amplitudes do *burst*. Nos segmentos [b] e [d], as amplitudes foram obtidas selecionando a soltura da fase de oclusão. Para a seleção do *burst* utilizou-se como referência as informações disponibilizadas pelo oscilograma e pela barra de explosão, quando presente (FIG. 2). Na ausência da barra de explosão utilizaram-se como referência as informações disponibilizadas pelo oscilograma. Dessa forma, selecionou-se o ponto onde se observa a modificação do traçado no oscilograma e o início da produção vocálica com a identificação de F2.

Após a identificação e a seleção do *burst*, foram obtidos os valores das amplitudes. Os valores dos formantes do *burst* foram obtidos da mesma forma como se deu para as fricativas. Outro parâmetro analisado foi a duração da fase de pré-sonorização e do *burst* dos segmentos plosivos [b] e [d]. Esses parâmetros são mostrados na Figura 3, onde é possível visualizar a barra de vozeamento, que corresponde à vibração das pregas vocais durante a articulação desses sons, a região de soltura da oclusão (*burst*) e a fase de pré-sonorização representada pela fase de oclusão da plosiva.

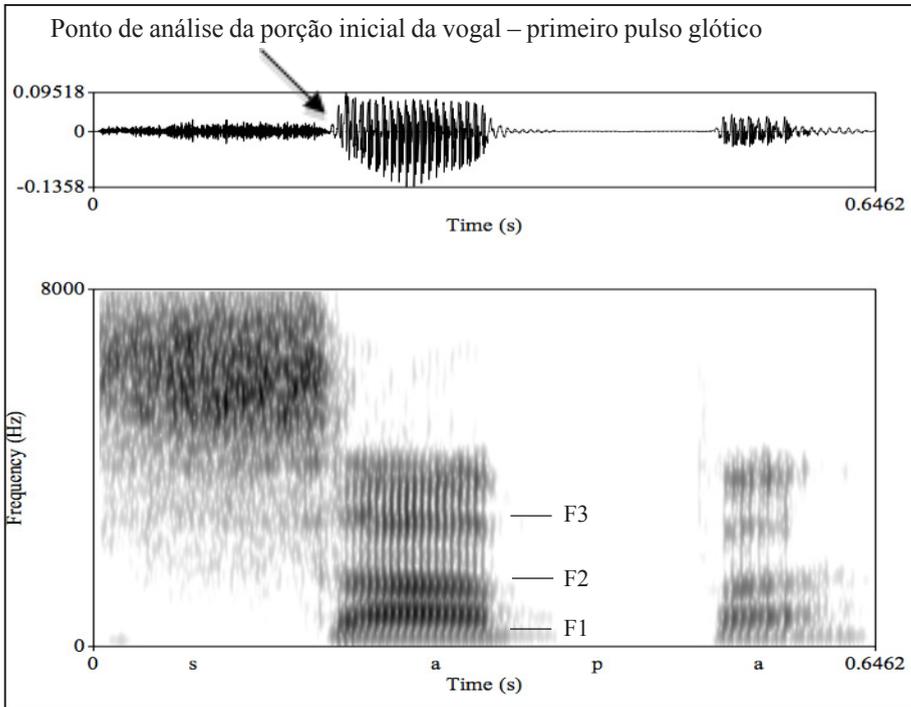
3.1.3 Propriedades acústicas da vogal [a] seguinte às plosivas e às fricativas

As medidas temporais e espectrais dos três primeiros formantes (F1, F2 e F3) foram obtidas manualmente por meio da análise do espectrograma e do oscilograma. Na análise espectral dos formantes vocálicos, os valores dos formantes foram obtidos por LPC com base nos seguintes parâmetros de análise: amplitude (Hz) = 0 a 8 kHz; janela(s) = 0,005; e amplitude dinâmica (dB) = 70 dB.

Os valores dos formantes vocálicos foram extraídos automaticamente pelo Praat, pelo método LPC, em dois principais momentos: 1) na porção inicial da vogal (*onset* vocálico; porção de transição dos formantes entre a consoante e a vogal) e na porção estável da vogal, ou seja, no momento em que se observou pequena variação entre os valores dos formantes.

O ponto inicial da vogal, ou seja, de transição dos formantes, foi selecionado com base nas informações oferecidas pelo oscilograma e espectrograma. No oscilograma, é possível visualizar uma mudança abrupta na amplitude entre a consoante e a vogal (FIG. 3). Após a identificação desse ponto no oscilograma, marcou-se o ponto no qual se observou o início do primeiro pulso glótico e o início do segundo formante vocálico visualizado no espectrograma. Nesse ponto, foram obtidos os valores das frequências dos formantes (F1, F2 e F3), que representaram a porção inicial da vogal.

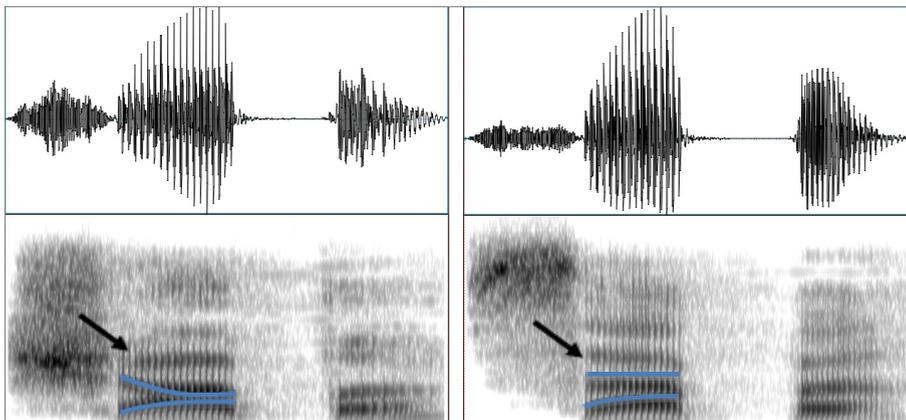
FIGURA 2 – Sinal de fala e espectrograma da palavra /sapa/. Ponto de análise da porção inicial da vogal [a] e seus formantes vocálicos (F1, F2 e F3)



Fonte: Elaborada pelos autores.

As figuras a seguir mostram o movimento do F1 e F2 na região de transição dos formantes entre as fricativas [ʃ] e [s] e a vogal [a], nas palavras “chapa” e “sapa”. Na figura da esquerda, observa-se um decréscimo do valor do segundo formante (F2) entre a fricativa [ʃ] e a vogal [a], conforme esperado.

FIGURA 3 – F1 e F2 da vogal [a] seguinte à fricativa [ʃ], na palavra “chapa” (à esquerda) e F1 e F2 da vogal [a] seguinte à fricativa [s], na palavra “sapa” (à direita)



Fonte: Elaborada pelos autores.

No caso da vogal precedida pelas fricativas, selecionou-se uma porção de 0,03 s de ruído e 70 ms de vogal, com o intuito de auxiliar no estabelecimento do ponto inicial da porção vocálica para extração dos valores dos formantes. Em seguida, solicitou-se o “*Formant listing*” disponibilizado pelo Praat, nas configurações em “*Formant*”.

Os valores dos formantes da porção estável da vogal foram definidos pela visualização do movimento dos formantes no espectrograma (pontos vermelhos) e pelos pulsos glóticos regulares no oscilograma. Além disso, com o auxílio do “*Formant listing*”, selecionou-se 20 ms da parte estável da vogal para obtenção do F1, F2 e F3.

Diante da vogal [a] precedida pelas plosivas, os valores dos formantes foram obtidos da mesma forma como se deu nas fricativas. O ponto inicial da vogal foi estabelecido a partir do primeiro pulso glótico vocálico, logo após a barra de soltura da oclusão da consoante plosiva, e com base na identificação do segundo formante (F2). Já os valores dos formantes da porção estável da vogal foram definidos conforme a descrição anterior.

As medidas de duração da vogal [a] foram obtidas selecionando a vogal a partir do primeiro pulso glótico regular até o último pulso regular dessa mesma vogal.

3.2 Resultados e considerações do estudo 1

Os resultados desse estudo são apresentados a seguir. Optou-se por descrever os valores de cada informante, para possibilitar posterior reanálise ou outra utilização desses dados. Apresentam-se também as médias do grupo para cada parâmetro acústico, bem como os desvios-padrão (dp). A normalidade dos dados foi estimada pelo teste de Shapiro-Wilks, e todos os resultados relatados aqui não violam a suposição de normalidade, exceto quanto dito em contrário. A homogeneidade de variância foi estimada pelo teste de Levene, e todos os resultados aqui relatados também não violam essa suposição. As diferenças entre médias foram avaliadas por meio de testes t de Student para amostras pareadas (graus de liberdade = 6) ou pelo teste dos postos sinalizados de Wilcoxon, nos poucos casos em que os dados violaram a normalidade.³

3.2.1 Propriedades acústicas das fricativas [s] e [ʃ]

3.2.1.1 Duração das fricativas

A duração total dos segmentos fricativos pode ser observada na tabela abaixo.

TABELA 1 – Duração das fricativas [s] e [ʃ] (ms)

Informante	[s]	[ʃ]
1	192	194
2	160	145
3	176	189
4	183	177
5	158	171
6	141	142
7	178	197
Média	169,7	173,6
dp	17,5	22,5

Fonte: Elaborada pelos autores.

Legenda: dp = desvio padrão

³ Todas as análises estatísticas foram realizadas no software R (R CORE TEAM, 2016) utilizando os seguintes pacotes: *stats*, *car* (FOX; WEISBERG, 2011) e *pastecs* (GROSJEAN; IBANEZ, 2014).

A Tabela 1 mostra que os valores médios de duração das fricativas [s] e [ʃ] foram 169 ms e 173 ms; respectivamente. Esses valores são próximos daqueles verificados por Santos (1987): em posição inicial de palavra seguido da vogal [a], a autora obteve valores médios de 165 ms para [s] e 180 ms para [ʃ].

A duração das fricativas é considerada um parâmetro acústico robusto para diferenciar as fricativas vozeadas das não-vozeadas. Em geral, as fricativas não vozeadas são mais longas do que as vozeadas. Esse fato foi observado nos trabalhos de Samczuk e Gama-Rossi (2004) e Haupt (2007) para o português brasileiro. Entretanto, os resultados aqui apresentados não possibilitam afirmar que a duração seja uma pista segura para diferenciar fricativas vozeadas das não-vozeadas, já que a diferença entre elas não foi significativa ($t = 0,85$; $p = 0,43$).

3.2.1.2 Formantes e picos de amplitude do ruído fricativo

A Tabela 2 apresenta os valores dos seis formantes (F1-F6) das fricativas [s] e [ʃ]. Esses valores não apresentam grande importância para a síntese desses sons. No entanto, só é possível obter os valores dos picos de amplitude dos sons fricativos a partir dos valores dos formantes. Para a síntese das fricativas, o que importa são os valores dos picos de amplitude A2F a A6F, que definirão as regiões do espectro que serão excitadas pelo ruído de fricção. Dessa forma, a discussão desse tópico vai concentrar-se nos valores de amplitude, e a divulgação da Tabela 2 é apenas informativa. A última linha da tabela apresenta os resultados dos testes de diferença entre as médias. Os valores de F6 não condizem com normalidade de dados, nem para [s] ($W = 0,76$; $p = 0,016$) nem para [ʃ] ($W = 0,624$; $p < 0,001$).

