

**PERCEPÇÃO DAS FRICATIVAS ESTRIDENTES: A PISTA ACÚSTICA PARA A
DISTINÇÃO ENTRE ALVEOLARES E PALATAIS UTILIZADAS PELOS
OUVINTES PESSOENSES**

Gabriela Belo da SILVA³⁹

Pedro Felipe de Lima HENRIQUE⁴⁰

Leonardo Wanderley LOPES⁴¹

Resumo: Este trabalho tem o objetivo de avaliar o pico espectral e a transição formântica como pistas para a percepção das fricativas [s] e [ʃ] pelos ouvintes de João Pessoa-PB. Para tanto, dois falantes nativos dessa cidade, foram treinados para produzir um gradiente de fricativas seguidas da vogal /a/, que foram apresentadas a 22 ouvintes pessoenses. Percebeu-se que as fricativas com primeiro pico nas regiões de frequência de 4,0 KHz para baixo foram associadas à palatal, e as com pico igual ou acima de 6,5 KHz, à alveolar. A transição formântica parece atuar como pista apenas nesse intervalo.

Palavras-chave: Pistas acústicas, pico espectral, transição formântica, percepção das fricativas, Português Brasileiro.

Abstract: This paper aims to evaluate the spectral peak and formant transition as cues to the perception of fricatives [s] and [ʃ] by listeners of João Pessoa. To this end, two native speakers of this city were trained to produce a spirants gradient of fricatives followed of the vowel / a /, which were presented to 22 listeners from João Pessoa. The results show that the fricatives with the first peak in the frequency regions of 4.0 KHz down were associated with palatal, and peaking at or above 6.5 kHz, the alveolar. The formant transition seems to act as track just in that range.

Keywords: *acoustic tracks, spectral peak, formant transition, perception, fricatives, Brazilian Portuguese.*

³⁹Doutoranda do Programa de Pós-Graduação em Linguística (PROLING) na Universidade Federal da Paraíba (UFPB), João Pessoa-PB, Brasil, gabesigo@gmail.com.

⁴⁰Mestrando do Programa de Pós-Graduação em Linguística (PROLING) na Universidade Federal da Paraíba (UFPB), João Pessoa-PB, Brasil, pedrofelipelh@hotmail.com.

⁴¹Pesquisador Doutor do Programa de Pós-Graduação em Linguística (UFPB) e professor do Departamento de Fonoaudiologia na Universidade Federal da Paraíba (UFPB), João Pessoa-PB, Brasil, lwlopes@hotmail.com.

Introdução

Considerando a escassez de trabalhos de percepção no português brasileiro, temos como objetivo avaliar a relação entre o pico espectral e a percepção dos fones [s] e [ʃ], pelos ouvintes de João Pessoa-PB. Para tanto, selecionamos dois falantes nativos de João Pessoa-PB, um do sexo masculino e outro do sexo feminino, que foram treinados para produzir um gradiente de fricativas com constrictões dos articuladores que vão do alvéolo até palato, seguidas de uma vogal central baixa [a]⁴².

Destarte, entendemos que este trabalho é relevante pois contribuirá para que seja possível descrever, de forma mais precisa, a relação entre o contínuo fonético e a categorização fonológica a partir de uma análise da realidade psicológica do som. Além disso, ao concentrarmo-nos especificamente nesses valores de fronteira, os resultados obtidos a partir desta pesquisa fornecerão uma indicação clara e mais precisa do que acontece no processamento perceptual como uma função da experiência, na distinção dos fonemas em análise. Outra contribuição é um panorama da visão do *como* o espaço fonético está sendo remapeado, em João Pessoa. Enquanto referencial teórico, nos embasaremos em Jhonson e Mulennix (1997), Gibson (1991[1977]), Edwards (1999) e Boersma (2011), quanto às teorias de percepção da fala, e Kent e Read (2015), Jongman et al. (2000) e Manrique e Massone (1981), sobre as pistas acústicas para a distinção de fricativas e sobre os resultados obtidos referentes à estudos envolvendo como elas são utilizadas pelos ouvintes de algumas línguas do mundo.

Estudos envolvendo a descrição acústica dos fones e sua relação com a respectiva percepção dessas produções podem contribuir também no contexto de avaliação, descrição e tratamento dos desvios de fala, sejam eles de origem fonética ou fonológica (BRASIL, et al., 2012; WIETHAN e MOTA, 2014; WIETHAN, et al., 2015). Descrições acústicas e seus impactos no nível perceptivo podem trazer uma maior compreensão dos mecanismos articulatorios subjacentes, que podem emergir como facilitadores ou não do input que é

⁴² A vogal /a/ foi escolhida em detrimento das outras vogais porque seus articuladores estão em uma posição mais neutra, o que diminui o efeito da coarticulação e possibilita construir melhor o gradiente entre a posição da língua mais anteriorizada, ponta discretamente elevada e corpo da língua baixo no [s] prototípico e o [ʃ] com posição de língua mais posteriorizada, ponta da língua baixa e corpo da língua elevado. A escolha da vogal /a/ permite partir de uma posição mais neutra para criar o gradiente (Cf. KENT e READ, 2015).

fornecido às crianças que apresentam desvios na fala, sejam eles caracterizados por omissões, substituições ou distorções, durante a abordagem terapêutica, trazendo maior objetividade ao trabalho que é realizado.

A produção realizada por esses falantes foi analisada acusticamente no programa PRAAT (BOERSMA e WEENNINK, 2012) de forma a obtermos realizações fonéticas das fricativas com picos espectrais entre as frequências de 3,56 e 9,12 KHz para a informante do sexo feminino e de 2,58 e 9,46 KHz para o informante do sexo masculino. Após esse procedimento, os áudios passaram por um processo de normalização, no programa Sound Forge 10.0.

