

# A percepção auditiva e visual das fricativas do Português Brasileiro diante da manipulação do sinal acústico<sup>1</sup>

Audinéia Ferreira-Silva\*

Vera Pacheco\*\*

Luís Carlos Cagliari\*\*\*

## Resumo

Neste artigo, nosso objetivo é avaliar o papel das informações auditiva e visual na percepção das fricativas do no Português Brasileiro, quando apresentadas com ambiguidade / manipulação do sinal acústico. Para tanto, montamos um *corpus* de palavras dissílabas, com estrutura silábica  $C_1V_1C_2V_2$ , onde  $C_1$  é uma das fricativas,  $C_2$  é a oclusiva surda e  $V_1$  e  $V_2$  são uma das vogais /a/, /i/ ou /u/. Após a gravação do *corpus*, o sinal acústico das fricativas foi manipulado em termos de duração do ruído e frequência do espectro. Nossos resultados evidenciam que os índices de identificação das fricativas foram, de maneira geral, maiores quando elas apresentavam a informação audiovisual. Nossos achados atestam que, quando as fricativas têm seu sinal manipulado, em termos de duração e frequência, seu desempenho perceptual aumenta nos casos em que a informação visual é apresentada com a auditiva, ou seja, diante da manipulação do sinal, as fricativas apresentam médias de recuperação mais altas com a informação audiovisual do que com a informação, apenas, auditiva. Esses resultados corroboram os tradicionais estudos de Sumby e Pollack (1954) e McGurk e MacDonald (1976), que afirmam que, na percepção, os ouvintes integram visão e audição. Assim, e considerando os pressupostos do FLMP de Massaro (1987), nossos resultados evidenciam que, na percepção da fala, cada fonte de informação é mais influenciável, na medida em que a outra fonte é mais ambígua.

Palavras-chave: Percepção. Fricativas. Informação auditiva e visual

Recebido em: 30/01/2017

Aceito em: 23/03/2017

---

<sup>1</sup> Esse trabalho é resultado da tese de Doutorado do 1º autor, fomentado pelo CNPq, Processo nº. 142061/2012-3.

\* Universidade Estadual do Sudoeste da Bahia-UESB. Doutora em Linguística e Língua Portuguesa.

\*\* Universidade Estadual do Sudoeste da Bahia-UESB. Doutora em Linguística.

\*\*\* Universidade Estadual Paulista "Júlio de Mesquita Filho"- UNESP. Doutor em Linguística.

## 1 Introdução

A visão é um dos cinco sistemas sensoriais e um dos mais significativos para os seres humanos. Do ponto de vista biológico, a relevância desse sistema pode ser atestada pelo fato de que cerca de metade do córtex cerebral é destinada ao processamento visual (SCHIFFMAN, 2005). No nível comportamental, por sua vez, o predomínio do sistema visual é verificado, em certa medida, quando a visão entra em conflito com outro sistema sensorial. Assim, na captura visual, a percepção é conduzida pela forma como as coisas parecem visualmente, mesmo nos casos em que há outra informação sensorial conflitante (SCHIFFMAN, 2005).

No âmbito da percepção da fala, a visão ganhou atenção e relevância a partir do momento em que a percepção deixou de ser vista como um processo apenas monomodal. Ao longo dos anos, alguns trabalhos evidenciaram que a percepção é, de fato, um processo multimodal. E isso tem sido observado, sobretudo, na comunicação face a face, ou seja, a partir da ação da visão. Estudos experimentais revelaram que nossa percepção e compreensão da fala são influenciadas pelos rostos e gestos dos falantes, tanto quanto são influenciados pelo som, ou seja, a informação visual exerce influência semelhante àquela exercida pela informação auditiva, no processo perceptual (MASSARO, 1998).

O primeiro trabalho a investigar a influência da visão para a percepção da fala foi realizado por Sumbly e Pollack (1954). Nesse trabalho, os autores realizaram um experimento em que algumas palavras, mascaradas com ruído de fundo, foram apresentadas sincronizadamente com a imagem da rosto do falante. Os resultados evidenciaram que a percepção dessas palavras melhorou, substancialmente, quando a informação visual foi apresentada aos ouvintes. Os resultados obtidos por Sumbly e Pollack (1954) revelaram que os ouvintes podem explorar, quando necessário, (e, assim, se beneficiar da correspondência informacional entre) aspectos visuais e acústicos do sinal da fala.

