

DESCRIÇÃO FONÉTICO-ACÚSTICA DAS FRICATIVAS NO PORTUGUÊS BRASILEIRO: CRITÉRIOS PARA COLETA DE DADOS E PRIMEIRAS MEDIDAS ACÚSTICAS

Ingrid SAMCZUK (Pontifícia Universidade Católica de São Paulo)
ingridbs@uol.com.br

Aglael GAMA ROSSI (Pontifícia Universidade Católica de São Paulo)
gamarossi@uol.com.br

ABSTRACT: The paper presents the criteria adopted to compose a corpus and to collect data to an acoustic study of Brazilian Portuguese fricatives (/f-v, s-z, ʃ-ʒ, x-ɣ or h-ɦ, the last two pairs been dependent of the subject's dialect). Preliminary results of acoustic measures, mainly frequency and duration, are also presented.

KEYWORDS: Brazilian Portuguese fricatives, acoustic analysis, method design.

0. Introdução

O objetivo deste trabalho é descrever as decisões metodológicas de um projeto de Iniciação Científica (PIBIC-CEPE) que visa realizar a descrição fonético-acústica dos sons fricativos no português brasileiro (PB), uma vez que a mesma ainda não existe.

As fricativas que compõem o sistema fonológico do PB distinguem-se quanto aos seguintes pontos de articulação: lábio-dentais (/f v/), alveolares (/s z/), e palatais (/ʃ ʒ/). Fricativas velares (/x ɣ/) e glotais (/h ɦ/) têm sido utilizadas como formas variantes dos *erres* dialetais. Em cada um desses pontos de articulação, ocorrem os pares vozeado e não-vozeado.

De acordo com Kent e Read (1992:121), o principal traço articulatório de uma fricativa é a formação de uma constrição estreita num dado ponto do trato vocal. Quando o ar passa através dessa constrição, numa quantidade de fluxo suficiente, há a formação de uma turbulência, o que significa que o movimento da partícula da corrente de ar torna-se altamente complexo, formando pequenos redemoinhos na região constrita. A condição aerodinâmica de turbulência está associada à geração de um ruído turbulento no sinal acústico. As fricativas, portanto, são identificadas (1) pela formação de uma constrição estreita no trato vocal, (2) pelo desenvolvimento de fluxo de ar turbulento, e (3) pela geração de um ruído de turbulência. Comparadas com as demais classes de sons, elas possuem duração relativamente longa e é esse intervalo de energia aperiódica que as diferencia enquanto uma classe.

1. Coleta de dados

Para o estudo dessa classe de sons, elaborou-se um *corpus* de palavras e logatomas (palavras sem sentido, passíveis de serem pronunciadas) no qual as

fricativas ocorrem no contexto: sílaba tônica de dissílabos paroxítonos, com estrutura silábica ${}^1C_1V_1.C_2V_2$, sendo V_1 qualquer uma das sete vogais orais (/a e e i o u/) ou das cinco vogais nasais /aŃ eŃ iŃ oŃ uŃ/) do PB; C_2 uma consoante dissimilatória, ou seja, que difere da fricativa-alvo (C_1) em ponto e modo de articulação (C_2 é sempre uma oclusiva), mas não quanto ao vozeamento; e V_2 sempre /a/, uma vez que essa é a vogal de maior frequência de ocorrência na pós-tônica de substantivos dissílabos paroxítonos CV.CV (Gama Rossi e Madureira, 2002).