Nesse ínterim, as fricativas isoladas e as sílabas formadas por estas e a vogal central baixa [a] foram apresentadas para vinte 22 ouvintes pessoenses, universitários, através do Aplicativo para Testes de Percepção (TP) (RAUBER et al., 2014). Com base na análise de cada som produzido, os ouvintes foram associando os segmentos ao grafema “s” ou “x”, e as sílabas, aos grafemas “sa” ou “xa”. Nossa hipótese é que os falantes associarão produções com picos espectrais abaixo de 4.5 KHz a sílaba aos grafemas “x” e “xa”, e produções com picos espectrais acima de 6 KHz, aos grafemas “s” e “sa”. Além disso, acreditamos que, nos intervalos espectrais com picos de 4.5 a 6 KHz, os falantes terão dificuldade por escolher um dos grafemas essa decisão não será categórica para nenhum das opções.

Sobre percepção da fala

Segundo Lopes (2012, p. 41), o estudo de percepção é um tema bastante pesquisado na psicologia social e tem sido gradualmente inserido nos estudos linguísticos. De acordo com Jhonson e Mulennix (1997) apud Lopes (2012, p. 12), os pesquisadores da percepção de fala estão empenhados em compreender as maneiras pelas quais “os seres humanos percebem, processam e codificam a linguagem falada” e, o papel dos diferentes tipos de variação no processamento da linguagem. Vale ressaltar que, no que tange a fala, o objetivo da percepção é apreender sua estrutura linguística, mais especificamente, a estrutura fonética (NITTROUER, 2002).

Nesse sentido, estamos compreendendo o conceito de percepção conforme Gibson (1991[1977], p. 417), para quem a percepção consiste na extração de informações sobre as coisas do mundo, ou seja, perceber significa retirar dados relevantes sobre os eventos que

tornem o ouvinte capaz de fazer um julgamento sobre o que é percebido. No entanto, apesar das pistas necessárias para a compreensão estarem sempre lá, isso não significa que todos os indivíduos recorrem às mesmas informações do mesmo modo. Na verdade, vários são os fatores que influenciam nesse processo de percepção, dentre os quais, o autor elenca três como sendo os mais importantes durante o processo: “a espécie do organismo, a maturidade de desenvolvimento e aprendizagem”.

Gibson (1991[1977]) apud Nittrouer (2002, p. 771) aponta que, quando se trata de compreender como as crianças aprendem a perceber o sinal de fala, o interesse maior está no papel que a aprendizagem desempenha para o desenvolvimento dessa competência, já que esta é o único dos três fatores que pode sofrer influência. Para defender sua tese, o autor aponta que pesquisas com falantes/ouvintes de diferentes línguas demonstraram, de forma robusta, que as informações extraídas do sinal de voz são altamente dependentes da língua de origem do indivíduo. Claramente, então, as crianças aprendem que informações dela devem ser extraídas, mas o que determina o que a criança precisa aprender a extrair? Gibson (1991[1977]) relata que, para responder a essa pergunta, é necessário pensar sobre o objetivo da percepção para a fala, que é “apreender a estrutura linguística” (apud NITTROUER, 2002, p. 771). Dessa forma, no que diz respeito à percepção de fala, crianças têm de aprender a extrair a informação que permita acesso a estrutura fonética em sua língua nativa.

Na acepção de Edwards (1999, apud LOPES, 2012), a percepção constitui-se como um filtro por meio do qual os dados sensoriais são analisados e posteriormente registrados, considerando o contexto de produção em que se insere o indivíduo e o conjunto de experiências que esse sujeito possui.

Boersma (2011, p.65) descreve a percepção, em termos estritamente linguísticos, como o mapeamento de um dado sensorio bruto para uma representação abstrata. O seu papel no sistema fonológico seria o mapeamento, para os ouvintes, de uma representação auditiva contínua bruta para uma discreta estrutura fonológica de superfície. Esse mapeamento pode ser descrito através do que os foneticistas no laboratório chamam de teste de identificação (BOERSMA, 2011, p.65).

Nesse sentido, é a partir de um dado contínuo acústico, os ouvintes têm a tarefa de associá-lo a uma determinada categoria discreta (fonema). O autor aponta a resistência de alguns fonólogos que tendem a argumentar que a fonética não é relevante (Hale e Reiss, 1998), ou que ela pode ser relevante, mas que sua modelagem não é importante para a teoria

fonológica (HAYES, 1999). Entretanto, com base na abundante existência de processos aparentemente fonéticos em fonologia segmental, Boersma (2011) cita que alguns fonólogos têm, não obstante, tentado incluir considerações fonéticas de força articulatória e contraste auditivo dentro do usual modelo fonológico de dois níveis: a forma subjacente e a forma de superfície (como a Teoria da Otimalidade Clássica e a Gramática Harmônica). Se apenas esses dois níveis forem considerados, deve-se propor que a estrutura fonológica de superfície de algum modo inclui detalhes do contínuo fonético. Não é essa, entretanto, a opinião do autor. Ele defende a existência de um nível fonético separado da estrutura fonológica de superfície, e este estabelece interface com aquela através de pistas acústicas. Para formalizar essa interface, utiliza a teoria do realismo direto da percepção da fala (FOWLER, 1986), propondo que a forma auditiva é diretamente interpretada em termos de gestos articulatórios e que sua percepção é conectada à fonologia.

Experimentos de percepção sobre distinção entre as estridentes

Para se analisar quais e como as pistas acústicas são utilizadas por ouvintes para acessar a estrutura fonológica de sua língua materna, é necessário considerar que propriedades acústicas podem ser associadas a gestos articulatórios distintos. No que se refere às propriedades acústicas das fricativas, Kent e Read (2015, p.270) afirmam que:

Vários candidatos podem ser considerados, incluindo momentos espectrais específicos (Behrens e Blumstein, 1988^a), pico espectral (Jongman et al., 2000) e medidas de inclinação espectral (Evers et al., 1998). Um eventualmente pode emergir como a característica preferida para todas as línguas em que o contraste é relevante. Entretanto, no momento, pode-se dizer que [s], comparado a [ʃ], tende a ter um pico espectral de frequência mais alta, maior assimetria (mas não uniformemente em todos os estudos), mais energia na região de frequência de 3,5-5,0 kHz (em oposição à região de frequência 2,5-3,5 kHz) e uma inclinação mais rasa para o envelope espectral abaixo de 2,5 kHz.