Outra importante evidência da participação da visão na percepção da fala foi apresentada por McGurk e MacDonald (1976) em um trabalho que buscava investigar a influência da informação visual na percepção da fala. Nesse experimento, os autores apresentaram o estímulo visual da sílaba [ga] simultaneamente ao estímulo auditivo (som) da sílaba [ba]. Essa sobreposição de estímulos resultou em uma ‘má combinação’, chamada de *mismatch*. O resultado do experimento em condição *mismatch* evidenciou que os informantes

recuperaram a sílaba [da], ou seja, um som que não correspondia a nenhum dos sons presentes nos estímulos visual, [ga], e auditivo [ba].

Os achados de McGurk e MacDonald (1976) apontam para uma ilusão perceptiva, decorrente da existência de uma percepção fonética que não corresponde aos estímulos auditivo e visual que foram apresentados. Tais evidências apontam para a não prevalência de uma informação sobre a outra, o que ocorre é a percepção de um novo elemento, resultado do efeito ilusório. Essa ilusão perceptiva ficou largamente conhecida como “efeito McGurk” (PACHECO, 2006).

Pacheco (2006) salienta que os resultados encontrados pelos autores demonstram que, na percepção da fala, os juízes integram informações auditivas e visuais presentes na face do falante. Em função disso, muitos trabalhos objetivam verificar se a integração desses dois sentidos, visão e audição, ocorreria na percepção fonológica e se afetaria no grau de inteligibilidade da fala.

Ainda de acordo com Pacheco (2006), apesar de certas situações restringirem a integração auditiva e visual, há muitas outras nas quais é possível observar esse processo, como, por exemplo, nas pesquisas comportamentais que têm realizado estudos que exploram as regiões do cérebro envolvidas no processamento do sinal auditivo e visual da fala e os resultados obtidos acenam para uma possível integração dessas duas fontes.

Embora estudos demonstrem que os ouvintes usem tanto a informação auditiva quanto a informação visual, não há, ainda, um consenso sobre como essas duas fontes são utilizadas durante a percepção da fala. Pesquisas têm mostrado que a “fala visível” é particularmente importante nos casos em que o estímulo auditivo é, em alguma medida, prejudicado (MASSARO, 1987). Contudo, a forte influência da informação visual não está limitada apenas às situações de comprometimento do estímulo auditivo; ela ocorre mesmo quando a informação visual é apresentada simultaneamente com o estímulo auditivo perfeitamente inteligível.

### **1.1 O modelo Fuzzy da percepção da fala**

O modelo lógico da percepção difusa (FLMP), modelo *Fuzzy* da percepção da fala, foi proposto por Oden e Massaro (1978) como uma descrição das fontes integradas na percepção da fala, isto é, a descrição de como múltiplas pistas acústicas são integradas para alcançar um único percepto acústico. A premissa dessa teoria é que a percepção ocorre de acordo com um algoritmo geral,

independente da modalidade ou natureza particular dos padrões. O modelo *Fuzzy* defende que a percepção da fala ocorreria a partir de três operações: a avaliação, integração e decisão, sendo que o processo de categorização ocorreria na última operação.

Na perspectiva desse modelo, o sistema sensorial transforma um evento físico e disponibiliza várias fontes de informação chamadas traços. Esses traços, valorizados continuamente, são avaliados, integrados e comparados com descrições prototípicas na memória. Assim, a decisão de identificação é feita na base do valor relativo da correspondência entre a informação do estímulo e as descrições prototípicas relevantes.