TABELA 2 – Frequência dos formantes das fricativas [s] e [ʃ] (Hz)

Informante	Fricativa	F1	F2	F3	F4	F5	F6
1	[s]	1904	3369	4635	5396	6224	7116
	[ʃ]	2282	2999	3753	4913	5791	6810
2	[s]	1562	2719	3951	5063	5752	6390
	[ʃ]	1721	2360	3108	4399	5307	6225
3	[s]	1430	2612	3856	5104	5762	6377
	[ʃ]	1943	2554	3627	5430	5541	6270
4	[s]	1547	2813	4148	5120	5837	6440
	[ʃ]	2248	2846	3410	4439	5490	6242
5	[s]	1734	2908	4106	5120	5742	6406
	[ʃ]	1976	2781	3460	4722	5397	6153
6	[s]	1416	2973	4126	4723	5401	6180
	[ʃ]	1993	2451	3287	4432	5519	6167
7	[s]	1674	2980	4337	5061	5892	6505
	[ʃ]	1885	2662	3426	4594	5462	6208
Média	[s]	1609,5	2910,5	4165,5	5083,8	5801,4	6487,7
	[ʃ]	2006,8	2664,7	3438,7	4704,1	5501	6296,4
dp	[s]	174,2	242,9	257,2	196,4	243,5	294,5
	[ʃ]	198,1	226,9	211,3	369,7	150,6	230
Diferença		5,11	-3,27	-8,09	-2,949	-3,967	28*
		(0,002)	(0,017)	(0,0002)	(0,025)	(0,007)	(0,015)

Fonte: Elaborada pelos autores.

Legenda: dp = desvio padrão;

Diferença = estimada pelo teste t de Student para amostras pareadas (entre parênteses, a probabilidade associada ao teste t), exceto (*) estimada pelo teste de Wilcoxon.

A tabela abaixo apresenta os picos de amplitude das fricativas [s] e [ʃ], propriedade acústica que, tanto quanto se saiba, ainda não foi investigada em nenhum estudo realizado para o Português Brasileiro.

TABELA 3 – Amplitude dos formantes das fricativas [s] e [ʃ] (dB)

Informante	Fricativa	A2	A3	A4	A5	A6
1	[s]	17	29	39	38	29
	[ʃ]	43	40	35	34	24
2	[s]	33	34	44	50	41
	[ʃ]	61	50	37	38	30
3	[s]	35	39	44	44	34
	[ʃ]	44	44	39	35	27
4	[s]	31	34	45	44	36
	[ʃ]	55	48	39	37	28
5	[s]	17	23	46	43	28
	[ʃ]	49	41	35	36	26
6	[s]	36	47	52	43	39
	[ʃ]	59	52	46	44	40
7	[s]	30	41	44	47	39
	[ʃ]	57	58	45	40	37
Média	[s]	28,4	35,2	44,8	44,1	35,1
	[ʃ]	52,5	47,5	39,4	37,7	30,2
dp	[s]	8	7,9	3,8	3,7	5
	[ʃ]	7,3	6,4	4,4	3,4	5,9
Diferença		8,762	5,944	-3,99	-4,172	-3,104
		(<0,001)	(0,001)	(0,007)	(0,006)	(0,021)

Fonte: Elaborada pelos autores.

Legenda: dp = desvio padrão;

Diferença = estimada pelo teste t de Student para amostras pareadas (entre parênteses, a probabilidade associada ao teste t).

A Tabela 3 mostra que as fricativas [s] e [ʃ] diferem substancialmente em relação aos valores dos picos de amplitude. A fricativa [s] apresenta as maiores amplitudes concentradas em regiões mais altas de

frequência quando comparada com [ʃ]. Assim, a fricativa [s] apresentou valores mais altos da quarta e da quinta amplitude (A4F e A5F), enquanto a fricativa [ʃ] apresentou valores mais altos da segunda e da terceira amplitude (A2F e A3F).

Os informantes 1 e 5 apresentaram valores mais baixos da segunda e terceira amplitudes (A2F e A3F) para a fricativa [s] em relação aos demais informantes. Dessa forma, o desvio padrão foi maior nesse caso. Porém, tal fato não foi relevante, uma vez que se analisou a tendência geral do grupo. Se esses informantes fossem excluídos da análise, teríamos os valores de 33 dB para A2F e 39 dB para A3F e, ainda assim, a quarta e a quinta amplitude apresentariam os maiores valores. Tanto o informante 1 quanto o informante 5 apresentaram A4F e A5F mais altos em relação aos demais valores de amplitude (A2F, A3F e A6F). O mesmo ocorreu com a fricativa [ʃ]. Esses informantes apresentaram valores mais baixos da segunda e terceira amplitudes quando comparados ao grupo e, conseqüentemente, o desvio padrão foi maior.

Comparando as tabelas acima, observa-se que, como a fricativa [s] apresentou valores maiores da quarta e da quinta amplitude (A4F e A5F), os picos espectrais mais proeminentes desse segmento estão entre 5 kHz e 5,8 kHz. Esses valores de frequência foram extraídos da Tabela 2 e representam os valores médios da quarta e da quinta frequência (F4 e F5) da fricativa [s] (5083,8 Hz e 5801,4 Hz respectivamente). Analisando a fricativa [ʃ], observa-se que os picos espectrais mais proeminentes concentram-se entre 2910,5 Hz e 3438,7 Hz, ou seja, em regiões de frequências mais baixas quando comparada com [s]. Essa questão será confirmada adiante com a extração dos valores do centro de gravidade.

3.2.1.3 Centro de gravidade do ruído fricativo

De acordo com a descrição realizada na seção 2, o centro de gravidade (CG) pode ser definido como a frequência abaixo (ou acima) da qual estão concentradas 50% da energia de determinado som. A Tabela 4 mostra os valores desse parâmetro para as fricativas [s] e [ʃ].

TABELA 4 – Centro de gravidade das fricativas [s] e [ʃ] (Hz)

Informante	Fricativa [s]	Fricativa [ʃ]
1	5717	3499
2	5349	2601
3	4994	3200
4	5649	3097
5	5395	3461
6	4401	2745
7	5112	3166
Média	5230	3109,8
dp	449,2	336,2

Fonte: Elaborada pelos autores.

Legenda: dp = desvio padrão

Conforme esperado, a fricativa [s] apresentou valor significativamente maior de CG do que a fricativa [ʃ] ($t = -13,91$, $p < 0,001$). Quando comparado com os demais estudos, observa-se que os valores do CG obtidos com o estudo piloto são inferiores em relação aos valores divulgados na pesquisa de Jongman, Wayland e Wongs (2000) para o inglês americano. O CG da fricativa [s] foi 6882 Hz, e o da fricativa [ʃ] foi 3712 Hz. Os valores obtidos aqui são mais próximos dos valores descritos no trabalho de Gordon, Barthmaier e Sands (2002). O valor maior de CG para a fricativa [s] foi 5463 Hz, e o menor para a fricativa [ʃ] foi 4134 Hz. Embora não relatem valores de CG para o português brasileiro, Russo e Behlau (1993) observaram que a fricativa [s] tem faixas de frequência acima de 4,5 kHz chegando a 8 kHz, enquanto a fricativa [ʃ] apresenta uma faixa de frequência entre 2,5 e 6 kHz, valores compatíveis com os alcançados neste estudo.

3.2.2 Propriedades acústicas das plosivas [b] e [d]

3.2.2.1 Duração absoluta dos segmentos

A medida de duração absoluta dos segmentos plosivos foi obtida por meio da análise da fase de pré-sonorização e da fase de soltura da oclusão, ou seja, *burst*, conforme mostra a Tabela 5.

TABELA 5 – Duração das fases de pré-sonorização e de soltura da oclusão em [b] e [d] (ms)

Informante	Pré-sonorização		Soltura da oclusão	
	[b]	[d]	[b]	[d]
1	146	96	13	13
2	74	52	11	12
3	101	56	12	15
4	97	80	12	9
5	93	75	11	10
6	92	72	9	12
7	108	73	12	11
Média	101,6	72	11,4	11,7
dp	22,2	14,75	1,3	1,9

Fonte: Elaborada pelos autores.

O valor de duração absoluta da fase de pré-sonorização encontrado neste estudo foi maior para a plosiva [b] quando comparado com a plosiva [d] ($t = 5,71$; $p = 0,0012$). Outros trabalhos do PB também revelaram resultados semelhantes. Melo *et al.* (2011) investigou e comparou as características acústicas das plosivas vozeadas e não vozeadas na fala de crianças com desenvolvimento fonológico típico e de adultos, que pertenciam ao grupo controle. Os pesquisadores analisaram a fala de 11 adultos e obtiveram os valores de duração da fase de oclusão de 98 ms para [b] e 91 ms para [d], em posição de *onset* medial. Barbosa (1999) também verificou valores da fase de pré-sonorização maiores para a plosiva [b] (86 ms) quando comparada com a plosiva [d] (71 ms). Da mesma forma, em seu estudo dos parâmetros acústicos de plosivas realizado com cinco informantes do sexo feminino, oriundas da região de Criciúma, no sul do Estado de Santa Catarina, Alves (2015) observou que a plosiva [b] apresentou o valor de duração absoluta de 101 ms, enquanto a plosiva [d] apresentou a duração de 94 ms.

Por fim, a Tabela 5 mostra os valores de duração da fase de soltura da oclusão para as plosivas [b] e [d], onde observa-se que os valores desse parâmetro foram muito próximos, sem diferenças significativas entre esses dois segmentos ($t = -0,34$; $p = 0,74$). Lieberman e Blumstein

(1988) verificaram que a duração do *burst* pode ser de 5 a 15 ms, valores corroborados pelos achados deste estudo.

3.2.2.2 Picos de amplitude e formantes da soltura da oclusão

Nesta seção apresenta-se uma análise qualitativa dos parâmetros de amplitude e dos formantes do *burst* das plosivas [b] e [d]. Essa análise tornou-se necessária pela curta duração desse parâmetro. Dessa forma, em 11 produções da plosiva [b] e em 5 produções da plosiva [d] não foi possível obter o valor exato do sexto formante (F6). No entanto, como as plosivas não têm valores altos de frequência que ultrapassem os 6 kHz, o valor do sexto formante não foi considerado relevante para a posterior síntese dos estímulos. Assim como nos sons fricativos, são os picos de amplitude que irão determinar as regiões do espectro das plosivas que serão excitadas pelo ruído de fricção.

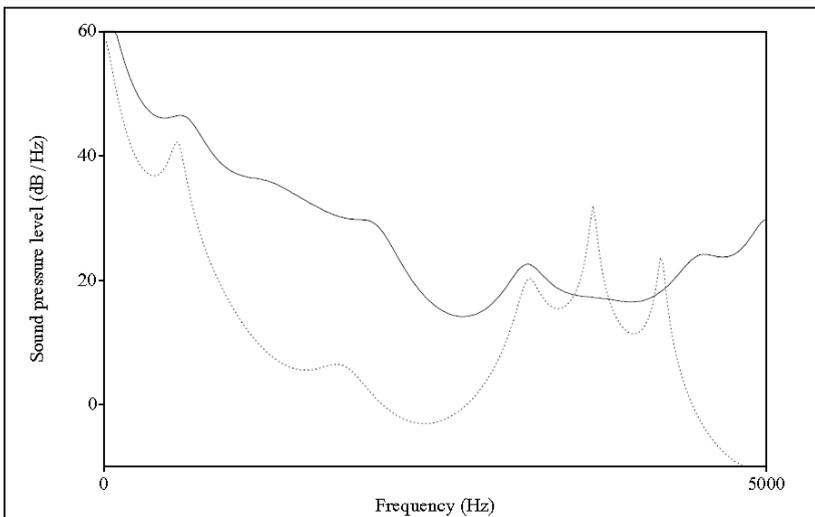
Experimentos perceptuais realizados com as plosivas sintetizadas no inglês, mostraram que o *burst* é considerado uma pista auditiva importante para que os ouvintes façam a distinção entre essas categorias de sons. Esses estudos também verificaram que o pico espectral do *burst* apresentou valores mais altos para as plosivas alveolares (acima de 4 kHz) quando comparado com as plosivas bilabiais (entre 0,5 kHz e 1,5 kHz) (STEVENS; BLUMSTEIN, 1977; HALLE; HUGHES; RADLEY, 1957).