A partir dos dados apontados por Kent e Read, percebe-se que não há consenso sobre a existência de apenas uma propriedade acústica essencial para a distinção entre as fricativas alveolares e palatais, o que significa que mais de uma delas pode ser utilizada para a distinção fonológica entre /s/ e /ʃ/, apesar de não sabermos qual e se existe uma mais importante. Para a realização do nosso experimento, uma das variáveis selecionadas, dentro do conjunto de

propriedades citadas, foi o pico espectral, já que sobre ele parece haver mais dados. O estudo de Jongman et al. (2000) é um dos poucos em larga escala que analisou as pistas acústicas associadas ao ponto de articulação das fricativas. Os autores levaram em conta, nesse trabalho, três propriedades espectrais estáticas - pico espectral, duração do ruído e amplitude do ruído -, e, dentre elas, apenas o pico mostrou-se relevante para a distinção entre as fricativas estridentes: o valor médio de frequência associado a [s, z] foi 6882 Hz, enquanto que o valor médio associado a [ʃ, ʒ] foi de 3712 Hz (JONGMAN et al., 2000, p. 199).

Kent e Read (2015, p.268) apresentaram um estudo realizado por Manrique e Massone (1981) com falantes de espanhol como língua materna, no qual verificou-se que “a identificação de [s] pareceu depender em picos de energia em cerca de 5 a 8 kHz, enquanto a identificação de [ʃ] foi relacionada a um pico em cerca de 2,5 kHz”.

Analisando as características acústicas das fricativas surdas no Português Europeu (PE) a partir de testes perceptuais, Lacerda (1982) concluiu que o /s/ é melhor percebido quando o estímulo auditivo possui um alto nível de intensidade e picos espectrais na região de 5 kHz. O /ʃ/, por outro lado, é normalmente associado a altos níveis de intensidade associado a picos espectrais na região de 3 kHz. Outra análise envolvendo a produção dos mesmos segmentos no PE foi a realizada por Jesus (1999, apud HAUPT, 2007, p.40), cujos dados revelaram que o /s/ e /z/ têm seu pico principal na região dos 5 kHz e picos secundários entre 10 e 15 kHz, o /ʃ/ tem seu pico mais alto em torno de 2,5 kHz, e o /ʒ/ na faixa dos 2,7 kHz, com um pico secundário na região dos 11 kHz. Esses resultados ratificam que a localização da constricção é responsável pelas diferenças acústicas entre palatais e alveolares e o parâmetro se comporta da mesma forma para surdas e sonoras. Outro dado importante trazido à tona pelo estudo de Jesus (1999) foi a influência do contexto vocálico nas propriedades acústicas das fricativas, alterando na região de proeminência espectral desses segmentos. Segundo os resultados:

O pico do espectro do /s/ em início de palavra em contexto /ɔ/ é 1 kHz mais baixo em relação ao pico no espectro de um /s/ em contextos de /i/, /i/ e /e/. O pico do /z/ em início de palavra é 1,5 kHz mais baixo em contextos vocálicos /ɔ/ e /o/, e 500 Hz mais baixo quando seguido de /u/, em relação ao pico do espectro de /z/ no contexto vocálico de /i/ e /e/. O espectro do /ʒ/ em início de palavra seguido de /ε/ tem seu pico na faixa dos 4 kHz, com amplitude mais alta que o mesmo pico nos contextos de /ɔ/, /o/ e /u/ (JESUS, 1999, apud HAUPT, 2007, p.38).

Assim como, em dados de produção, as vogais parecem influenciar no espectro das fricativas adjacentes, em dados de percepção elas parecem auxiliar na identificação dos segmentos no que diz respeito ao ponto de articulação, dado que pode ser observado através da transição formântica. Quanto a essa pista para percepção das fricativas, Kent e Read (2015, p. 271) apontam que ela é “provavelmente secundária ao espectro de ruído como uma pista para a percepção das estridentes”. Abaixo, seguem dois espectrogramas, um de uma sílaba prototipicamente associada a “xa” e outro de uma sílaba prototipicamente associada a “sa”, de acordo com os padrões de pico apontados pela literatura, produzidos pelo mesmo falante pessoense e que serviu de estímulo para o teste de percepção deste trabalho:

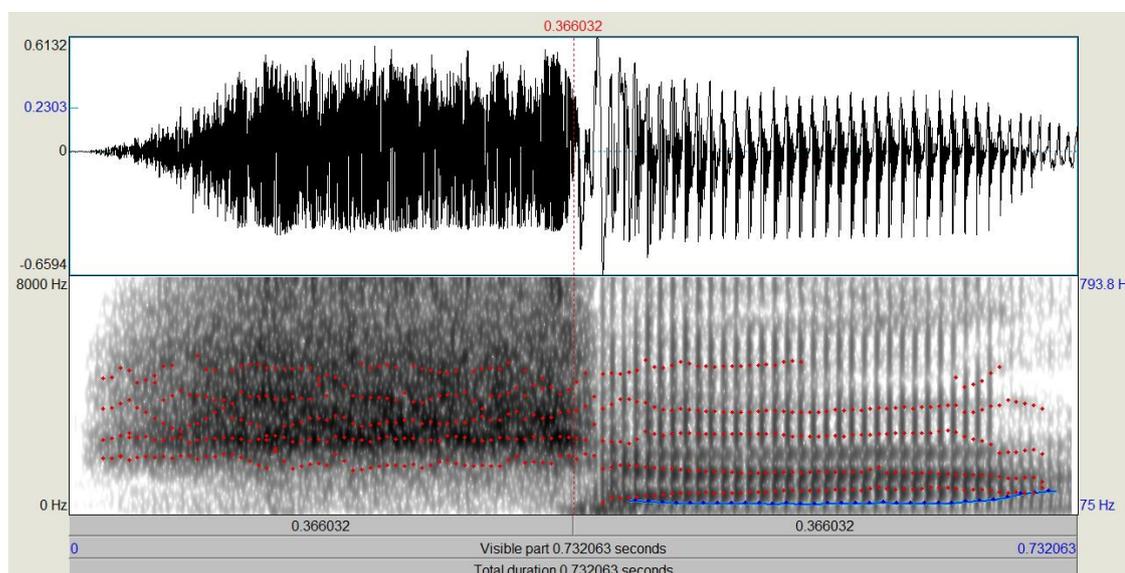


Figura 1: Espectrograma e oscilograma da sílaba /sa/, produzida por um falante pessoense do sexo masculino, com pico espectral de 2,5 KHz para a fricativa /s/.
Fonte: Pesquisa direta, 2015.

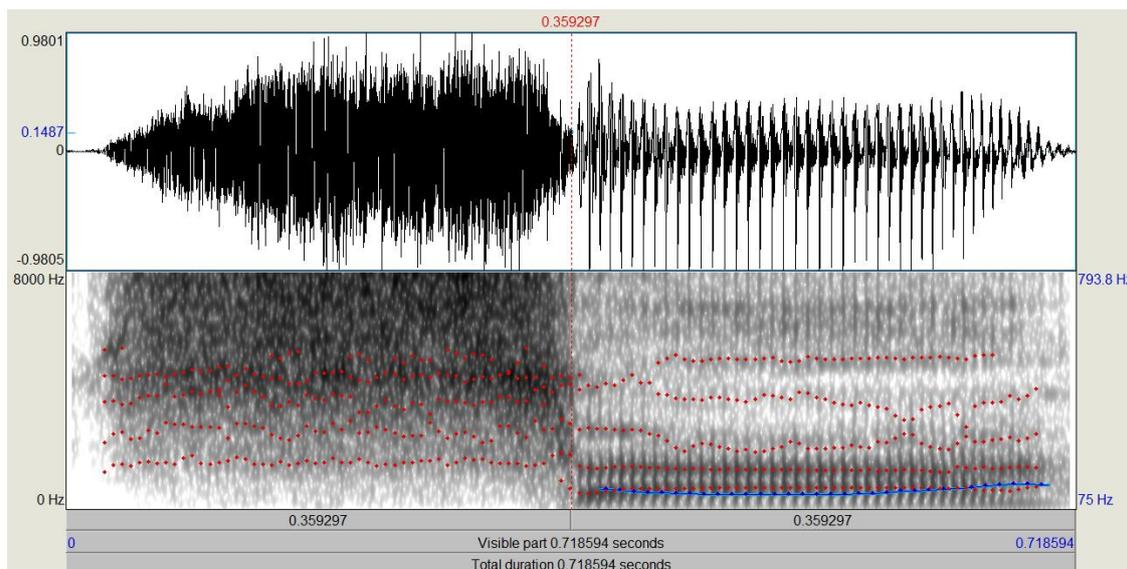


Figura 2: Espectrograma e oscilograma da sílaba /ʃa/, produzida por um falante pessoense do sexo masculino, com pico espectral de 2,5 KHz para a fricativa /ʃ/.
Fonte: Pesquisa direta, 2015.

Observando a transição formântica entre os segmentos fricativos e a vogal a eles adjacente pode-se fazer algumas considerações seguindo os apontamentos de Kent e Read (2015). Na figura 1, o limite de frequência mais baixa, da maior região de ruído primária para [ʃ], está perto da frequência de F3 para a vogal. Na figura 2, o limite de frequência mais baixa, da maior região de ruído, está perto da frequência de F4 para a vogal. Segundo os autores, como cada fricativa ocorre em uma sílaba CV, é conveniente comparar a região de ruído da fricativa com o padrão formântico da vogal.

Citando estudos como os de Harris (1958), que utilizou estímulos editados em que diferentes pistas estavam disponíveis e percebeu que o segmento de ruído para [s] não foi uma prova tão efetiva no contexto de [i] quanto foi nos contextos de [a] ou [u], Kent e Read (2015, p.271) apontam que, apesar das estridentes poderem ser bem identificadas apenas com seus segmentos de ruídos, as transições formânticas podem exercer um papel secundário na melhoria do reconhecimento das fricativas.

Metodologia

Da produção dos estímulos

Os estímulos para o teste de percepção foram gravados por dois falantes nativos de João Pessoa, um do sexo masculino e outro do sexo feminino, no Laboratório de Variação Linguística da Paraíba (VAL-PB), ambiente com ruído inferior a 50 dB. A coleta foi realizada com o microfone de mesa pedestal, marca Multilaser, modelo AC00040ML, com tecnologia antirruído, sensibilidade 38dBV/Pa, faixa de frequências de 50 - 16kHz (+/- 3db) e taxa de amostragem de até 48 kHz, acoplado diretamente, via cabo USB, a um *notebook* HP Pavilion 14-n010br, com placa de som integrada. A gravação foi feita por meio *software* PRAAT 4.1.44 (BOERSMA e WEENNINK, 2012), numa taxa de amostragem de 44.100 KHz.

Ambos os falantes foram treinados para produzir um gradiente de fricativas estridentes com constrictões feitas com o ápice da língua e o articulador fixo, seguidas de uma vogal central baixa [a]. Essa constrictão começava no alvéolo, na primeira produção, e terminava no palato, na última produção. A primeira gerou um efeito auditivo prototipicamente associado a um /s/, e a segunda, um efeito auditivo geralmente associado a um /ʃ/. Como foi discutido na seção anterior, algumas pistas acústicas parecem ser importantes para a discriminação desses dois segmentos, dentre elas a duração da consoante, o nível de energia em determinadas regiões de frequência e o pico espectral em determinadas regiões de frequência e a transição formântica das estridentes para a vogal.