O critério da dissimilação de pontos e modos de articulação entre a fricativa-alvo e a consoante da pós-tônica foi adotado por dois motivos: em primeiro lugar, o PB pretere consoantes de mesmo ponto e modo de articulação em sílabas contíguas, conforme demonstrado por Albano (2001); em segundo lugar, caso fosse fixada uma única consoante oclusiva na pós-tônica para todo o *corpus*, ela assumiria distâncias diferentes de ponto de articulação em relação a cada fricativa. Por exemplo, se tivesse sido fixada a oclusiva alveolar não-vozeada (/t/) na posição pós-tônica, por ser anterior, ela assumiria diferentes relações com as fricativas no eixo ântero-posterior do trato vocal: lábio-dental (*fata*), alveolar (*safa*), palatal (*chata*) e velar/glotal (*rata*). Resumindo, o motivo por trás da adoção do critério da dissimilação refere-se à influência da articulação da consoante pós-tônica sobre a articulação da fricativa-alvo, uma vez que, na fala, ocorre sobreposição entre sons, conhecida como coarticulação (grau de acomodação entre segmentos fonéticos adjacentes, o que reflete o fato de que, no planejamento e na execução neuromuscular da fala, há uma tendência para que a realização de um segmento antecipe as características articulatórias de um ou mais segmentos fonéticos adjacentes, de acordo com Laver, 1994: 151).

Adotaram-se, então, as seguintes relações dissimilatórias entre a fricativa-alvo e a consoante pós-tônica: fricativa lábio-dental e oclusiva velar (/f/-/k/ e /v/-/g/); fricativa alveolar e oclusiva velar (/s/-/k/ e /z/-/g/); fricativa palatal e oclusiva bilabial (/ʃ/-/p/ e /ʒ/-/b/). No caso das fricativas velares e glotais, por não se saber previamente, na composição do *corpus*, qual variante seria empregada pelo sujeito, essas possíveis variantes foram combinadas com as oclusivas bilabiais vozeada e não-vozeada, respectivamente (/x/-/p/ e /x/-/b/).

Como forma de controlar a curva entoacional de cada palavra ou logatoma, dissipando a curva ascendente característica que eles adquiririam caso fossem repetidos isoladamente em forma de lista, os mesmos foram inseridos na frase-veículo “Diga _____ pra ele”, da qual foram posteriormente recortados para estudo.

O sujeito da pesquisa é do sexo feminino, nascido na cidade de São Paulo e com 41 anos de idade na data da gravação. Para esta, foram elaboradas fichas que continham a frase-veículo preenchida com o logatoma ou palavra, num total de 96 fichas [8 fricativas x 12 (7 vogais orais + 5 vogais nasais)]. As fichas foram embaralhadas antes de cada repetição com vistas a atenuar o efeito de aprendizagem ou memorização caso fosse mantida a ordem fixa de apresentação dos estímulos. Foram feitas 6 repetições do conjunto de 96 palavras ou logatomas, isto porque padrões acústicos dos sons da fala são apreendidos com base na análise estatística, daí a necessidade de uma grande quantidade de dados.

A gravação foi realizada no Estúdio de Rádio e TV da PUC-SP, monitorada pelo técnico Ernesto Foschi, diretamente no computador: Pentium III; com 2 HD de 20Gb cada; 256Mb de memória RAM e processador de 700MHz. Foram utilizados ainda: microfone Sennheiser e835; mesa Shure M267; placa de áudio Delta 44 e o software Sound Forge 6.0. Os dados foram capturados em uma taxa de amostragem de 22.050Hz. Da placa de som do computador que gerencia o estúdio, o *corpus* foi gravado diretamente em CD-ROM, de modo que cada repetição correspondesse a uma faixa.

Cada frase-veículo, incluindo as 6 repetições, foi dividida num arquivo digital contendo do estouro do /g/ pós-tônico de “diga” até o estouro do /p/ de “pra”. O programa de análise acústica que vem sendo utilizado é o Praat, obtido livremente no *site* www.praat.org.

2. Critérios para as medidas acústicas

2.1. Medidas de frequência

De acordo com Ladefoged e Maddieson (1996:139), a estrutura acústica das fricativas varia amplamente de um indivíduo a outro, o que reflete o fato de ainda não haver procedimentos acústicos padronizados para a descrição das mesmas. Não se sabe como apreender o que é constante e o que é lingüística e perceptualmente mais importante em termos acústicos.