No português brasileiro, Alves (2015) realizou um estudo das características espectrais do *burst* das plosivas vozeadas e não vozeadas, que envolveu a análise dos picos e dos momentos espectrais (centróide ou média, variância, assimetria e curtose). Direcionando-se à análise das plosivas vozeadas [b] e [d], assim como no inglês, a pesquisadora também observou que a plosiva bilabial apresentou pico de energia em regiões mais baixas de frequência quando comparada com a plosiva alveolar (1189 Hz e 2324 Hz; respectivamente). No entanto, apesar de a plosiva alveolar apresentar o valor do pico espectral mais elevado do que a plosiva bilabial, esse valor ficou abaixo dos valores obtidos no inglês.

Neste estudo, não se realizou uma análise detalhada dos momentos espectrais do *burst*, uma vez que essas informações não são relevantes para a posterior síntese dos estímulos. Conforme relatado, a análise do *burst* concentrou-se nos picos de amplitude, parâmetro fundamental para a geração dos estímulos sintéticos. Dessa forma, após a extração e a análise das amplitudes do *burst*, observou-se que a plosiva

bilabial [b] apresentou picos de energia mais intensos em regiões mais baixas de frequência, referente à região do segundo e terceiro formantes (A2F-A3F). Já a plosiva alveolar [d] apresentou picos de energia mais intensos em regiões mais altas de frequência, compatível com a região de frequência do quarto e do quinto formantes (A4F-A5F). A forma espectral da plosiva [b] foi mais plana quando comparada com a plosiva [d] (FIG. 4). À esquerda, observa-se a forma espectral do *burst* da plosiva [b] na palavra /bata/ e da plosiva [d] na palavra /data/. Essa tendência corrobora os achados descritos anteriormente para o inglês e para o PB.

FIGURA 4 – Forma espectral do *burst* da plosiva [b] (linha contínua) e da plosiva [d] (linha pontilhada) obtidas de um dos informantes do estudo



Fonte: Elaborada pelos autores.

A seguir, serão abordados os últimos parâmetros analisados neste estudo que são as medidas de duração vocálica e dos valores dos formantes da vogal [a] seguinte às fricativas e às plosivas. Os valores de duração da vogal foram obtidos apenas com o intuito de contribuir com a síntese de fala baseada nas propriedades acústicas dos falantes nativos do PB.

3.2.3 Propriedades acústicas da vogal [a] seguinte às fricativas

A Tabela 6 mostra os valores de duração da vogal [a] seguinte às fricativas [s] e [ʃ]. Observa-se que a vogal [a] seguinte às fricativas [s] e [ʃ] apresentou valores de duração próximos, porém estatisticamente diferentes ($t = 4,38$; $p = 0,005$).

TABELA 6 – Duração da vogal [a] seguinte às fricativas (ms)

Informante	[s_]	[ʃ_]
1	127	129
2	142	155
3	173	178
4	150	155
5	119	131
6	145	149
7	170	177
Média	146,6	153,4
dp	20,1	19,5

Fonte: Elaborada pelos autores.

Legenda: [s_] = valores de duração da vogal [a] seguinte a [s];
 [ʃ_] = valores de duração da vogal [a] seguinte a [ʃ];
 dp = desvio padrão

A Tabela 7 apresenta os valores dos três primeiros formantes da vogal [a], na região de transição dos formantes e na região estável da vogal. Os valores dos formantes na região de transição foram retratados na tabela com a palavra “inicial”, ou seja, os valores do F1, F2 e F3 inicial relacionam-se à região de transição dos formantes entre o segmento fricativo e a vogal. Além disso, a descrição [s_] na coluna “Contexto” refere-se aos valores dos formantes da vogal [a] seguinte a [s], enquanto a descrição [ʃ_] refere-se aos valores dos formantes da vogal [a] seguinte a [ʃ].

TABELA 7 – Frequência dos formantes da vogal [a] seguinte às fricativas [s] e [ʃ] (Hz)

Informante	Contexto	F1 inicial	F1 estável	F2 inicial	F2 estável	F3 inicial	F3 estável
1	[s_]	526	717	1433	1413	3084	2837
	[ʃ_]	505	687	1550	1396	3077	2844
2	[s_]	529	739	1211	1210	2491	2447
	[ʃ_]	502	732	1530	1228	2331	2335
3	[s_]	496	661	1263	1208	2570	2397
	[ʃ_]	441	631	1607	1218	2602	2242
4	[s_]	554	679	1375	1415	2595	2503
	[ʃ_]	545	670	1694	1536	2775	2767
5	[s_]	616	768	1389	1366	2863	2584
	[ʃ_]	613	819	1579	1369	2878	2444
6	[s_]	537	662	1285	1311	2453	2310
	[ʃ_]	543	696	1511	1370	2375	2300
7	[s_]	582	768	1224	1348	2400	2366
	[ʃ_]	540	797	1641	1362	2984	2379
Média	[s_]	548,5	713,4	1311,4	1324,4	2636,6	2492
	[ʃ_]	527	718,8	1587,4	1354,1	2717,4	2473
dp	[s_]	39,7	46,9	87,1	86,7	247,8	176,8
	[ʃ_]	52,7	68,2	64,8	107,8	291,4	236,6
Diferença		-2,63	0,44	7,1	1,7	0,87	-0,35
		(0,039)	(0,673)	(<0,001)	(0,14)	(0,416)	(0,74)

Fonte: Elaborada pelos autores.

Legenda: [s_] = valores dos formantes da vogal [a] seguinte a [s];

[ʃ_] = valores dos formantes da vogal [a] seguinte a [ʃ];

dp = desvio padrão.

Conforme esperado, observa-se uma pequena variação nos valores do primeiro formante (F1), tanto na região de transição dos formantes, em que a diferença foi significativa, quanto na porção estável da vogal. Como este estudo analisou somente a vogal [a], esse fato era esperado, uma vez que o F1 relaciona-se ao grau de abertura de uma vogal e a altura da língua.

O valor do segundo formante (F2) na região de transição dos formantes foi maior para a vogal [a] seguinte à fricativa [ʃ] quando comparada com a fricativa [s] (1578 Hz e 1311 Hz, respectivamente), uma diferença altamente significativa. Esse fato também era esperado porque o valor do F2 está diretamente relacionado com o grau de anteriorização da língua. Como na produção do segmento [ʃ] a língua encontra-se mais posterior na cavidade oral quando comparado com o segmento [s], espera-se que valor do F2 seja mais alto para [ʃ]. No entanto, observa-se uma pequena variação nos valores médios da frequência do F2 da vogal [a] seguinte a fricativa [s], quando se compara os valores da região de transição dos formantes e da região estável da vogal (1311-1324 Hz).

Nos estudos de percepção da fala que investigam a ponderação de pistas auditivas,⁴ desenvolvidos por Nittrouer no inglês americano, a pesquisadora produziu a vogal [a] sintetizada que seguiu as fricativas [s] e [ʃ] com os seguintes valores (inicial e estável) dos formantes: F1 (450 Hz-650 Hz), F2 (1250 Hz-1130 Hz) e F3 (2464 Hz-2300 Hz) para a vogal [a] seguinte a fricativa [s]; F1 (450 Hz-650 Hz), F2 (1570 Hz-1130 Hz) e F3 (2000 Hz-2300 Hz) para a vogal [a] seguinte a fricativa [ʃ] (NITTROUER, 2002). Esses valores foram obtidos por meio da análise das propriedades acústicas de um falante nativo do inglês americano. Dessa forma, constata-se uma variação maior dos valores do F2 (inicial e estável) da vogal [a] seguinte a fricativa [ʃ] (440 Hz) quando comparado com a fricativa [s] (120 Hz), assim como se verificou no presente estudo. Porém, observa-se que essa variação nos valores do F2 (inicial e estável) da vogal [a] seguinte a fricativa [s], verificada neste estudo (1311 Hz-1324 Hz), foi menor quando comparada com os estudos desenvolvidos por Nittrouer (1250 Hz-1130 Hz).

⁴ A ponderação de pistas auditivas pode ser definida como o peso perceptivo atribuído pelos ouvintes às pistas auditivas ao distinguir os sons da fala.

3.2.4 Propriedades acústicas da vogal [a] seguinte às plosivas

A Tabela 8 mostra que a vogal [a] seguinte às plosivas [b] e [d] apresentou valores de duração muito próximos, cuja diferença não se revelou significativa ($t = 0,81$; $p = 0,45$).

TABELA 8 – Duração da vogal [a] seguinte às plosivas (ms)

Informante	[b_]	[d_]
1	185	133
2	143	101
3	165	188
4	145	151
5	149	189
6	171	165
7	204	157
Média	166	154,8
dp	22,7	30,9

Fonte: Elaborada pelos autores.

Legenda: [a_] /ba/ = valores de duração da vogal [a] seguinte a [b];

[a_] /da/ = valores médios da duração da vogal [a] seguinte a plosiva [d];

dp = desvio padrão

Os valores dos três primeiros formantes (F1, F2 e F3) da vogal [a] seguinte às plosivas [b] e [d] podem ser visualizados na Tabela 9.

TABELA 9 – Frequência dos formantes da vogal [a] seguinte às plosivas [b] e [d] (Hz)

Informante	Contexto	F1	F1	F2	F2	F3	F3
		inicial	estável	inicial	estável	inicial	estável
1	[b_]	603	715	1268	1374	2725	2804
	[d_]	419	721	1814	1608	2568	2396
2	[b_]	513	751	1062	1223	2397	2505
	[d_]	487	669	1717	1596	2297	2183
3	[b_]	457	625	1045	1016	2502	2417
	[d_]	439	750	1972	1537	2572	2572
4	[b_]	576	692	1199	1356	2233	2441
	[d_]	422	625	2090	1317	2708	2562
5	[b_]	524	744	1146	1258	2490	2480
	[d_]	518	803	1826	1742	2436	2546
6	[b_]	508	739	1111	1265	2352	2474
	[d_]	465	640	1996	1563	2806	2424
7	[b_]	586	720	1166	1288	2349	2321
	[d_]	544	889	1701	1485	2968	2310
Média	[b_]	538,1	712,3	1142,4	1254,3	2435,4	2491,7
	[d_]	470,6	728,1	1873,7	1549,7	2622,1	2427,5
dp	[b_]	52	43,4	77,9	118,1	157,1	150,2
	[d_]	48,2	94,7	148,1	129,8	226	145,8
Diferença		2,517	-0,397	-11,537	-4,103	1,545	0,776
		(0,045)	(0,705)	(<0,0001)	(0,006)	(0,17)	(0,467)

Fonte: Elaborada pelos autores.

Legenda: [b_] = valores dos formantes da vogal [a] seguinte a [b];

[d_] = valores dos formantes da vogal [a] seguinte a [d];

dp = desvio padrão

A Tabela 9 mostra que, na região de transição dos formantes, o valor de F2 para [da] é maior quando comparado com [ba]. Isso ocorre,

pelo fato da plosiva alveolar [d] apresentar ponto de articulação mais posterior do que a plosiva bilabial [b].