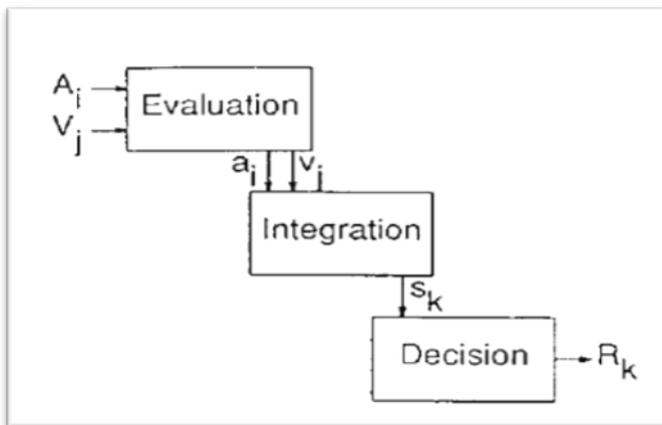
Durante a primeira operação envolvida na percepção da fala, a **avaliação**, os traços dos estímulos são avaliados em termos de protótipos que são gerados para a tarefa em mãos. Para cada traço e para cada protótipo, traços avaliados fornecem informações sobre o grau em que o traço do sinal coincide com o valor do traço do protótipo (traços já avaliados fornecem informações sobre o grau de correspondência entre o traço do sinal e o traço do protótipo). Durante a segunda operação, **integração**, os traços correspondentes a cada protótipo são combinados (ou unidos em termos lógicos). O resultado da integração dos traços consiste do nível a que cada protótipo corresponde a um estímulo. Na terceira operação, **decisão**, o valor de cada protótipo relevante é avaliado em relação à soma de todos os valores dos protótipos relevantes.

O valor relativo das correspondências entre estímulo e identificação dá a proporção das vezes em que o estímulo foi identificado como um modelo do protótipo ou uma decisão de classificação que indica o grau em que o estímulo coincide com a categoria. Uma forte predição da FLMP é que a contribuição de uma fonte de informação para desempenho (identificação/ categorização) cresce com o aumento da ambiguidade das outras fontes de informação disponíveis. Por exemplo, no caso da integração das informações auditiva (voz) e visual (face) na percepção, cada fonte de informação é mais influenciável, à medida que a outra fonte é mais ambígua, ou seja, para a percepção, quanto mais ambígua a informação auditiva mais influenciável é a informação visual, e vice-versa.

Na perspectiva do FLMP, o reconhecimento perceptual é visto como tendo múltiplas fontes de informação disponíveis, apoiando a identificação e interpretação de um *input*. As afirmações centrais desse modelo são: (1) cada fonte de informação é avaliada em um nível contínuo em que, para cada fonte

específica, há diversas alternativas; (2) as fontes de informação são avaliadas independentemente das outras; (3) as fontes são integradas para fornecer um nível geral de apoio para cada alternativa; e (4) identificação e interpretação perceptual seguem o nível relativo de força entre as alternativas.

**Figura 1: Representação dos três processos envolvidos no reconhecimento perceptual**



Fonte: (MASSARO, 1987, p. 314)

No FLMP, as fontes de informação auditiva ( $A_i$ ) e visual ( $V_j$ ) são processadas na fase de avaliação e transformadas em valores psicológicos ( $a_i$  e  $v_j$ ). Essas fontes são integradas, dando um nível geral do suporte,  $s_k$ , para cada alternativa da fala,  $k$ . A operação de decisão mapeia o poder de integração de algumas alternativas de respostas,  $R_k$ . A resposta pode tomar a forma de uma decisão discreta ou um índice do nível do quanto a alternativa é apropriada / provável.

Os teóricos desse paradigma defendem que o FLMP permite determinar como uma fonte de informação é processada e integrada com outras fontes de informação. De acordo com Massaro (1987), esse paradigma já provou ser efetivo nos estudos sobre a integração auditiva e visual na percepção da fala. Assim, esse paradigma não objetiva apenas investigar como diferentes fontes de informação são avaliadas e integradas, ele objetiva descobrir quais fontes de informação são realmente usadas na percepção da fala.

Diante do exposto, nosso objetivo é avaliar a percepção das fricativas em relação aos estímulos auditivo e visual; buscamos verificar, também, se, em caso de manipulação do sinal acústico, essas informações são integradas na percepção das fricativas ou se há a prevalência de um ou outro estímulo na percepção desses sons.