O primeiro problema encontrado refere-se a que tipo de medidas de frequência realizar e com base em qual método. Para Kent e Read (1992:70-2), a escolha depende do tipo de amostra em estudo: a análise de Fourier, FFT, (*Fast Fourier Transform*) assume a existência de uma estrutura periódica ou harmônica, como aquela encontrada em vogais e soantes, nas quais se observa, nos espectrogramas, linhas verticais que correspondem aos ciclos de abertura e fechamento das pregas vocais. Já a análise de LPC (*Linear Prediction* ou *Linear Predictive Coding*) não assume tal estrutura e baseia-se no fato de que qualquer ponto de uma amostra de fala digitalizada é parcialmente predito a partir do que imediatamente o precedeu, uma vez que a fala não varia amplamente de um ponto a outro. Portanto, a predição linear refere-se à hipótese de que qualquer ponto é uma função linear daquilo que o precede. Uma vez que as fricativas constituem sinais aperiódicos, decidiu-se por empregar o método de LPC na extração de seus componentes de frequência, mesmo no caso das fricativas vozeadas, formadas pela sobreposição entre o ruído (sinal aperiódico) gerado na cavidade oral a partir de uma obstrução parcial da mesma e a vibração das pregas vocais (sinal periódico).

De acordo com Ladefoged e Maddieson (1996:139), uma vez que não há um modelo adequado para a acústica das fricativas, aposta-se que as medidas importantes para caracterizá-las sejam: a intensidade geral¹, a frequência do ponto mais baixo do espectro, e algo que corresponda ao centro de gravidade² e à dispersão dos componentes espectrais acima de um certo limiar. No presente trabalho, a medida das frequências das fricativas começou por localizar o ponto médio delas (a partir de sua duração total), estabelecendo-se uma janela de 50ms para cada um dos lados desse ponto, totalizando 100ms (Nittrouer, Studdert-

Kennedy, McGowan, 1989). Conforme será visto na tabela 3, a duração média das fricativas não-vozeadas é em torno de 200ms, lembrando que as mesmas estão em posição de acento. Dentro da janela de 100ms, foram extraídos quinze valores para cada um dos quatro primeiros formantes da fricativa, totalizando 90 ocorrências (6 repetições x 15 valores) de cada um deles. Os valores das frequências foram extraídos automaticamente pelo *Praat* a partir do método de LPC, com base nos seguintes parâmetros: *New sample rate* (Hz): 11000; *Precision* (samples): 50; *Prediction order*: 10; *Analysis width* (seconds): 0,025; *Time step* (seconds): 0,010; *Pre-emphasis from* (Hz): 50. A realização da análise de LPC com base em tais parâmetros teve como objetivo captar as frequências mais baixas dos ruídos fricativos. A extração dos valores das frequências dos mesmos, a partir de seus pontos centrais, numa janela de 100 ms, teve como objetivo extrair valores na parte mais estável dos ruídos, ou menos suscetível aos efeitos coarticulatórios das vogais pós-tônica (/a/) precedente ([dʒigɐ]) e da tônica seguinte (/a, ɛ, e, i, ɔ, o, u/). Serão ainda extraídas as frequências de uma janela a ser delimitada no final de cada ruído fricativo, para verificar se há efeito da vogal tônica sobre eles, ou seja, se há efeitos de coarticulação antecipatória (Nittrouer *et al.*, *op. cit.*).

Numa primeira tentativa de diferenciação entre o ponto de articulação das fricativas não-vozeadas, a partir das frequências extraídas no centro das mesmas, elas foram comparadas entre si apenas antes de /a i u/, dentro dos logatomas/palavras *faca, fica, fuca, saca, sica, suca, chapa, chipa, chupa*, e não para as sete vogais orais do PB conforme previsto para a versão final do trabalho. A observação³ dos valores mínimo, máximo, de média e desvio-padrão das frequências mostrou uma grande variação, o que leva à questão do erro gerado na extração automática das mesmas pelo método de LPC. Isso cria a necessidade de revisão da tabela com os valores gerados para a retirada dos que são espúrios e posterior realização da análise estatística. Como exemplo da variação das medidas, tem-se para /ʃ/ de *chapa*, com 90 ocorrências de cada frequência (6 repetições x 15 valores da mesma), a média de F1 em 1587Hz, com desvio-padrão de 570Hz, sendo que o valor mínimo foi 804Hz e o máximo 3024Hz. Por outro lado, para /f/ de *fuca*, a média encontrada para F1 foi 1380Hz, com desvio-padrão de 88Hz, e valores mínimo e máximo, respectivamente, de 1130Hz e 1668Hz.