Nos estudos que investigaram a percepção do ponto de articulação das plosivas do inglês americano, desenvolvidos por Stevens e Blumstein (1978) e Walley e Carrell (1983), os pesquisadores produziram a vogal [a] sintetizada, que seguiu as plosivas [b] e [d], com os seguintes valores (inicial e estável) dos segundo e do terceiro formantes: F2 (900 Hz-1240 Hz) e F3 (2000 Hz-2500 Hz) para a vogal [a] seguinte a plosiva [b]; F2 (1700 Hz-1240 Hz) e F3 (2800 Hz-2500 Hz) para a vogal [a] seguinte a plosiva [d]. Estes estudos verificaram valores maiores do F2 para a vogal [a] seguinte a plosiva [d]. Além disso, observa-se um aumento de 340 Hz dos valores do F2 (inicial e estável) da vogal [a] seguinte a plosiva [b] e uma diminuição de 460 Hz dos valores do F2 da vogal [a] seguinte a plosiva [d].

No presente estudo, nota-se um aumento de 112 Hz dos valores médios do F2 (inicial e estável) da vogal [a] seguinte a plosiva [b] e uma diminuição de 324 Hz dos valores do F2 da vogal [a] seguinte a plosiva [d]. Por fim, a Tabela 9 revela um decréscimo nos valores do terceiro formante (inicial e estável) da vogal [a] seguinte a plosiva [d], o que foi verificado nos estudos desenvolvidos com o inglês (STEVENS; BLUMSTEIN, 1978; WALLEY; CARRELL, 1983). Assim como o F2, o terceiro formante também foi considerado uma pista importante para distinção do ponto de articulação das plosivas. Entretanto, no presente estudo, a diferença nos valores de F3 não foi significativa nem na posição inicial nem na parte estável da vogal.

3.3 Resumo dos achados da análise acústica

Neste estudo, analisaram-se os dados de fala de sete falantes nativos do PB com o intuito de caracterizar as propriedades acústicas das fricativas [ʃ] e [s] e das plosivas [b] e [d] seguidas da vogal [a] em posição inicial de palavra. Essa análise foi desenvolvida enfocando duas questões: quais as principais pistas auditivas que auxiliam os ouvintes a diferenciar esses sons e que parâmetros acústicos contribuem para uma posterior síntese desses sons em estudos perceptivos?

Com base na análise dos dados, verificou-se que as fricativas [ʃ] e [s], por serem sons não-vozeados, apresentaram valores de duração próximos. Isso é condizente com a observação que se nota na literatura, ou seja, de que a duração é utilizada principalmente para distinguir

vozeamento, característica não analisada aqui. No caso das plosivas [b] e [d], observou-se que a duração absoluta da fase de pré-sonorização foi maior para a plosiva [b] quando comparada com a plosiva [d]. No entanto, foram observados valores próximos de duração da fase de soltura da oclusão.

Em relação às principais pistas auditivas que contribuem para a distinção das fricativas [ʃ] e [s], ou seja, os picos de amplitudes dos sons fricativos e a região de maior concentração de energia, verificou-se que a fricativa [s] apresentou picos de amplitude mais proeminentes em regiões mais altas de frequência quando comparada com a fricativa [ʃ] e apresentou também valor médio maior de CG do que a fricativa [ʃ] (5231 Hz e 3110 Hz; respectivamente). Quanto às plosivas, observou-se que a plosiva alveolar [d] apresentou picos de energia mais intensos em regiões mais altas de frequência quando comparada com a plosiva bilabial [b].

Por fim, os resultados relacionados à vogal [a] mostraram pequenas diferenças nos valores de F1 e de F3 da vogal [a] seguinte às fricativas [ʃ] e [s], enquanto F2 apresentou valores mais altos para a vogal [a] seguinte a fricativa [ʃ]. Em relação a vogal [a] seguinte às plosivas, observaram-se valores próximos do F1 da vogal [a] tanto diante da plosiva [b] quanto da plosiva [d]. No entanto, verificou-se uma variação nos valores inicial e estável do F2 e do F3, com valores maiores desses formantes para a vogal [a] seguinte a plosiva [d].

Após descrição e análise das propriedades acústicas dos estímulos, na próxima seção será abordado o experimento de classificação.

4 Experimento de classificação

Este estudo tem o objetivo de investigar, por meio de uma tarefa de classificação, as pistas auditivas para diferenciar as palavras /ʃapa/-/sapa/ e /bata/-/data/. Essa tarefa induz a utilização de representações fonológicas, o que possibilita aos ouvintes demonstrarem percepção categórica. Com as informações obtidas no estudo anterior, de análise acústica, é possível conhecer valores que descrevem as variáveis em foco. Porém, a fala é variável, e é preciso, além disso, conhecer também qual o limite no qual esses parâmetros podem variar sem comprometer a identificação do segmento. Por meio da técnica apresentada nesta seção, é possível estimar até que ponto o ouvinte tolera a mudança em determinado parâmetro, a partir do qual ele/ela começa a ouvir outro segmento.

4.1 Métodos

4.1.1 Participantes do estudo

Participaram oito adultos do sexo feminino, falantes nativos do português brasileiro de Belo Horizonte, com faixa etária entre 20 e 45 anos de idade, sem histórico significativo de otite média, de colocação de tubo de ventilação ou de perda auditiva. Qualquer comprometimento da audição foi descartado por meio de triagem auditiva, utilizando um audiômetro AD28. Os participantes assinaram o Termo de Consentimento Livre e Esclarecido, após serem informados sobre os aspectos gerais da tarefa de classificação.

4.1.2 Estímulos

A síntese de estímulos foi realizada de acordo com Nittrouer e Miller (1996). Criaram-se dois contínuos de sons da fala para investigar a distinção de fricativas e dois, para a distinção de plosivas. Isto é, para as fricativas criaram-se duas séries de dez sons cada, com o primeiro som mais semelhante a [ʃ], e o último mais próximo a [s]. Para plosivas, duas séries entre [b] e [d]. Em cada série, um som difere pouco a pouco do som anterior, formando um contínuo que varia ao longo de uma única dimensão. Assim, manipulam-se duas variáveis: uma ao longo do contínuo e outra, entre os contínuos. Esse tipo de desenho experimental possibilita uma investigação do efeito perceptual das duas pistas. Um ouvinte que não é influenciado pela pista que muda entre os contínuos, ou seja, a pista de transição dos formantes para as fricativas e a pista de amplitudes do *burst* para as plosivas, perceberá os dois contínuos como sendo uma mesma série. Ao contrário, um ouvinte que é influenciado pela pista que muda entre os contínuos perceberá diferença entre as duas séries. Além disso, a pista que muda lentamente ao longo do contínuo possibilita avaliar em que ponto da série (isto é, a partir de quais características acústicas) o ouvinte passa a identificar o som como pertencente a outra categoria sonora.⁵

⁵ Os detalhes da investigação da percepção da fala por meio de tarefas psicoacústicas estão além dos limites do presente trabalho, que pode ser complementado com a discussão mais técnica sobre as tarefas, em SCHOUTEN; GERRITS; VAN HESSEN (2003). Uma breve introdução ao assunto em português encontra-se em SILVA; ROTHE-NEVES (2009).

Todos os estímulos foram sintetizados por meio do *software* Sensyn, implementação de KLATT (KLATT, 1990). O sintetizador de formantes proposto por Klatt (1979) requer dois tipos de parâmetros de controle: 12 parâmetros que permanecem constantes ao longo de todo o enunciado sintetizado e 48 parâmetros variáveis em que os valores podem ser modificados ao longo do tempo. Os parâmetros variáveis são atualizados a cada 5 ms. Os valores de referência desses parâmetros, utilizados para a síntese de todos os estímulos, foram baseados na amostra de fala de um dos informantes que participaram do estudo citado anteriormente (informante 5, Tabelas 8 e 10), que apresentou os valores dos formantes vocálicos mais próximos da média.

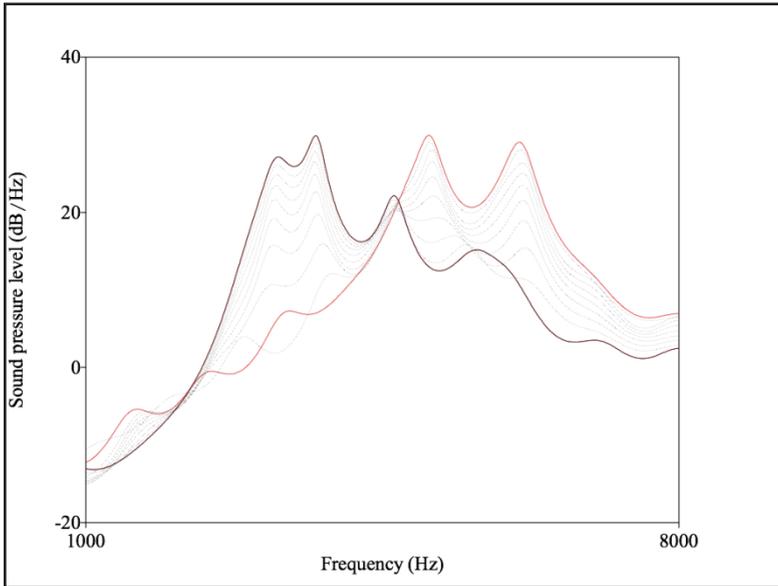
Os sons da primeira sílaba de cada palavra foram sintetizados separadamente, uma vez que as pistas auditivas dinâmicas e estáticas do experimento foram manipuladas somente na primeira sílaba de cada palavra. Por exemplo, a fricativa [s] e a vogal [a] da primeira sílaba da palavra /sapa/ foram sintetizadas separadamente, enquanto a segunda sílaba /pa/, da palavra /sapa/, foi sintetizada em conjunto. Feito isso, os sons foram concatenados no Praat. A segunda sílaba das palavras foi sintetizada somente com o intuito de produzir a fala de forma natural, sendo que nenhuma pista auditiva foi manipulada.

4.1.2.1 Contínuos /sapa/-/fapa/

No caso das fricativas, a altura da frequência do ruído fricativo foi manipulada ao longo do contínuo e, entre os contínuos, a transição de F2 e F3 no início da vogal. Manipulou-se a altura da frequência do ruído fricativo a partir de uma frequência apropriada para [s] (5633 Hz) para uma frequência apropriada para [ʃ] (3830 Hz). Os demais oito ruídos do contínuo foram gerados por interpolação, amostra por amostra, entre as amplitudes relativas dos envelopes espectrais desses dois sons de ruído.⁶ O primeiro e o décimo ruído constituíram as extremidades do contínuo fricativo. Observa-se, na Figura 5, que o ruído 1 (em preto, linha contínua) apresenta os picos de amplitude em regiões de frequência mais baixas quando comparado ao ruído 10 (em vermelho). Dessa forma, o ruído 1 representou a fricativa [ʃ], enquanto o ruído 10 representou a fricativa [s]. A duração dos ruídos fricativos foi de 160 ms, e a intensidade foi ajustada em 55 dB com base na análise dos estímulos naturais de fala.

⁶ O script pode ser encontrado em <http://www.holgermitterer.eu/HM/sample_interpolation.praat>. Acessado em: 17 mar. 2017.