Como a maioria dos estudos feitos até agora (JHONSON e MULENNIX, 1997; GIBSON, 1991[1977]; EDWARDS, 1999; BOERSMA, 2011; KENT e READ, 2015; JONGMAN et al., 2000; MANRIQUE e MASSONE, 1981) parece convergir para a efetividade das duas últimas pistas, essas foram as variáveis escolhidas para a análise aqui proposta. É importante destacar, entretanto, a importância de outros experimentos que controlem as outras pistas, dado que nenhum experimento de formalização de pista para as fricativas estridentes foi realizado para o português brasileiro e os dados que apresentamos levam em conta a realidade de discriminação de outras línguas. Desta forma, há a possibilidade de outras pistas serem relevantes para o português além das que aqui serão

analisadas. As variáveis “duração” e “loudness” foram controladas com a edição dos sons no programa SoundForge 10.0⁴³.

Após cada produção do gradiente pelos informantes, o segmento era analisado acusticamente no PRAAT para averiguar a frequência em que ocorria o primeiro pico espectral. Foram selecionados os segmentos que apresentavam distância entre os picos entre 0,5 e 1 KHz.

Das gravações das fricativas seguidas da vogal, foram recortadas apenas as fricativas. Desta forma, tivemos dois grupos de estímulo. O primeiro composto por sílabas e o segundo, de segmentos. Esse procedimento foi adotado porque também é nosso objetivo observar se a transição formântica da consoante para a vogal contribui para a acurácia na identificação do segmento como /s/ ou /ʃ/.

Do teste de percepção

Como foi exposto na seção anterior, haverá dois grupos de estímulos: o composto pelas consoantes seguidas de vogal produzidas pelos informantes, e o composto pelas consoantes apenas, recortadas dos estímulos originais. Os estímulos de cada grupo estão dispostos segundo as variáveis controladas, como é apresentado no quadro abaixo:

Quadro 1 – Estímulos

	Sexo do Inf.	Frequência do primeiro pico	Tipo de estímulo
Estímulo 1	Feminino	3.564 Hz	Apenas a fricativa
Estímulo 2	Feminino	4.000 Hz	Apenas a fricativa
Estímulo 3	Feminino	5.350 Hz	Apenas a fricativa
Estímulo 4	Feminino	5.565 Hz	Apenas a fricativa

⁴³ A edição dos estímulos para o controle da duração ocorreu através a operação cortar, a partir da qual normalizamos a duração de todas as fricativas para aproximadamente 0,35s. Quanto ao *loudness*, este foi controlado através da seleção do segmento e a aplicação da função "normalize - 6dB".

Estímulo 5	Feminino	6.023 Hz	Apenas a fricativa
Estímulo 6	Feminino	7.226 Hz	Apenas a fricativa
Estímulo 7	Feminino	8.002 Hz	Apenas a fricativa
Estímulo 8	Feminino	9.125 Hz	Apenas a fricativa
Estímulo 9	Masculino	2.585Hz	Apenas a fricativa
Estímulo 10	Masculino	3.238 Hz	Apenas a fricativa
Estímulo 11	Masculino	3.564 Hz	Apenas a fricativa
Estímulo 12	Masculino	4.084 Hz	Apenas a fricativa
Estímulo 13	Masculino	4.522 Hz	Apenas a fricativa
Estímulo 14	Masculino	5.193 Hz	Apenas a fricativa
Estímulo 15	Masculino	6.517 Hz	Apenas a fricativa
Estímulo 16	Masculino	8.466 Hz	Apenas a fricativa
Estímulo 17	Feminino	3.564 Hz	Fricativa + vogal
Estímulo 18	Feminino	4.000 Hz	Fricativa + vogal
Estímulo 19	Feminino	5.350 Hz	Fricativa + vogal
Estímulo 20	Feminino	5.565 Hz	Fricativa + vogal
Estímulo 21	Feminino	6.023 Hz	Fricativa + vogal
Estímulo 22	Feminino	7.226 Hz	Fricativa + vogal
Estímulo 23	Feminino	8.002 Hz	Fricativa + vogal
Estímulo 24	Feminino	9.125 Hz	Fricativa + vogal
Estímulo 25	Masculino	2.585Hz	Fricativa + vogal

Estímulo 26	Masculino	3.238 Hz	Fricativa + vogal
Estímulo 27	Masculino	3.564 Hz	Fricativa + vogal
Estímulo 28	Masculino	4.084 Hz	Fricativa + vogal
Estímulo 29	Masculino	4.522 Hz	Fricativa + vogal
Estímulo 30	Masculino	5.193 Hz	Fricativa + vogal
Estímulo 31	Masculino	6.517 Hz	Fricativa + vogal
Estímulo 32	Masculino	8.466 Hz	Fricativa + vogal

Depois de etiquetados e normalizados quanto ao *loudness* e duração das fricativas, os segmentos foram fornecidos como *inputs* para o teste de percepção. Ele foi elaborado e executado a partir *software* TP. O experimento foi dividido em duas partes: a primeira para avaliar a percepção das fricativas seguidas da vogal e a segunda, para análise das fricativas isoladamente.

Na primeira parte do experimento, os ouvintes deveriam ler a instrução: “Que sílaba você escuta?” e logo após, escolher a opção correta. À direita, ficou disposta a sílaba “sa” e à esquerda, a “xa”. O segundo experimento seguiu o mesmo roteiro. Os ouvintes deveriam ler a orientação: “Que som você escuta?” e escolher a opção correta. As opções dadas foram “s” e “x”. Para avaliar a atenção dos ouvintes, no teste, todos os estímulos foram duplicados e aleatorizados pelo programa. Assim, contamos ao todo com sessenta e quatro estímulos para serem analisados pelos ouvintes. Desse modo os ouvintes que marcaram, de forma diferente, mais de 50% das vezes o mesmo estímulo, foram excluídos da amostra. Cada estímulo poderia ser repetido três vezes.