## 2 Metodologia

### 2.1 Montagem do corpus

Para este trabalho, foi montado um *corpus* composto por dissílabas com estrutura silábica  $C_1V_1C_2V_2$ , onde  $C_1$  é uma das seis fricativas do Português Brasileiro, doravante PB, a saber: /f/, /v/, /s/ z/, /ʃ/ e /ʒ/,  $C_2$  é uma oclusiva e  $V_1$  e  $V_2$  é uma das vogais /a/, /i/ ou /u/. Desta forma, o *corpus* foi constituído de pares mínimos em que a oposição é dada em relação ao vozeamento e ponto de articulação das fricativas. Nos casos em que não foi possível encontrar pares mínimos, optamos por utilizar pares análogos, como pode ser observado no quadro abaixo.

**Quadro 1: Exemplos do corpus**

Labiodentais	Alveolares	Palatoalveolares
FACA / VACA	SAGA / ZAGA	CHAPA / JAPA
FIGA / VIGA	SICA / ZICA	CHIPE / JIPE
FUGA / VUBA <sup>i</sup>	SUCO / ZUCO <sup>ii</sup>	CHUCA / JUTA <sup>i</sup>

<sup>i</sup> Par análogo; <sup>ii</sup> Logatoma.

### 2.2 Gravação do corpus

As palavras do *corpus*, formado por dissílabos, foram inseridas na frase-veículo “Digo X baixinho” e gravadas por um sujeito do sexo feminino, de 20 anos de idade, sem distúrbios de fala. A gravação foi realizada em ambiente acusticamente tratado, com vistas a evitar que o ruído do ambiente compromettesse a qualidade sonora, e, conseqüentemente, as análises. Essas gravações ocorreram no Laboratório de Pesquisa em Estudos Fonéticos e Fonológicos (LAPEFF), da Universidade Estadual do Sudoeste da Bahia – UESB, de Vitória da Conquista/

BA, por meio de um MacBookPro11,1, OSX, 10.9.4 (13E28), 2,4 GHz Intel Core i5, que possui câmera de alta resolução, Apple Camera VendorID\_0x106B ProductID\_0x1570, e microfone interno com controle de ruído. Assim, foi possível obter som e imagem de qualidade satisfatória.

### **2.3 Manipulação do sinal acústico**

Como evidenciado por pesquisadores como Harris (1958) Heinz e Stevens (1961), entre outros, para o inglês, e Silva (2012), para o PB, as características do ruído fricativo, especialmente a duração e a frequência, são pistas importantes para a percepção desses sons. Assim, realizamos manipulações da duração do ruído e do espectro de frequência das fricativas com o objetivo de investigar qual o papel das informações auditiva e visual para a percepção desses sons, quando eles apresentam ambiguidade do sinal acústico.

A manipulação da duração do ruído fricativo foi realizada por meio da ferramenta *Manipulate* > *To manipulation* do Software Praat (BOERSMA; WEENINK, 2002). Para a redução do ruído fricativo, foi mensurada a duração absoluta do ruído fricativo, extraído da palavra, e sobre esse valor foram aplicadas três taxas de redução: 25%, 50% e 75. Após a redução do ruído fricativo a 25%, 50% e 75%, as três novas fricativas foram inseridas individualmente na frase no áudio original. Dessa forma, cada palavra do *corpus* passou a contar com três versões de estímulos sonoros, além do sinal original.

A manipulação do espectro de frequência das fricativas, por sua vez, foi realizada a partir da utilização da ferramenta *modify*>*In-line filters*>*Filters with one formant (in-line)* do Praat (BOERSMA; WEENINK, 2002). Essa ferramenta possibilita a manipulação da frequência e da largura de banda dos formantes. Os valores de frequência do ruído fricativo foram manipulados com base em duas referências principais, Stevens (1960) e Ferreira-Silva et al. (2015). O primeiro autor definiu os valores de frequência mínimo e máximo para cada uma das fricativas de acordo com o ponto de articulação. Já Ferreira-Silva et al. (2015) definiram os valores médios de frequência para as fricativas do PB com base na análise do primeiro momento espectral, centroide. Observados os valores do espectro de frequência das fricativas, foram aplicadas três taxas de manipulação da frequência: valor mínimo, médio e máximo para cada fricativa. Tal como para a manipulação da duração, após a manipulação da taxa de frequência do

ruído fricativo, as fricativas foram reinseridas individualmente na frase do áudio original. Dessa forma, cada palavra do *corpus* passou a contar com três versões de estímulos sonoros com frequência manipulada, além do sinal original.