Seguindo Funatsu (1995, *apud* Jesus e Shadle, 2002:439), verificou-se se as frequências dos quatro primeiros formantes, medidas no início da transição do segundo formante da vogal tônica, ajudariam a caracterizar as fricativas quanto ao ponto de articulação. Tal medida foi realizada para as palavras/logatomas descritos acima. A tabela 1 mostra os valores mínimo, máximo, de média e desvio-padrão de /a i u/ antes de /f s ʃ/.

Tabela 1: Valores mínimo, máximo, média e desvio-padrão das quatro primeiras frequências formânticas medidas no início de F2 das vogais tônicas /a i u/ depois de /f s ʃ/

		Valor Mínimo	Valor Máximo	Média	Desvio-Padrão
/a/ depois de /f/	F1	826	973	900	50
	F2	1324	1406	1370	34
	F3	2591	2746	2660	58
	F4	4199	4420	4303	91
/a/ depois de /s/	F1	664	826	744	56
	F2	1405	1526	1461	44
	F3	2701	2807	2773	40
	F4	4344	4433	4388	34
/a/ depois de /ʃ/	F1	633	742	704	44
	F2	1379	1494	1461	44
	F3	2714	2987	2906	100
	F4	4172	4504	4345	128
/i/ depois de /f/	F1	304	311	308	2
	F2	2500	2558	2531	25
	F3	3008	3095	3063	39
	F4	4012	4230	4163	81
/i/ depois de /s/	F1	303	319	309	6
	F2	2387	2510	2468	55
	F3	2812	2963	2906	59
	F4	4093	4260	4191	71
/i/ depois de /ʃ/	F1	293	307	303	5
	F2	2415	2508	2475	37
	F3	2881	3019	2945	55
	F4	2866	4249	4062	140
/u/ depois de /f/	F1	346	427	381	28
	F2	730	983	882	85
	F3	2674	2929	2836	92
	F4	3927	4079	4003	53
/u/ depois de /s/	F1	313	334	324	8
	F2	1070	1286	1157	75
	F3	2884	3001	2930	47
	F4	4049	4138	4090	39
/u/ depois de /ʃ/	F1	331	356	344	11
	F2	1071	1843	1417	283
	F3	2851	3053	2939	80
	F4	3234	4120	3711	316

Tabela 2: Valores de p para comparações entre as quatro primeiras frequências formânticas medidas no início de F2 das vogais tônicas /a i u/ depois de /f s ʃ/

	F1	F2	F3	F4
faca - saca	0,000	0,972	0,388	0,996
saca - chapa	0,738	1	0,192	1
chapa - faca	0,000	0,974	0,000	1
fica - sica	1	0,997	0,058	1
sica - chipa	1	1	0,998	0,942
chipa - fica	1	0,999	0,328	0,987
fuca - suca	0,273	0,026	0,647	0,995
suca - chupa	0,996	0,044	1	0,009
chupa - fuca	0,904	0,000	0,531	0,105