Ao final do experimento, o programa apresentava para cada participante o total de estímulos ouvidos ao longo do teste, o número de acertos e de erros, além do tempo gasto por ele ao longo de toda realização da atividade. Ao mesmo tempo, o programa gerava uma planilha para análise do teste com as seguintes informações: o estímulo apresentado, a resposta fornecida pelo informante, o resultado correto esperado, e o tempo, em segundos, gasto pelo participante para responder cada estímulo.

Resultados e discussão dos dados

Os resultados expostos abaixo foram contabilizados e organizados através dados fornecidos pelo programa no qual os testes de percepção foram aplicados, dentre eles, o número de acertos e erros e sua porcentagem, a descrição da opção tomada a partir de cada estímulo exposto e o tempo que o participante levou para decidir entre as opções oferecidas. Para esta análise, consideraremos apenas a porcentagem de acertos para os estímulos dados, e as respostas associadas a cada estímulo.

Quanto às porcentagens de acerto de cada resposta, elas estão descritas na tabela abaixo:

Tabela 1: Porcentagens de acertos de cada resposta

Alternativa de resposta	Porcentagem de acerto
s	64,39%
x	88,96%
sa	80,55%
xa	87,98%

A definição da resposta certa para cada estímulo foi determinada pelos pesquisadores a partir do que a literatura aponta acerca dos dados de produção e percepção sobre segmento (JONGMAN et al. 2000; MANRIQUE e MASSONE, 1981; HARRIS, 1995). Desta forma, estímulos com picos de frequência abaixo de 4,5 KHz foram associados as opções “x” e “xa”, enquanto que estímulos com picos acima de 5 KHz foram associados a “s” e “sa”. Feita essa ressalva, podemos analisar a tabela acima observando que o índice de acertos foi alto para a associação das frequências mais baixas à fricativa palatal, assim como para a sílaba por ela composta. Esse índice cai consideravelmente quanto a associação de determinados picos às fricativas alveolares, voltando a crescer, entretanto, quando essa fricativa está seguida de uma vogal, formando com ela uma sílaba. Esse dado é uma pista para a confirmação da hipótese de que a transição formântica pode ser um dado importante para a percepção de fricativas alveolares, principalmente as que são produzidas com pico em uma faixa de frequência um pouco inferior às prototipicamente a ela associadas.

Faz-se necessário, pois, a partir da constatação acima, verificar a que opção os ouvintes associaram os estímulos oferecidos, dispostos num gradiente de picos espectrais,

formados por segmentos e formados por sílabas, características que correspondem às pistas que estão sendo consideradas para a análise. Esses dados estão descritos na tabela abaixo:

Tabela 2: Porcentagens de associação dos estímulos fornecidos e as opções feitas pelos ouvintes em contexto sem vogal adjacente e com vogal adjacente.

Estímulos produzidos pelos informantes	Opção escolhida			
	s	x	sa	xa
Pico espectral em 2.585Hz	13,6%	86,4%	6,8%	93,2%
Pico espectral em 3.238 Hz	6,8%	93,2%	0,0%	100,0%
Pico espectral em 3.564 Hz	11,3%	88,7%	4,5%	95,5%
Pico espectral em 3.564 Hz	6,8%	93,2%	0,0%	100,0%
Pico espectral em 4.000 Hz	9,0%	91,0%	2,2%	97,8%
Pico espectral em 4.084 Hz	9,0%	91,0%	22,7%	77,3%
Pico espectral em 4.522 Hz	18,1%	81,9%	52,2%	47,8%
Pico espectral em 5.193 Hz	59,0%	41,0%	77,2%	22,8%
Pico espectral em 5.350 Hz	15,9%	84,1%	6,8%	93,2%
Pico espectral em 5.565 Hz	59,0%	41,0%	84,0%	16,0%
Pico espectral em 6.023 Hz	38,6%	61,4%	90,9%	9,1%
Pico espectral em 6.517 Hz	68,1%	31,9%	77,2%	22,8%
Pico espectral em 7.226 Hz	70,4%	29,6%	100,0%	0,0%
Pico espectral em 8.002 Hz	88,6%	11,4%	95,4%	4,6%
Pico espectral em 8.466 Hz	84,0%	16,0%	95,4%	4,6%
Pico espectral em 9.125 Hz	90,9%	9,1%	100,0%	0,0%

A partir de uma leitura atenta dos dados expostos acima, podemos fazer algumas observações importantes. Percebe-se que, até a faixa do 4,5 KHz, os ouvintes atribuem o estímulo produzido à fricativa palatal, tanto individual, quando acompanhada de vogal. Essa realidade muda a partir dessa faixa de frequência até a faixa dos 6,0 KHz, em que há confusão na escolha entre a fricativa alveolar e a palatal, tanto isoladas quanto acompanhadas de vogal. A partir dos 6,0 KHz, a associação com a fricativa alveolar, nas duas modalidades aqui consideradas, passa a ser mais produtiva.