FIGURA 5 – Envoltórios espectrais obtidos em 50 ms em torno do ponto central do primeiro (em preto, linha contínua) e décimo (em vermelho) ruídos fricativos

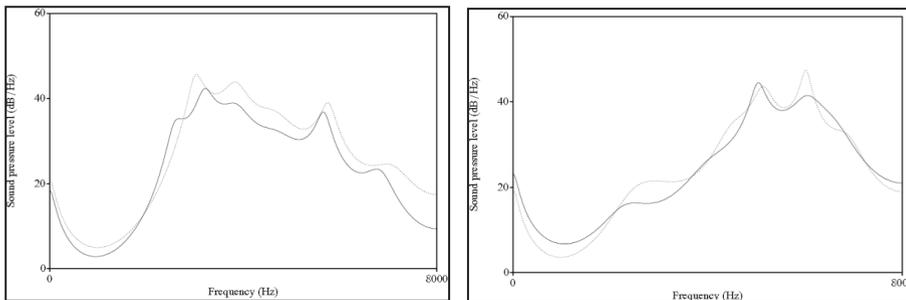


Fonte: Elaborada pelos autores.

Essa técnica de interpolação dos envelopes espectrais está de acordo com o fato de que é o todo que parece ser importante para a percepção da distinção de ponto, mais a transição de formantes subsequente, e não uma ou outra característica espectral (WHALEN, 1991). Por outro lado, isso coloca um problema para a manipulação experimental. A investigação psicofísica envolve a manipulação de uma variável contínua que provoca efeitos descontínuos (ou categóricos, neste caso). Em geral, utiliza-se a frequência central (*center frequency*) para indicar a altura da frequência do ruído fricativo. Mas a interpolação resulta numa série de ruídos cujos envelopes espectrais formam, de fato, um contínuo (como se vê na figura anterior), sem que isso se reflita num contínuo de valores de centro de gravidade cujos estímulos se diferenciem em passos de igual tamanho. A Tabela 11 abaixo mostra os valores obtidos de CG (em Hz) para cada ruído obtido por interpolação entre o primeiro e o último da série. Nota-se que a diferença de um valor a outro de frequência (coluna “ Δ Hz”) nunca é a mesma. Por isso, optou-se por utilizar como variável independente contínua a ordem dos

estímulos no contínuo (coluna “Item”). Para manter a homogeneidade de procedimentos, o mesmo foi utilizado para a classificação das plosivas.

FIGURA 6 – Envoltórios espectrais obtidos em 50 ms em torno do ponto central dos ruídos fricativos de [ʃ] (esquerda) e de [s] (direita).



Fonte: Elaborada pelos autores.

Legenda: Em linha contínua, sons originais obtidos de um dos informantes do estudo; em linha pontilhada, sons sintetizados.

A Figura 6 possibilita apreciar a qualidade da síntese, comparando-se os espectros obtidos de um informante com os sons sintetizados.

TABELA 10 – Valores de CG (Hz) do ruído fricativo dos estímulos produzidos por interpolação e a diferença (Hz) entre um estímulo e o antecedente no contínuo

Item	Centro de Gravidade (Hz)	Δ Hz
1	3830	182
2	4012	400
3	4412	440
4	4852	336
5	5188	211
6	5399	121
7	5520	65
8	5585	33
9	5618	15
10	5633	-

Fonte: Elaborada pelos autores.

Quanto à transição de formantes, sintetizou-se um contínuo com transições vocálicas apropriadas, para acompanhar [s], e outro, para acompanhar [ʃ]. A pista auditiva manipulada no caso das vogais foi a transição do segundo e do terceiro formantes (F2 e F3). A Tabela 12 apresenta os valores inicial e estável das frequências dos formantes vocálicos (F1, F2 e F3).

TABELA 11 – Valores das frequências dos formantes da vogal [a] seguinte às fricativas [s] e [ʃ] em (Hz)

Contexto	F1		F2		F3	
	inicial	estável	inicial	estável	inicial	estável
[s_]	616	768	1389	1366	2863	2584
[ʃ_]	613	819	1579	1369	2878	2444

Fonte: Elaborada pelos autores.

A duração da vogal foi de 140 ms, e a intensidade foi ajustada em 70 dB. Os valores dos formantes estabilizaram em 50 ms. A f0 do informante, copiada para a síntese, apresentou o valor inicial de 124 Hz e o valor final de 112 Hz.

4.1.2.2 Contínuos /bata/-/data/

Os contínuos compostos de par mínimo /bata/-/data/ foram criados de maneira diferente dos sons anteriores. Isso porque, no caso das plosivas, as duas pistas auditivas manipuladas foram a amplitude da explosão de soltura (*burst*) e a transição de F2 e F3 no início da vogal. Criaram-se duas amplitudes de *burst* e nove vogais que diferiram em relação aos valores de transição do segundo e do terceiro formantes (F2 e F3). Ou seja, diferentemente das fricativas, a pista que variou de forma idêntica nos dois contínuos do par mínimo /bata/-/data/ foi a transição de F2 e F3, com valores dos formantes apropriados para [b] num extremo até valores dos formantes apropriados para [d] no outro. Os valores iniciais de F2 foram modificados em passos uniformes de 88 Hz, de uma frequência apropriada para acompanhar [b] (vogal 1) para uma frequência apropriada para acompanhar [d] (vogal 9). Da mesma forma, os valores iniciais de F3 foram manipulados em passos uniformes de 34 Hz. A tabela a seguir mostra os valores inicial e estável de F2 e F3.

Os valores estáveis do F2 e do F3 também foram idênticos para todos os exemplares de vogal do contínuo, em 1400 Hz e 2460 Hz, respectivamente (TAB. 12). Esses valores foram obtidos pela média dos valores estáveis do F2 e do F3, descritos no estudo anterior. O primeiro formante (F1) apresentou valores idênticos para todas as vogais, iniciando em 503 Hz e alcançando o valor de 710 Hz em 50 ms. A duração das vogais [a] foi de 140 ms, e a intensidade foi ajustada em 70 dB. Os valores dos formantes estabilizaram em 50 ms. A f0 iniciou em 132 Hz e manteve-se 128 Hz ao longo da vogal.

TABELA 12 – Valores de frequência dos formantes do contínuo vocálico [a] seguinte às plosivas [b] e [d] (em Hz)

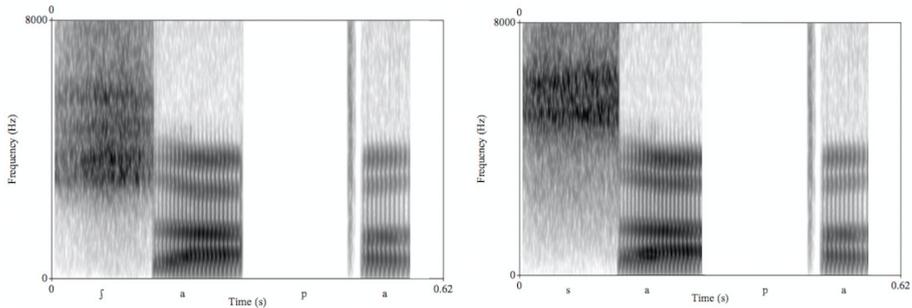
Valores das frequências dos formantes da vogal [a] sintetizada seguinte às plosivas				
Vogal	F2 inicial	F2 estável	F3 inicial	F3 estável
1	1111	1400	2352	2460
2	1199	1400	2386	2460
3	1287	1400	2420	2460
4	1375	1400	2454	2460
5	1463	1400	2488	2460
6	1551	1400	2522	2460
7	1639	1400	2556	2460
8	1727	1400	2590	2460
9	1815	1400	2624	2460

Fonte: Elaborada pelos autores.

A amplitude do *burst* foi a pista manipulada entre contínuos. Foram produzidas duas amplitudes do *burst*: uma amplitude com picos de energia mais intensos em regiões mais baixas de frequência (compatível com [b]), e outra com picos de energia mais intensos em regiões mais altas de frequência (compatível com [d]). A duração total dos estímulos para plosivas foi de 100 ms, isto é, 85 ms de fase de pré-sonorização e 15 ms de *burst*. A intensidade das plosivas foi ajustada em 55 dB.

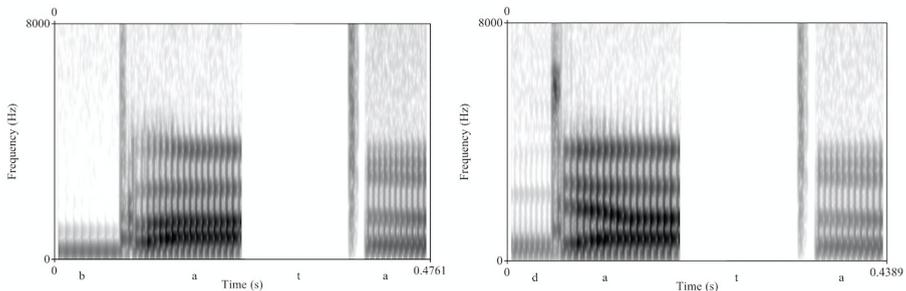
As figuras a seguir mostram os melhores exemplares dos pares mínimos /ɟapa/-/sapa/ e /bata/-/data/ após a síntese.

FIGURA 6 – Espectrogramas dos melhores exemplares das palavras fricativas sintetizadas. À esquerda /ʃapa/ e à direita /sapa/



Fonte: Elaborada pelos autores.

FIGURA 7 – Espectrogramas dos melhores exemplares das palavras plosivas sintetizadas. À esquerda /bata/ e à direita /data/



Fonte: Elaborada pelos autores.

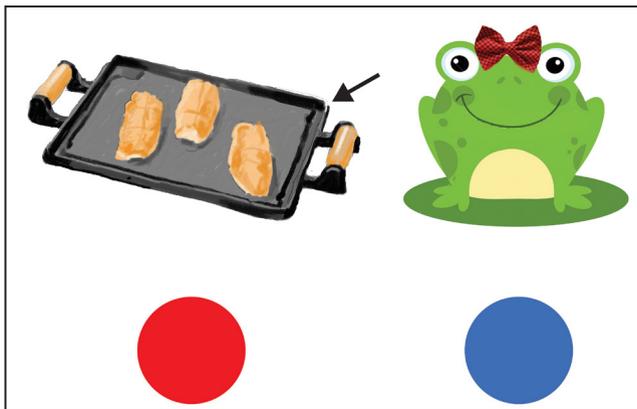
4.2 Procedimento experimental

Os estímulos foram apresentados aos participantes por meio de fones de ouvido *Philips*, acoplados a um *laptop HP Pavilion dv2000*. Com o *software PercEval* (versão 3.0.5.0), realizou-se uma tarefa de classificação em que os participantes ouviam cada estímulo e tinham de decidir em qual categoria classificá-lo entre as possibilidades dispostas na tela do *laptop* (por exemplo, /ʃapa/ ou /sapa/). As Figuras 8 e 9 mostram a tela de exibição dos estímulos no *software PercEval* para os pares mínimos /ʃapa/-/sapa/ e /bata/-/data/. Abaixo das imagens, à esquerda, inseriu-se um círculo vermelho e, abaixo das imagens, à direita, um

círculo azul. Os mesmos símbolos coloridos também estavam dispostos nas teclas “ctrl” do lado direito e do lado esquerdo do *laptop*, utilizadas pelos participantes para registro de sua resposta.