A observação da tabela 1 mostra uma diferença entre as médias de F1 para a vogal tônica /a/ que se segue a /f/, de um lado, e que se segue a /s ʃ/, de outro. De acordo com Kent e Read (1992:23), o valor de F1 está inversamente relacionado à altura da língua: vogais altas têm F1 baixo. Sendo /a/ uma vogal central e baixa, e a média de F1 correspondente a 900Hz, 744Hz e 704Hz depois de /f s ʃ/, respectivamente, tem-se que a produção das fricativas /s ʃ/ faz com que a língua esteja numa posição mais elevada quando do início da produção da vogal. A tabela 2 confirma que há diferença significativa entre os valores de F1 tomados no início de F2 da vogal /a/ advinda de /f/ em oposição às que vêm de /s ʃ/. Esses resultados indicam que há coarticulação perseveratória, uma vez que efeitos dos pontos de articulação das fricativas /s ʃ/ ainda estão presentes no início da vogal tônica que a elas se segue. Note que a articulação da fricativa lábio-dental /f/ não interfere em articulações que requeiram a participação da língua, como é o caso das fricativas alveolar e palatal /s ʃ/.

No que se refere à vogal /u/ advinda das três fricativas, foram encontradas diferenças significativas para F2. Segundo Kent e Read (*op.cit.*), F2 relaciona-se ao eixo ântero-posterior do trato vocal, sendo que o avanço da língua resulta num aumento de seu valor. Como mostra a tabela 1, há um aumento de F2 de /u/ depois de /s ʃ/, mas não depois de /f/, o que sugere que os pontos de articulação das fricativas alveolar e palatal trazem para frente o ponto de articulação da vogal posterior /u/. Foram encontradas, conforme a tabela 2, diferenças significativas para F2 da vogal /u/ advinda de cada uma das três fricativas. Note ainda que o alto valor do desvio-padrão de seu F2 depois de /ʃ/ (283Hz) em oposição àqueles obtidos para /f s/ (85Hz e 75Hz, respectivamente) parece referir-se à variação inerente ao ponto de articulação da fricativa palatal, uma vez que nela uma área mais ampla da língua (suas bordas) faz contato com o palato, ao contrário do que ocorre nas fricativas lábio-dental (incisivos centrais superiores em contato com lábio inferior) e alveolar (ponta da língua em contato com os alvéolos dos incisivos centrais inferiores).

Já no que concerne à vogal /i/, não foram encontradas influências dos pontos de articulação das fricativas sobre a mesma.

A medida feita no início da transição de F2 da vogal tônica parece, de acordo com os dados das tabelas 1 e 2, ter servido para caracterizar as fricativas no que diz respeito à influência de seus pontos de articulação (lábio-dental, alveolar e palatal) sobre as vogais /a u/, o que vai de encontro ao que foi proposto por Funatsu (1995, *apud* Jesus e Shadle, 2002:439) na comparação entre japonês e russo. Portanto, essa mesma medida deve ser complementada para as demais vogais orais seguintes às fricativas não-vozeadas e também para todas as vogais orais seguintes às fricativas vozeadas.

2.2. Duração

Para realizar as medidas de duração, primeiramente a seqüência vogal pós-tônica – fricativa – vogal tônica foi recortada do arquivo de som contendo cada uma das frases-veículo, sendo o arquivo delimitado do estouro do /g/ (de “Diga”) até a metade do tempo de fechamento da consoante oclusiva da pós-tônica pertencente à palavra/logatoma. A partir daí, a duração de cada um dos três segmentos da seqüência foi delimitada segundo os critérios: primeiro pulso regular da vogal pós-tônica até o último pulso regular da mesma; último pulso regular da vogal pós-tônica até o primeiro pulso regular da vogal tônica; primeiro pulso regular da vogal tônica até o último pulso regular da mesma.

Foi feito o cálculo da média e desvio-padrão da duração das fricativas não-vozeadas com base em 42 ocorrências de cada (6 repetições da fricativa antes de cada uma das 7 vogais orais). Os dados analisados até o momento, tanto em valores absolutos quanto em *Z-score* (duração da fricativa menos a média de todas as suas ocorrências, dividido pelo desvio-padrão)⁴, mostram que não há diferença entre elas, o que pode ser observado na tabela 3, abaixo.