Na faixa de frequência do intervalo em que a confusão na associação dos estímulos acontece (4,0 a 6,5 KHz), percebe-se que, apesar da maior parte dos estímulos formados por segmentos isolados serem associados à fricativa palatal, o mesmo não acontece quando os segmentos são acompanhados de vogal. Nesse contexto, a porcentagem de associação para a fricativa alveolar cresce consideravelmente. O estímulo constituído de uma fricativa isolada

com pico espectral em 4,52 KHz foi associado a uma fricativa alveolar em apenas 18,1% das vezes que apareceu no teste. Quando essa mesma fricativa com pico espectral em 4,52 KHz é seguida de uma vogal central no estímulo, sua associação a uma fricativa alveolar sobe para 52,2%. O mesmo acontece com os estímulos compostos por fricativas com picos espectrais nas frequências 5,19 KHz, 5,35 KHz, 5,56 KHz, e 6,02 KHz. Este último, quando isolado, era associado a uma fricativa alveolar em apenas 38,6% das vezes que apareceu, ao contrário do estímulo com fricativa com pico espectral na mesma frequência seguida de vogal, que foi associado a uma sílaba com fricativa alveolar em 90,9% das pessoas que apareceu.

Algumas assimetrias entre a associação feita pelos ouvintes dentro do intervalo que aqui chamamos de “confuso” podem ser observadas na tabela acima. Destacamos, como exemplo, a escolha do “s” em 59% das vezes que estímulos compostos por fricativas com pico espectral em 5,19 KHz aparecem, assim como a escolha pela mesma opção em apenas 15,9% das vezes que estímulos compostos por segmentos com pico em 5,35 KHz são apresentados aos ouvintes. O esperado seria que, quanto maiores as frequências de ressonância do ruído, mais se associaria o estímulo a uma fricativa alveolar, ou seja, se o estímulo com pico em 5,19 KHz foi mais associado a “s”, o mesmo deveria acontecer com o estímulo com pico em 5,35 KHz.

Uma hipótese para essa pequena assimetria pode estar vinculada ao sexo dos informantes da amostra, já que os estímulos apresentados acima são os produzidos por ambos (um do sexo feminino e outro do sexo masculino). Abaixo, seguem os estímulos produzidos por cada um dos informantes e as representações gráficas a eles atribuídas.

Tabela 3: Porcentagens de associação dos estímulos produzidos pela informante do sexo feminino e as opções feitas pelos ouvintes em contexto sem vogal adjacente e com vogal adjacente

Estímulos produzidos pelo informante do sexo feminino	Opção escolhida			
	s	x	sa	xa
Pico espectral em 3.564 Hz	6,8%	93,2%	0,0%	100,0%
Pico espectral em 4.000 Hz	9,0%	91,0%	2,2%	97,8%
Pico espectral em 5.350 Hz	15,9%	84,1%	6,8%	93,2%
Pico espectral em 5.565 Hz	59,0%	41,0%	84,0%	16,0%
Pico espectral em 6.023 Hz	38,6%	61,4%	90,9%	9,1%
Pico espectral em 7.226 Hz	70,4%	29,6%	100,0%	0,0%
Pico espectral em 8.002 Hz	88,6%	11,4%	95,4%	4,6%
Pico espectral em 9.125 Hz	90,9%	9,1%	100,0%	0,0%

Tabela 4: Porcentagens de associação dos estímulos produzidos pela informante do sexo masculino e as opções feitas pelos ouvintes em contexto sem vogal adjacente e com vogal adjacente

Estímulos produzidos pelo informante do sexo masculino	s	x	sa	xa
Pico espectral em 2.585Hz	13,6%	86,4%	6,8%	93,2%
Pico espectral em 3.238 Hz	6,8%	93,2%	0,0%	100,0%
Pico espectral em 3.564 Hz	11,3%	88,7%	4,5%	95,5%
Pico espectral em 4.084 Hz	9,0%	91,0%	22,7%	77,3%
Pico espectral em 4.522 Hz	18,1%	81,9%	52,2%	47,8%
Pico espectral em 5.193 Hz	59,0%	41,0%	77,2%	22,8%
Pico espectral em 6.517 Hz	68,1%	31,9%	77,2%	22,8%
Pico espectral em 8.466 Hz	84,0%	16,0%	95,4%	4,6%

Agora podemos analisar como acontece a categorização dos estímulos em cada um dos informantes. Para a informante do sexo feminino, percebe-se que a faixa de frequência limite para a percepção das fricativas palatais foi de 5,35 KHz. Abaixo disso, os estímulos foram mais associados à fricativa palatal. O intervalo entre 5,35 KHz e 7,23 KHz pareceu difuso para a categorização. Acima disso, os estímulos foram associados à fricativa alveolar. Pode-se notar, também, que os estímulos na faixa de frequência 5,56 e 6,02 KHz apresentaram uma diferença considerável no que diz respeito a sua categorização, enquanto segmento isolado e enquanto sílaba, principalmente no que diz respeito a última faixa. A transição formântica parece, pois, exercer um papel importante para a identificação das fricativas alveolares nessas faixas de frequência.

No que diz respeito aos estímulos produzidos pelo informante do sexo masculino, a partir da faixa de 5,19 KHz para cima, os ouvintes classificaram os sons isolados como a letra “s” e os seguidos de vogal, como a sílaba “sa”, com um índice de acerto maior quando o seguimento foi seguido de vogal. Da faixa de frequência de 4,08 KHz para baixo, os ouvintes classificaram significativamente o estímulo a uma fricativa palatal e a sílaba formada por ela e uma vogal. A faixa de frequência de 4,5 pareceu difusa para os ouvintes, que classificaram o estímulo isolado como uma fricativa palatal e, quando o estímulo era composto por fricativa e vogal, a classificação como fricativa alveolar cresceu significativamente.

Considerações finais

A partir dos dados discutidos na seção anterior, algumas considerações podem ser postuladas no que diz respeito às pistas acústicas para a distinção entre as fricativas alveolares de palatais por falantes de João Pessoa. Quanto ao pico espectral, percebeu-se que as fricativas que apresentaram o primeiro pico de energia nas regiões de frequência de 4,0 KHz para baixo, tanto isoladas quanto acompanhadas de vogal, foram associadas a fricativa palatal. Por outro lado, as fricativas alveolares foram associadas a estímulos com picos espectrais iguais ou maiores que 6,5 KHz, tanto individuais quanto seguidos de vogal. O intervalo entre essas faixas de frequência parece confundir os ouvintes, e a pista da transição vocálica parece atuar nesse momento como pista relevante para a identificação das fricativas alveolares.