#### **2.4 Montagem do teste de percepção**

Após preparação dos estímulos, foi montado o teste de identificação perceptual (TI) para avaliarmos o papel das informações auditivas e visuais na percepção das fricativas do PB. Esse teste foi montado através do software TP (Treino/Teste de percepção), disponível no site [www.worken.com.br/tp/](http://www.worken.com.br/tp/) (RAUBER; RATO; KLUGE; SANTOS, 2012) que é uma ferramenta que permite criar, configurar e aplicar testes de identificação e discriminação perceptual.

Na tarefa de identificação perceptual, os estímulos das fricativas foram apresentados, em uma primeira condição, em sua versão original, sem nenhuma manipulação do sinal, com os estímulos visual, auditivo e audiovisual. Em uma segunda condição, as fricativas com a duração e a frequência manipuladas foram apresentadas com os estímulos auditivo e audiovisual (áudio manipulado + vídeo original). Os testes foram realizados com 13 juízes (8 mulheres e 5 homens), com idade entre 24 e 45 anos, todos naturais de Vitória da Conquista/BA, que não apresentavam queixa de perda da acuidade auditiva.

#### **2.5 Análise Estatística**

A análise estatística foi realizada por meio do software BioEstat (AYRES et al. 2005). O teste estatístico utilizado foi o de Análise de Variância – Kruskal-Wallis. Este é um teste não paramétrico, conhecido como Teste H, que se destina à comparação de três ou mais amostras independentes, do mesmo tamanho ou desiguais, cujos escores devem ser mensurados, pelo menos, em nível ordinal. Os dados que apresentaram valores significativos, ou seja,  $p \leq 0,05$ , para  $\alpha=0,05$ , foram submetidos aos testes de separação de médias, Dunn e o Student-Newman-Keuls. O primeiro teste, Dunn, foi nossa opção para a separação das médias por ser considerado mais conservador que o segundo, pois retifica o *Nível Alfa* conforme o número de amostras: *Nível Alfa* ( $k(k-1)$ ), sendo  $k$  o número de amostras (AYRES et al.).

### 3 Resultados e Discussão

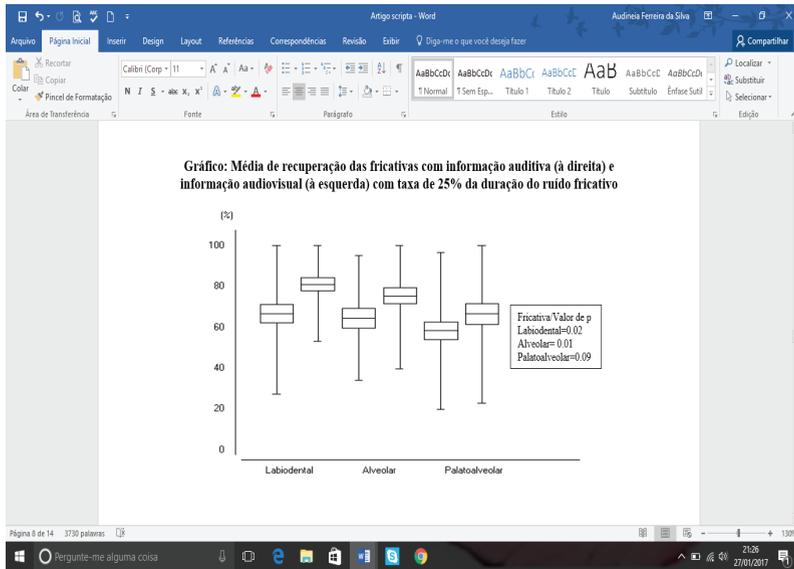
Ao avaliarmos a perceptibilidade das fricativas em função dos parâmetros duração segmental e frequência do espectro, observamos diferenças nos resultados para a identificação das fricativas em função do tipo de informação, auditiva ou audiovisual, com a qual as fricativas foram apresentadas no teste<sup>2</sup>. Assim, buscamos avaliar a perceptibilidade das fricativas com manipulação da duração e da frequência espectral em função do tipo estímulo com o qual as fricativas foram apresentadas.