Tabela 3 – Média e desvio-padrão para a duração das 42 ocorrências de cada fricativa não-vozeada (/f, s, ʃ/) em valores absolutos (ms) e em *Z-score*

	f		s		ʃ	
	Abs	Z	Abs	Z	Abs	Z
Média	201	0	222	0	220	0
Desvio-padrão	14	1	16	1	13	1

A comparação entre a duração das fricativas diante de cada vogal, a partir de uma análise estatística (ANOVA – teste de Scheffe, com nível de significância $p < 0,05$), mostrou poucas diferenças significativas.

3. Considerações finais

O objetivo desse artigo foi descrever os critérios adotados na elaboração de um *corpus* e na coleta de dados para o estudo fonético-acústico das fricativas do PB. Tendo apresentado tais critérios, foram mostrados os primeiros resultados

de algumas medidas acústicas realizadas com o objetivo de distinguir as fricativas. Uma medida que se mostrou promissora foi a das frequências formânticas da vogal tônica, extraídas no início da transição de seu segundo formante, conforme proposto por Funatsu (1995, *apud* Jesus e Shadle, 2002:439). Quando comparadas para suas durações, as fricativas não-vozeadas também não mostraram diferenças entre si. As análises aqui apresentadas não contemplam todas as fricativas e contextos vocálicos da pesquisa, nem todas as medidas acústicas que podem levar a uma caracterização mais completa das mesmas, as quais devem vir a ser feitas no sentido de complementar a caracterização das fricativas do PB.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALBANO, Eleonora Cavalcante. *O Gesto e suas Bordas. Esboço de Fonologia Acústico-Arteficial do Português Brasileiro*. Campinas: Mercado de Letras Edições e Livraria Ltda./Associação de Leitura do Brasil, 2001.
- GAMA-ROSSI, Aglael; MADUREIRA, Sandra. Julgamentos de aceitabilidade da fonotaxe do português brasileiro: primeira análise. *Estudos Linguísticos*, v.XXXI, Grupo de Estudos Linguísticos do Estado de São Paulo (GEL), publicação em CD-ROM, <file:///F:\htm\comunica\CiIII26a.htm>, 2002.
- JESUS, Luis; SHADLE, Christine. A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics*, 30: 437-464, 2002.
- KENT, Ray; READ, Charles. *The Acoustic Analysis of Speech*. San Diego: Singular Publishing Group Inc., 1992.
- LADEFOGED, Peter; MADDIESON, Ian. *The Sounds of the World's Languages*. Oxford: Blackwell Publishers Ltd., 1996.
- LADEFOGED, Peter. *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Oxford: Blackwell Publishers Ltd., 2003.
- LAVER, John. *Principles of Phonetics*. Cambridge: Cambridge University Press, 1994.
- NITTROUER, Susan; STUDDERT-KENNEDY, Michael; MCGOWAN, Richard. The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, vol.32: 120-132, March 1969.

NOTAS

¹ A intensidade geral de cada um dos ruídos fricativos não poderá ser medida uma vez que não foi controlada durante a gravação dos dados, já que ela é um parâmetro acústico que varia em função do volume de voz do locutor e da proximidade de sua boca em relação ao microfone, fixo sobre a mesa.

² Ladefoged (2003) coloca que o centro de gravidade refere-se ao primeiro momento espectral da forma definida pela curva espectral e suas fronteiras. Isso equivale ao ponto em que, se espetado em um alfinete, um pedaço de cartolina com a forma da curva espectral ficaria em equilíbrio. Para maiores detalhes sobre o cálculo do centro de gravidade, ver Ladefoged (*op. cit.*).

³ Agradecemos a Profa. Dra. Yara Castro, consultora em Estatística da PUC-SP, à análise estatística.

⁴ Os valores absolutos de duração foram transformados em valores de *Z-score* como forma de atenuar o fato de que a duração é um parâmetro sujeito a variação em função de: posição de acento, taxa de elocução ou número de sílabas pronunciadas por segundo, fronteira dentro das sentenças e entre sentenças.