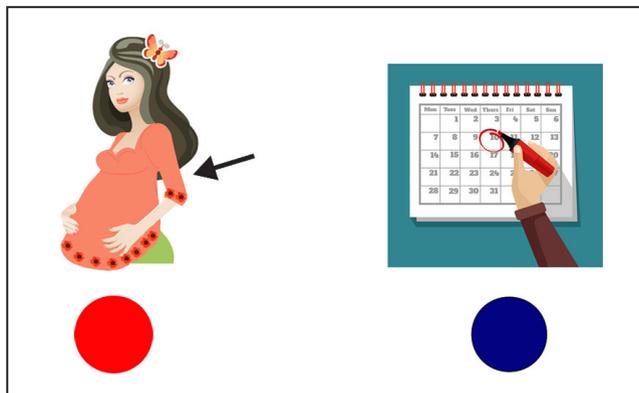
O intervalo para registro da resposta foi de cinco segundos, após o que, se o participante não respondesse, iniciava-se uma nova prova, com a tela vazia exibida por um segundo, antes da apresentação do próximo estímulo. Foram apresentadas seis repetições para cada estímulo, para ambos os pares mínimos.

FIGURA 8 – Tela de exibição do par mínimo /japa/-/sapa/



Fonte: Elaborada pelos autores.

FIGURA 9 – Tela de exibição do par mínimo /bata/-/data/



Fonte: Elaborada pelos autores.

4.3 Plano de análise

Para esse paradigma experimental, em que se obtêm respostas numa curva, a análise estatística mais apropriada é a regressão logística, que possibilita estimar a probabilidade de uma resposta binária (categórica) a partir da mudança em uma variável contínua. Utilizou-se a análise de regressão logística com função probito num modelo misto por meio do pacote *lme4* (BATES; MAECHLER; BOLKER; WALKER, 2015) do software R (R CORE TEAM, 2016).

Um modelo misto incorpora tanto efeitos fixos, que são os parâmetros associados a certos níveis reprodutíveis de fatores experimentais (isto é, a manipulação experimental), e fatores aleatórios, associados a unidades experimentais individuais amostradas aleatoriamente de uma população (isto é, aquela variação devida aos participantes da pesquisa) (PINHEIRO; BATES, 2000; QUENÉ; VAN DEN BERGH, 2004). Assim, modelos mistos levam em consideração a correlação entre as observações dentro de uma unidade experimental – ou medidas repetidas. Com isso, é possível modelar, ao mesmo tempo, a separação e a inclinação das curvas de resposta, bem como quantificar a importância da variância no nível dos sujeitos em apenas um passo analítico. Embora os modelos de efeitos mistos venham se tornando padrão na pesquisa linguística quantitativa (BAAYEN; BATES, 2008; JAEGER, 2008; JOHNSON, 2009; QUENÉ; VAN DEN BERGH, 2004, 2008), tanto quanto se saiba, apenas uma vez foi utilizado para investigar a ponderação de pistas na pesquisa de percepção da fala (CHRABASZCZ; WINN; LIN; IDSARDI, 2014).⁷

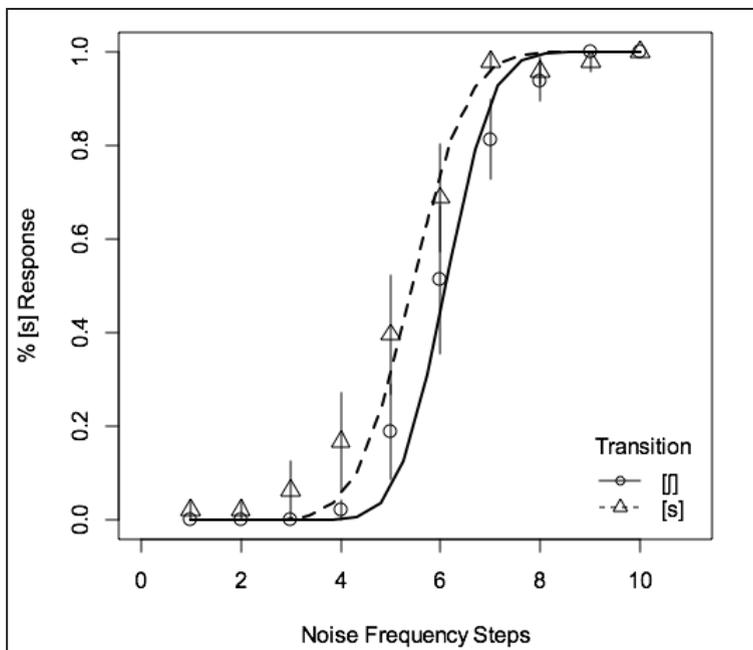
4.4 Resultados e considerações da tarefa de classificação

Os objetivos de uma tarefa de classificação são (1) verificar se a mudança progressiva na variável manipulada dentro do contínuo produziu o efeito de percepção categórica; (2) verificar se há diferença na inclinação da curva de respostas em função da variável manipulada entre os contínuos, o que indica interação entre as variáveis; e (3) verificar se a distância entre as curvas de respostas é significativa, o que indica

⁷ Mais recentemente, Nixon *et al.* (2016) investigaram a variação intracategórica na percepção de vozeamento e tom no cantonês, utilizando a mesma técnica de análise aplicada a dados obtidos a partir de outro método experimental.

a necessidade de dois níveis na variável entre contínuos (ou seja, que a manipulação entre contínuos se justifica).

FIGURA 10 – Curvas de resposta para classificação de fricativas



Fonte : Elaborada pelos autores.

Legenda: Círculos e triângulos representam a porcentagem média de resposta [(s)a] para cada nível de frequência de ruído em ambos os contínuos. Barras verticais indicam o erro padrão da média. Linhas contínua e pontilhada representam valores preditos pelo modelo não linear misto.

A Figura 10 apresenta as respostas dos participantes para classificação das fricativas, bem como resultados da análise estatística (erro padrão da média e curvas de valores preditos pelo modelo não linear misto). Nessa figura, apresenta-se a proporção em que os participantes indicaram ter ouvido /sapa/ entre todas as repetições de cada estímulo. Isso foi feito de modo a tornar possível visualizar um S alongado em função do aumento da frequência do ruído fricativo. Como já dito, a frequência do ruído fricativo é mais alta para [s] do que para [ʃ]. Os traços verticais representam o erro padrão da média, uma medida da

variabilidade de respostas dos participantes; note-se que os traços são mais longos no centro, na parte ascendente das curvas, que é a região de maior incerteza.

Consoante ao objetivo (1) supramencionado, a Figura 10 possibilita avaliar se a resposta dos participantes muda abruptamente de [ʃ] para [s]: os participantes apresentaram curvas de classificação íngremes, em formato de S alongado, o que indica o efeito de respostas categóricas. Além disso, observa-se que as curvas de classificação encontram-se mais íngremes quando se trata dos estímulos compostos das transições de formantes apropriadas para acompanhar [ʃ], e essa diferença de inclinação resulta em curvas que parecem mais próximas no topo da fase ascendente do que na sua base. Se confirmado pela análise estatística, isso quer dizer que há interação entre essas variáveis, a inclinação é diferente em função da transição de formantes que acompanha o ruído. Finalmente, o contínuo de estímulos com transições apropriadas para [s] mostra que o grupo atinge o limiar de 50% de respostas [s] entre o 5º e o 6º estímulos do contínuo, antes daquele com transições apropriadas para [ʃ], em que esse ponto é alcançado após o 6º estímulo. Isso quer dizer que, utilizando a transição adequada, ou seja, a pista que melhor serve à resposta [s], os mesmos ruídos são avaliados como pertencentes à outra classe, um passo antes no contínuo, o que indica o uso dessa pista auditiva pelos ouvintes. O modelo não linear misto serviu para avaliar estatisticamente essas observações.

O melhor modelo estatístico foi o que levou em consideração a interação entre Transição e Item nos efeitos fixos e incluiu diferentes interceptos e inclinações de curva para cada sujeito nos efeitos aleatórios. Em comparação a um modelo nulo que se obteve apenas com o intercepto e os efeitos aleatórios, a diferença para o melhor modelo foi muito significativa ($\chi^2(3) = 48,39$; $p < 0,0001$). Utilizando o teste das variáveis no modelo no formato mais comumente utilizado de análise de desvios (Type II Wald χ^2 tests), vê-se que tanto a variável Item ($\chi^2(1) = 46,7$; $p < 0,0001$) quanto a variável Transição ($\chi^2(1) = 22,7$; $p < 0,0001$) foram estatisticamente muito significativas, enquanto a interação entre elas foi marginalmente importante ($\chi^2(1) = 3,78$; $p < 0,052$), quase ficando abaixo do valor de $\alpha = 0,05$. Em outras palavras, tanto a mudança de altura das frequências mais intensas no ruído fricativo quanto a transição de formantes da vogal subsequente ao ruído foram pistas muito importantes para esse grupo de participantes distinguir entre /ʃapa/-/

sapa/, mas a interação entre essas pistas não foi significativa. Ou seja, o uso da altura de frequência não parece ter sido diferente em função da transição na vogal. Mesmo assim, o modelo que inclui a interação entre essas variáveis foi melhor para explicar a variância nos dados do que outro sem a interação ($\chi^2(1) = 4,12$; $p = 0,042$). Isso pode indicar que a quantidade de participantes não é suficiente para que o efeito da interação seja estatisticamente significativo.

Neste ponto, é oportuno lembrar que, na análise acústica, a diferença entre o início e o final da transição de formantes da vogal subsequente às fricativas foi maior para [ʃ] do que para [s]. Após [s], quase não se observou mudança nos valores de F2, entre a região de transição e a região estável, ao contrário dos falantes do inglês. Isso pode explicar a interação de pistas marginalmente significativa na percepção: seria natural que o uso dos formantes fosse mais importante para distinguir aquela categoria em que a transição é mais diferente.

Relacionada ao objetivo (3), já testado, a estimativa dos valores de fronteira entre as categorias sonoras /ʃ-s/ pode ser calculada por meio dos coeficientes do modelo, que incluiu a variação entre os sujeitos. Para cada participante, há um ponto no contínuo de altura de frequência do ruído fricativo em que se obteve 50% de respostas [s]. É o ponto de dúvida absoluta quanto à classificação do estímulo, que resulta numa resposta completamente aleatória. Na literatura, esse ponto é chamado de “fronteira de fonema”, a partir do qual a percepção de um estímulo como pertencente a uma categoria diminui e aumenta na outra. A Tabela 13 apresenta as fronteiras entre as categorias para cada participante, como um ponto no contínuo de estímulos apresentados e em valores de CG estimado (em Hz). O cômputo de CG foi realizado com base nos valores já apresentados, obtidos para cada estímulo do contínuo, bem como o tamanho da diferença entre eles. A grande variabilidade entre os participantes na região ascendente das curvas, já aparente no erro padrão das médias (FIG. 10), mostra-se aqui nos valores de diferença entre as fronteiras estimadas para cada participante em cada contínuo. Isso é comum em estudos de percepção da fala, pois o sistema perceptivo de cada um apresenta diferenças individuais impossíveis de controlar.