Os resultados para a percepção das fricativas com duração do ruído reduzido, em função do tipo de informação, indicam que as médias de recuperação das fricativas são maiores em situação de informação audiovisual, com exceção da fricativa alveolar, com 75% de duração, que apresentou média de recuperação maior com a informação auditiva. Podemos observar que a comparação entre as médias de recuperação das fricativas com informação auditiva e audiovisual se mostrou significativa com a taxa de 25% de duração para as fricativas labiodentais e alveolares, com taxa de 50% para as fricativas alveolares e palatoalveolares e com taxa de 75% para as palatoalveolares. Essa significância estatística evidencia que, diante da redução da duração, há um aumento considerável no desempenho perceptual das fricativas com a informação audiovisual.

---

<sup>2</sup> Para este artigo, possíveis diferenças na percepção das fricativas entre homens e mulheres não foram consideradas.

### Gráfico 1: Média de recuperação das fricativas com informação auditiva (à direita) e informação audiovisual (à esquerda) com taxa de 25% da duração do ruído fricativo



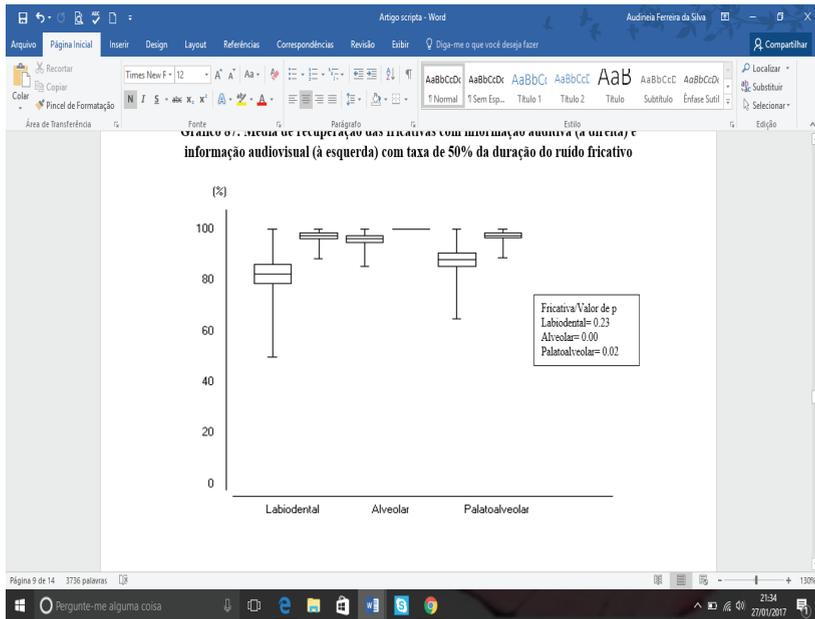
Fonte: Elaboração dos autores

No gráfico 1, podemos observar que, com a taxa de 25% de duração, as fricativas labiodentais, alveolares e palatoalveolares foram mais bem identificadas com a informação audiovisual do que com a informação auditiva. As fricativas labiodentais apresentaram média de 66.92% e 81.03%, respectivamente, com informação auditiva e audiovisual. As fricativas alveolares apresentaram, para a informação auditiva, a média de 64.59% e, para a informação audiovisual, 75.48%. Com informação auditiva, a palatoalveolar apresentou média de 58.32% e, com a informação audiovisual, apresentou média de 66.57%. Nota-se que, com essa taxa de duração, a diferença entre as médias de recuperação com a informação auditiva e audiovisual foi considerada significativa para as fricativas labiodentais e alveolares. Vemos que os valores de  $p$  para a comparação dessas médias foram, respectivamente, 0.02 e 0.01, ou seja,  $<0.05$ .