Quanto à análise dos informantes individualmente, os dados foram bastante parecidos. Entretanto, para o informante masculino, o intervalo de confusão para os ouvintes diminuiu para 1KHz (4,08 a 5,19 KHz) em relação ao intervalo para a informante do sexo feminino, que era de aproximadamente 2 KHz (5,35 a 7,23 KHz). Uma análise envolvendo identificação de estímulos produzidos por mais informantes de ambos os sexos em testes de percepção seria necessário para se chegar a conclusões mais precisas e confiáveis a esse respeito.

Ademais, testes de percepção como os que aqui foram realizados estão se tornando cada vez mais importantes para o entendimento das pistas acústicas que são utilizadas para os ouvintes para acessar categorias fonológicas discretas e como essas pistas podem interagir entre si, principalmente para perspectivas teóricas que integram fonética e fonologia através do mapeamento do continuum fonético e a formalização de restrições de pista acústica, como o modelo de processamento de fala bidirecional do Boersma (2011). Apesar de sua importância, poucos são os trabalhos desenvolvidos com o Português Brasileiro.

Outra vantagem desse tipo de trabalho é da relação entre um gesto articulatório e um correlato acústico relevante para a interpretação do sinal da fala, informação relevante, por exemplo, para a clínica fonoaudiológica e o trabalho com a identificação de desvios articulatórios e possíveis ajustes a serem realizados de forma a se atingir o padrão prototípico da língua alvo.

Referências

BOERSMA, P. *A programme for bidirectional phonology and phonetics and their acquisition and evolution*. In: Anton BENZ, A.; MATTAUSCH, J. (eds), *Bidirectional Optimality Theory*. Amsterdam: John Benjamins, p. 33-72. 2011.

_____; WEENINK, D. *PRAAT: doing phonetics by computer* [computer program], 2012. Disponível em <<http://www.praat.org/>>.

BRASIL, B. C.; MEZZOMO, C. L.; MOTA, H. B.; MELO, R. M.; LOVATTO, L.; ARZENO, L. Características acústicas dos fones [s] e [S] de adultos e crianças com desenvolvimento fonológico típico. *Revista da Sociedade Brasileira de Fonoaudiologia*. v.17, n.2, p.182-188, 2012.

EDWARDS, J. Refining our understanding of language attitudes. *Journal of Language and Social Psychology*, v.18, n.1, p.101-110, 1999.

FOWLER, C. A. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, v. 14, pp. 3–28. 1986.

GIBSON, E. J. How perception really develops: A view from outside the network, in *An Odyssey in Learning and Perception*. MIT Press, Cambridge, MA!, pp. 411–491. 1991.

HAUPT, Carine. As fricativas [s], [z], [ʃ] e [ʒ] do português brasileiro. *Revista Estudos Linguísticos XXXVI*(1), janeiro-abril, pp. 37-46, 2007.

HALE, M.; REISS, C. Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry*, v. 29, pp. 656–683. (1998).

HARRIS, K. Cue for discrimination of American English fricatives in spoken syllables. *Language and Speech*, n. 1, pp. 1-17, 1958.

HAYES, Bruce. Phonetically-driven phonology: the role of Optimality Theory and inductive grounding. In: DARNELL, M. et al. (Eds.), *Functionalism and Formalism in Linguistics*, Volume I: General Papers, Amsterdam: John Benjamins. pp. 243-285. 1999.

JESUS, Luís Miguel Teixeira. *Analysis of Portuguese Fricative Consonants*. Mini Thesis- Department of Electronics and Computer Science, University of Southampton. 70f. 1999.

JONGMAN, A.; WAYLAND, R.; WONG, S. Acoustics Characteristics of English Fricatives: I. Static Cues. *Working Papers of the Cornell Phonetics Laboratory*, v. 12, pp. 195-205, 1998.

JOHNSON, Keith; MULLENIX, John W. (Ed.) *Talker variability without in speech perception*. San Diego: Academic Press, p.1-8, 1997.

KENT, R. D.; READ, C. *Análise Acústica da Fala*. Tradução de Alexandro Rodrigues Meireles. 1 ed. São Paulo: Cortez, 2015.

LACERDA, Francisco P. Acoustic perceptual study of the Portuguese voiceless fricatives. *Journal of Phonetics*. London, v. 10, p. 11-22. 1982.

LOPES, L. W. *Preferências e atitudes dos ouvintes em relação ao sotaque regional no telejornalismo*. Tese (Doutorado) – Universidade Federal da Paraíba, 2012.

MANRIQUE, A. M. B.; MASSONE, M. I. Acoustic analysis and perception of Spanish fricative consonants. *Journal of the Acoustical Society of America*, n. 69, p. 1145-1153, 1981.

NITTROUER, S. *Learning to perceive speech: How fricative perception changes, and how it stays the same*. *Journal of Acoustic Society of America*. v.112, August 2002

RAUBER, A. et al. *TP: perception tests and perceptual training with immediate feedback*, versão 3.1. 2012. Disponível em: <[http://www.worken.com.br/tp_regfree .php](http://www.worken.com.br/tp_regfree.php)>, último acesso em 28 fev. 2014.

WIETHAN, F. M.; MOTA, H. B. A generalização a partir do tratamento com fricativas: ambientes favoráveis *versus* ambientes pouco favoráveis e neutros. *Revista CEFAC*. v.16, n.3, p. 1021-1029, 2014.

_____; CERON, M. I.; MARCHETTI, P.; GIACCHINI, V.; MOTA. H. B.

O uso da eletroglotografia, eletromiografia, espectrografia e ultrassom nos estudos de fala – revisão teórica. *Revista CEFAC*. v.17, n.Supl1, p. 115-125, 2015.