O valor identificado como “Geral”, na última linha da tabela, foi calculado utilizando-se apenas os efeitos fixos estimados, uma vez que a variação individual já foi apreciada nos efeitos aleatórios. Assim, não é um valor médio do grupo, mas um valor que representa a proporção

da variação comum ao grupo. Esses valores nos mostram que, embora a altura da frequência central típica para [ʃ] seja a do 1º estímulo do contínuo (3830 Hz), os falantes do português do Brasil podem tolerar até em torno de 5400 Hz e, ainda assim, classificar o som nessa categoria. Utilizando também a informação oferecida pela transição de formantes da vogal subsequente, porém, essa tolerância é menor, e, antes, o participante já começa a apresentar mais respostas na outra categoria, quando a altura da frequência central do ruído fricativo ultrapassa aproximadamente 5280 Hz.⁸

TABELA 13 – Estimativas de fronteira de fonema [ʃ-s] para cada participante

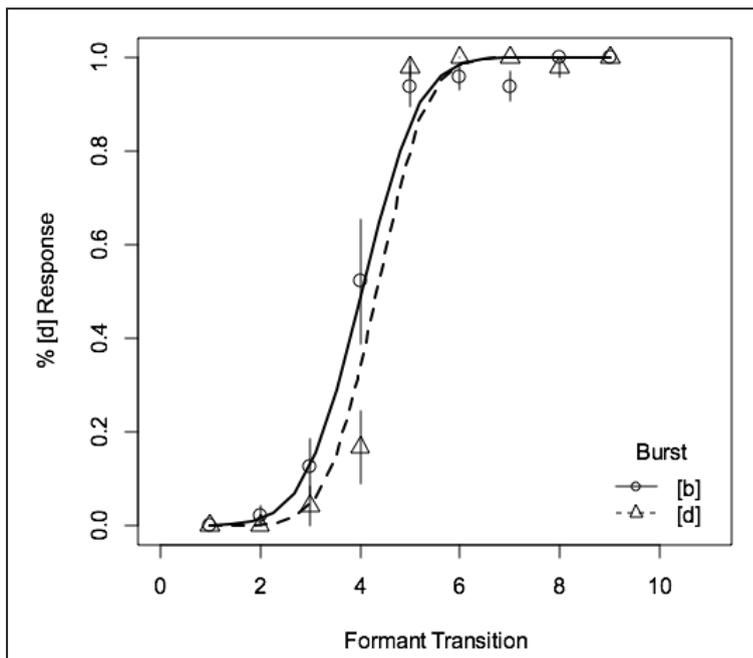
Participantes	Contínuo [ʃ]		Contínuo [s]		Δ Hz
	Step	CF (Hz)	Step	CF (Hz)	
CC	6,59	5470	6,10	5411	59
GF	6,21	5424	5,41	5274	150
LB	5,92	5308	5,39	5270	38
LD	4,57	5043	3,07	4443	600
MP	7,04	5523	6,53	5463	60
PS	5,63	5321	5,01	5190	131
RC	5,88	5374	4,41	4990	384
TM	6,54	5464	6,02	5401	63
Geral	6,10	5411	5,43	5279	132

Fonte: Elaborada pelos autores.

Na classificação das plosivas, os participantes também demonstraram percepção categórica, como pode ser visto na figura a seguir.

⁸ A bem da completude, um modelo que não leva em consideração a variável Transição possibilita estimar a fronteira em 5350 Hz. A estimativa da fronteira ($p=0,5$) é feita dividindo-se o valor negativo do intercepto estimado pelo modelo probito pelo valor estimado para Item ($-\beta_1/\beta_2$), ajustado para a diferença de Transição e interação, quando for o caso.

FIGURA 11 – Curvas de resposta para classificação de plosivas



Fonte: Elaborada pelos autores.

Legenda: Círculos e triângulos representam a percentagem média de resposta [(d)a] para cada nível de frequência inicial da transição de formantes em ambos os contínuos.

Barras verticais indicam o erro padrão da média. Linhas contínua e pontilhada representam valores preditos pelo modelo não linear misto

Consoante ao objetivo (1), a Figura 11 também possibilita afirmar que a resposta dos participantes muda abruptamente de [b] para [d], mostrando o efeito de respostas categóricas. Contudo, aqui as curvas de classificação não se encontram mais íngremes quando se trata dos estímulos compostos dos ruídos transientes (*burst*) apropriados para acompanhar [d]. De fato, a curva com os estímulos associados a [b] atingem um pouco antes o limiar de 50%. Se confirmado pela análise estatística, isso quer dizer que a pista do *burst* é de pouca valia para os falantes de português. Nesse sentido, é importante notar que as duas curvas não resultam muito separadas.

Numa primeira análise, verificamos se o modelo estatístico utilizado para avaliar os dados de classificação de fricativas também seria adequado para a classificação de plosivas. As variáveis independentes foram Burst (entre os contínuos) e Item, que aqui codifica a variação contínua da frequência inicial de F2 e F3 na transição de formantes que inicia a vogal subsequente à plosiva. Como fatores aleatórios, utilizaram-se diferentes interceptos e inclinações de curva para cada sujeito. Comparado ao modelo nulo, em que apenas a média geral dos sujeitos e sua variação individual são utilizados como preditores, o modelo com interação entre os fatores foi altamente significativo ($\chi^2(1) = 14,8$; $p = 0,00012$). Entretanto, o melhor modelo estatístico foi o que considerou como variável explicativa apenas Item, desprezando a variação entre os contínuos. Em comparação ao modelo com interação, o modelo simples (apenas Item) foi significativo, ou seja, explica melhor a variância dos dados para a quantidade de parâmetros utilizados ($\chi^2(2) = 6,14$; $p = 0,046$). Desse modo, o melhor modelo para explicar a tarefa de classificação foi aquele em que apenas a mudança ao longo do contínuo de formantes (variável Item) possibilita estimar a probabilidade de classificação em /bata/ ou /data/ ($\chi^2(1) = 36,63$; $p < 0,00001$). Isso quer dizer que os falantes adultos do português do Brasil não utilizaram a pista de ruído transiente (*burst*) para a classificação de plosivas que variam em ponto de articulação, ao contrário dos falantes do inglês.

TABELA 14 – Estimativas de fronteira [b-d] para cada participante

Participante	Fronteira	F2	F3
CC	4,82	1447	2482
GF	3,78	1356	2447
LB	3,89	1365	2450
LD	4,28	1400	2464
MP	4,09	1383	2457
PS	4,36	1407	2466
RC	4,32	1403	2465
TM	4,20	1393	2461
Geral	4,17	1390	2460

Fonte: Elaborada pelos autores.

A Tabela 14 apresenta as estimativas de fronteira de fonema entre /bata/ e /data/ obtidas com base nos parâmetros calculados para cada participante usando os efeitos fixos e os efeitos aleatórios. Neste caso, a variável contínua foi a ordem dos estímulos no contínuo. Assim, o modelo estatístico possibilita prever o ponto imaginário numa linha contínua de itens em que a curva atinge a probabilidade de 50% de respostas /data/ (coluna “Fronteira”). Importante lembrar que cada estímulo foi produzido com um ou outro ruído transiente (*burst*), como variável entre contínuos, e com diferentes valores para o ponto inicial das transições de F2 e F3. Para isso, produziram-se duas séries de valores, uma para F2, variando em passos de 88 Hz, e outra, para F3, em passos de 34 Hz (TAB. 12). Com base nos valores originais de F2 e F3 do estímulo, estimaram-se os valores dos pontos correspondentes a 50% de respostas /data/ (colunas “F2” e “F3”, respectivamente). A linha “Geral” ao final da tabela foi calculada utilizando-se os valores previstos levando-se em conta apenas os efeitos fixos do modelo final, sem considerar diferença entre contínuos. Portanto, os valores estimados representam os valores iniciais de F2 e F3 em transições de formantes na vogal subsequente às plosivas, para a qual cada falante teria 50% de chance de responder /pata/ ou /bata/.

Assim como para as fricativas, aqui também temos variação entre os indivíduos, que foi capturada pelo modelo estatístico. Os participantes GF e LB, à exceção dos demais, mudam de categoria antes do 4º estímulo do contínuo. Para o participante CC, a fronteira entre categorias está mais próxima do 5º estímulo do contínuo. A fronteira geral, obtida após levar em conta as diferenças individuais, resultou bastante próxima, mas justamente após o 4º estímulo do contínuo.

4.5 Resumo dos achados do experimento de classificação

O segundo estudo deste artigo investigou as pistas auditivas que podem ser utilizadas pelos ouvintes para diferenciar as palavras /japa-/ /sapa/ e /bata-/ /data/. Por meio de uma tarefa de classificação, buscou-se verificar: (1) se a mudança progressiva na variável manipulada dentro do contínuo produziu o efeito de percepção categórica; (2) se há diferença na inclinação da curva de respostas em função da variável manipulada entre os contínuos e (3) se a distância entre as curvas de respostas é significativa.

Os participantes apresentaram curvas de classificação íngremes, em formato de S alongado, diante do par mínimo /ʃapa/-/sapa/ e do par mínimo /bata/-/data/. Esse fato indicou o efeito de respostas categóricas. No par de palavras compostas das fricativas, observou-se que as curvas de classificação foram mais íngremes quando se tratava dos estímulos compostos das transições de formantes apropriadas para acompanhar [ʃ]. Além disso, o contínuo de estímulos com transições apropriadas para [s] mostrou que o grupo atingiu o limiar de 50% de respostas [s] entre o 5º e o 6º estímulos do contínuo, antes do que no outro contínuo, com transições apropriadas para [ʃ], em que esse ponto é alcançado após o 6º estímulo. No par de palavras compostas das plosivas, as curvas de classificação não foram mais íngremes quando se tratava dos estímulos compostos dos ruídos transientes (*burst*) apropriados para acompanhar [d], e a distância entre as curvas de resposta não foi significativa.

Em relação à importância que os ouvintes atribuem às pistas auditivas para classificar o par /ʃapa/-/sapa/ observou-se que tanto a mudança de altura das frequências mais intensas no ruído fricativo quanto a transição de formantes da vogal subsequente ao ruído foram pistas muito importantes para o grupo de participantes deste estudo distinguir entre /ʃapa/-/sapa/, mas a interação entre essas pistas não chegou a ser significativa. Diante do par /bata/-/data/ verificou-se que os falantes adultos do português brasileiro não utilizaram a pista de ruído transiente (*burst*) para a classificação de plosivas que variam em ponto de articulação.

5 Conclusões

Este artigo foi dividido em dois estudos. O primeiro teve o objetivo de analisar e descrever as propriedades acústicas das fricativas [ʃ] e [s] e das plosivas [b] e [d] associadas à vogal [a] em posição inicial de palavra. Feita a análise, os estímulos foram sintetizados. O segundo teve o objetivo principal de investigar, por meio de uma tarefa de classificação, as pistas auditivas utilizadas para diferenciar as palavras /ʃapa/-/sapa/ e /bata/-/data/.

Os resultados do primeiro estudo mostraram que as fricativas [ʃ] e [s] apresentaram valores de duração próximos, assim como a fase de soltura da oclusão (*burst*) das plosivas [b] e [d]. Já a fase de pré-sonorização foi maior para a plosiva [b] quando comparada com a plosiva

[d]. Em consonância com os estudos realizados com as fricativas do inglês americano, do português europeu, do espanhol argentino e do italiano, observou-se que a fricativa [s] apresentou picos de amplitude mais proeminentes em regiões mais altas de frequência quando comparada com a fricativa [ʃ] e apresentou também valor médio maior do CG do que a fricativa [ʃ]. Diante das plosivas, verificou-se que a plosiva alveolar [d] apresentou picos de energia mais intensos em regiões mais altas de frequência quando comparada com a plosiva bilabial [b].

Os resultados direcionados à análise do segundo formante vocálico (F2) mostrou que a diferença entre o início e o fim da transição de formantes da vogal subsequente às fricativas foi maior para [ʃ] do que para [s]. Após [s], quase não se observou mudança nos valores de F2, entre a região de transição e a região estável, ao contrário dos estudos que se basearam nas propriedades acústicas dos falantes do inglês. Parece que, no inglês, a transição do F2 seguinte às fricativas é mais marcada, o que pode tornar a pista de transição do F2 mais informativa nesta língua quando comparada com o PB.

Por fim, os resultados do segundo estudo revelaram que os participantes apresentaram curvas de classificação íngremes, diante dos pares mínimos /ʃapa/-/sapa/ e /bata/-/data/, com efeito de respostas categóricas. No que diz respeito às pistas auditivas utilizadas para diferenciar as palavras /ʃapa/-/sapa/ e /bata/-/data/, observou-se que, para o par /ʃapa/-/sapa/, tanto a transição dos formantes vocálico quanto a altura das frequências mais intensas no ruído fricativo auxiliaram os ouvintes na distinção desse par. Para o par /bata/-/data/, verificou-se que os participantes desse estudo não utilizaram a pista de ruído transiente (*burst*) para a classificação das plosivas [b] e [d], como fazem os falantes nativos do inglês. Assim, os resultados deste estudo indicam diferenças linguísticas que afetam a percepção da distinção de ponto de articulação das fricativas e das plosivas por falantes do PB.

Agradecimentos

Ao CNPq, pela bolsa Pq 312277/2015–6 para o primeiro autor, e à CAPES, pela bolsa de estágio sanduíche no exterior para a segunda autora.

Referências

ALVES, M. A. *Estudo dos parâmetros acústicos relacionados à produção das plosivas do Português Brasileiro na fala adulta: análise acústico-quantitativa*. 2015. Tese (Doutorado) – Universidade Federal de Santa Catarina, Florianópolis, 2015.

BARBOSA, P. A. Revelar a estrutura rítmica de uma língua construindo máquina falantes: pela integração da ciência e tecnologia de fala. In: SCARPA, E. (Org.). *Estudos de prosódia*. Campinas: Unicamp, 1999. p. 21-52.

BARBOSA, P. A.; MADUREIRA, S. *Manual de fonética acústica experimental*. São Paulo: Cortez, 2015.

BATES, D.; MAECHLER, M.; BOLKER, B.; WALKER, S. Fitting Linear Mixed Effects Models Using lme4. *Journal of Statistical Software*, Innsbruck, Austria, v. 67, p. 1-48, 2015. Doi: 10.18637/jss.v067.i01

BAAYEN, R. H.; DAVIDSON, D. J.; BATES, D. M. Mixed-effects Modeling with Crossed Random Effects for Subjects and Items. *Journal of Memory and Language*, Elsevier, v. 59, p. 390-412, 2008. Doi: <https://doi.org/10.1016/j.jml.2007.12.005>

BORZONE DE MANRIQUE, A. M.; MASSONE, M. I. Acoustic Analysis and Perception of Speech Fricative Consonants. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 69, p. 1145-1153, 1981. Doi: <https://doi.org/10.1121/1.385694>

BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer*. Versão 5.2.25. 2011. Disponível em: <www.praat.org>.

CHRABASZCZ, A.; WINN, M.; LIN, C. Y.; IDSARDI, W. J. Acoustic Cues to Perception of Word Stress by English, Mandarin, and Russian Speakers. *Journal of Speech, Language, and Hearing Research*, American Speech-Language-Hearing Association, v. 57, p.1468-1479, 2014. Doi: 10.1044/2014_JSLHR-L-13-0279

FANT, G. *Speech Sounds and Features*. Cambridge, MA: MIT Press, 1973.

FOX, John; WEISBERG, Sanford. *An {R} Companion to Applied Regression*. 2. ed. Thousand Oaks CA: Sage, 2011.

GORDON, M.; BARTHMAIER, P.; SANDS, K. A Cross-linguistic Acoustic Study of Voiceless Fricatives. *Journal of the International Phonetic Association*, Cambridge, v. 32, p. 141-174, 2002. Doi: <https://doi.org/10.1017/S0025100302001020>

GROSJEAN, Philippe; IBANEZ, Frederic. pastecs: Package for Analysis of Space-Time Ecological Series. R package version 1.3-18. Disponível em: <<https://CRAN.R-project.org/package=pastecs>>.

HALLE, M.; HUGHES, G. W.; RADLEY, J. P. A. Acoustic Properties of Stop Consonants. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 29, p. 107-116, 1957. Doi: <https://doi.org/10.1121/1.1908634>

HARRIS, K.S. Cues for the Discrimination of American English Fricatives in Spoken Syllables. *Language and Speech*, Sage Journals, v. 1, p. 1-7, 1958.

HAUPT, C. As fricativas [s], [z], [ʃ] e [ʒ] do português brasileiro. *Estudos Linguísticos*, Florianópolis, UFSC, v. XXXVI, n. 1, p. 37-46, 2007.

HEINZ, J. M.; STEVENS, K. N. On the Properties of Voiceless Fricative Consonants. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 33, p. 589-596, 1961. Doi: <https://doi.org/10.1121/1.1908734>

HIXON, T. Turbulent Noise Sources for Speech. *Folia Phoniatica*, Bethesda, USA, v. 18, p. 168-182, 1966. Doi: <https://doi.org/10.1159/000263090>

ISTRE, G. A fonética acústica. In: ISTRE, G. *Fonologia transformacional e natural: uma introdução crítica*. Florianópolis: Núcleo de Estudos Linguísticos, UFSC, 1983. p. 37-72.

JAEGER, T.F. Categorical Data Analysis: Away from ANOVAs (Transformation or Not) and Towards Logit Mixed Models. *Journal of Memory and Language*, Elsevier, v. 59, p. 434-446, 2008. Doi: <https://doi.org/10.1016/j.jml.2007.11.007>

JONGMAN, A.; WAYLAND, R. WONGS, S. Acoustic Characteristics of English Fricatives. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 108, p. 1252-1263, 2000. Doi: <https://doi.org/10.1121/1.1288413>

JOHNSON, D. E. Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-effects Variable Rule Analysis. *Language and Linguistics Compass*, Wiley Online Library, v. 3, p. 359-383, 2009. Doi: 10.1111/j.1749-818X.2008.00108.x

KLATT, D. H. Software for Cascade/Parallel Formant Synthesizer. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 67, p. 971-995, 1979. Doi: <https://doi.org/10.1121/1.383940>

KLATT, D. H.; KLATT, L. C. Analysis, Synthesis and Perception of Voice Quality Variations Among Male and Female Talkers. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 87, p. 820-856, 1990. Doi: <https://doi.org/10.1121/1.398894>

LACERDA, F. P. Acoustic Perceptual Study of the Portuguese Voiceless Fricative. *Journal of Phonetics*, v. 10, n. 1, p. 11-22, 1982.

LADEFOGED, P.; MADDIESON, I. *The Sound's of the World's Languages*. Massachusetts: Blackwell, 1996. p. 47-101.

LIEBERMAN, P.; BLUMSTEIN, S. E. *Speech Physiology, Speech Perception and Acoustic Phonetics*. Cambridge: Cambridge University Press, 1988. Doi: <https://doi.org/10.1017/CBO9781139165952>

MARCHAL, A.; REIS, C. *Produção da fala*. Belo Horizonte: Editora UFMG, 2012.

MELO, R. M.; MOTA, H. B.; MEZZOMO, C. L.; BRASIL, C. B.; LOVATTO, L.; ARZENO, L. Caracterização acústica da sonoridade dos fones plosivos do português brasileiro. *Revista CEFAC - Speech, Language, Hearing Sciences and Education Journal*, São Paulo, v. 14, p. 487-499, 2011. Doi: <http://dx.doi.org/10.1590/S1516-18462011005000143>

NITTROUER, S. Learning to Perceive Speech: How Fricative Perception Changes, and How it Stays the Same. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 112, p. 711-719, 2002. Doi: <https://doi.org/10.1121/1.1496082>

NIXON, J. S.; VAN RIJ, J.; MOK, P.; BAAYEN, R. H.; CHEN, Y. The Temporal Dynamics of Perceptual Uncertainty: Eye Movement Evidence from Cantonese Segment and Tone Perception. *Journal of Memory and Language*, Elsevier, v. 90, p. 103-125, 2016. Doi: <https://doi.org/10.1016/j.jml.2016.03.005>

PINHEIRO, J. C.; BATES, D. M. *Mixed-Effects Models in S and S-PLUS*. New York: Springer, 2000. Doi: <https://doi.org/10.1007/978-1-4419-0318-1>

QUENÉ, H.; VAN DEN BERGH, H. On Multi-level Modeling of Data from Repeated Measures Designs: A Tutorial. *Speech Communication*, Elsevier, v. 43, p. 103-121, 2004. Doi: <https://doi.org/10.1016/j.specom.2004.02.004>

QUENÉ, H.; VAN DEN BERGH, H. Examples of Mixed-effects Modeling with Crossed Random Effects and with Binomial Data. *Journal of Memory and Language*, Elsevier, v. 59, p. 413-425, 2008. Doi: <https://doi.org/10.1016/j.jml.2008.02.002>

R CORE TEAM. R: A language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria. Disponível em: <<https://www.R-project.org/>>.

REPP, B. H. [1984]. Categorical Perception: Issues, Methods, Findings. In: LASS, Norman J. (Ed.). *Speech and Language: Advances in Basic Research and Practice*. New York: Academic Press, 2016. p. 243-335.

RUSSO, I. C. P.; BEHLAU, M. S. *Percepção da fala: análise acústica do português brasileiro*. São Paulo: Lovise, 1993.

SAMCZUK, I. B.; GAMA-ROSSI, A. *Descrição fonético-acústica das fricativas do português brasileiro*. São Paulo: PUC, 2004.

SANTOS, M. T. M. *Uma análise espectrográfica dos sons fricativos surdos e sonoros do Português Brasileiro*. 1987. Monografia (Especialização em Fonoaudiologia) – Escola Paulista de Medicina, São Paulo, 1987.

SCHINDLER, O. Material fonemico. In: _____. *Manuale di audiofono: logopedia*. Torino: Editrice Omega, 1974.

SCHOUTEN, B.; GERRITS, E.; VAN HESSEN, A. The End of Categorical Perception as We Know It. *Speech Communication*, Elsevier, v. 41, p. 71-80, 2003. Doi: [https://doi.org/10.1016/S0167-6393\(02\)00094-8](https://doi.org/10.1016/S0167-6393(02)00094-8)

SILVA, D. M. R.; ROTHE-NEVES, R. Um estudo experimental sobre a percepção do contraste entre as vogais médias posteriores do português brasileiro. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, São Paulo, v. 25, n. 2, p. 319-345, 2009. Doi: <https://doi.org/10.1590/S0102-44502009000200005>

STEVENS, K. N.; BLUMSTEIN, S. E. Invariant Cues for Place of Articulation in Stop Consonants. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 64, p. 1358-1368, 1978. Doi: <https://doi.org/10.1121/1.382102>

STEVENS, K. N.; KLATT, D. H. *Studies of Acoustic Properties of Speech Sound*. Massachusetts: Air Force Cambridge Research Laboratories, 1968.

WALLEY, A. C.; CARRELL, T. D. Onset Spectra and Formant Transitions in Adult's and Child's Perception of Place of Articulation, *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 73, p. 1011-1022, 1983. Doi: <https://doi.org/10.1121/1.389149>

WHALEN, D. H. Perception of the English/s/-/ʃ/Distinction Relies on Fricative Noises and Transitions, Not on Brief Spectral Slices. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 90, p. 1776-1785, 1991. Doi: <https://doi.org/10.1121/1.401658>

WRIGHT, R. A Review of Perceptual Cues and Cue Robustness. In: BRUCE, H.; ROBERT, K.; DONCA, S. *Phonetically Based Phonology*. Cambridge: Cambridge University Press, 2004. p. 34-57. Doi: <https://doi.org/10.1017/CBO9780511486401.002>