Com a taxa de 50% de duração, no gráfico 2, observa-se que as fricativas apresentaram desempenho perceptual melhor com a informação audiovisual do que com a auditiva. As fricativas labiodentais registraram média de 82.38%, para

a informação auditiva, e média de 86.28%, para a informação audiovisual. As alveolares apresentaram 92.21% e 98.69%, respectivamente, para informação auditiva e audiovisual. Para as palatoalveolares, por sua vez, foram encontradas médias de 85.78% e 95.64%, para as fricativas com informação auditiva e audiovisual, respectivamente, como podemos verificar no gráfico 2.

**Gráfico 2: Média de recuperação das fricativas com informação auditiva (à direita) e informação audiovisual (à esquerda) com taxa de 50% da duração do ruído fricativo**



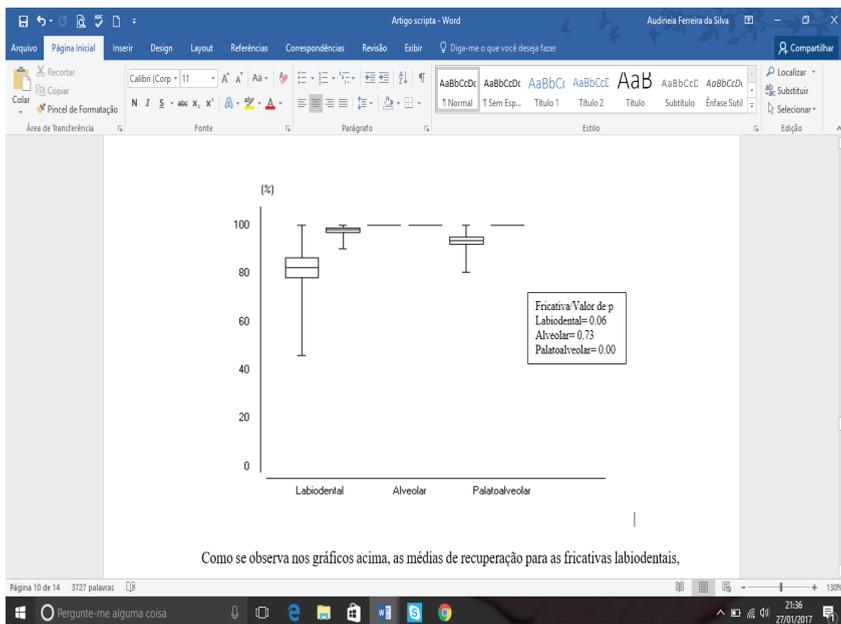
Fonte: Elaboração dos autores

Podemos observar, no gráfico 2, que a comparação entre as médias de recuperação com informação auditiva e audiovisual para as fricativas se mostrou significativa para as alveolares e palatoalveolares, com valores de  $p$  de 0.00 e 0.02, respectivamente.

No gráfico 3, observamos que, à semelhança do que ocorreu com as fricativas com taxas de 25% e 50% de duração, com a taxa de 75% de duração, a informação audiovisual favoreceu a identificação perceptual das fricativas. As fricativas labiodentais apresentaram 82.43% de recuperação com informação

auditiva e 93.53% com a informação audiovisual. As palatoalveolares, por sua vez, apresentaram médias de 92.17% e 99.56%, respectivamente, com as informações auditiva e audiovisual. Apenas no caso da fricativa alveolar, a média de recuperação com a informação auditiva é maior que a média com informação audiovisual, 98.25% e 97.82%, respectivamente, mas, como vemos, essa diferença não chega a 1%, como ilustrado no gráfico a seguir.

**Gráfico3 Média de recuperação das fricativas com informação auditiva (à direita) e informação audiovisual (à esquerda) com taxa de 75% da duração do ruído fricativo**



**Fonte: Elaboração dos autores**

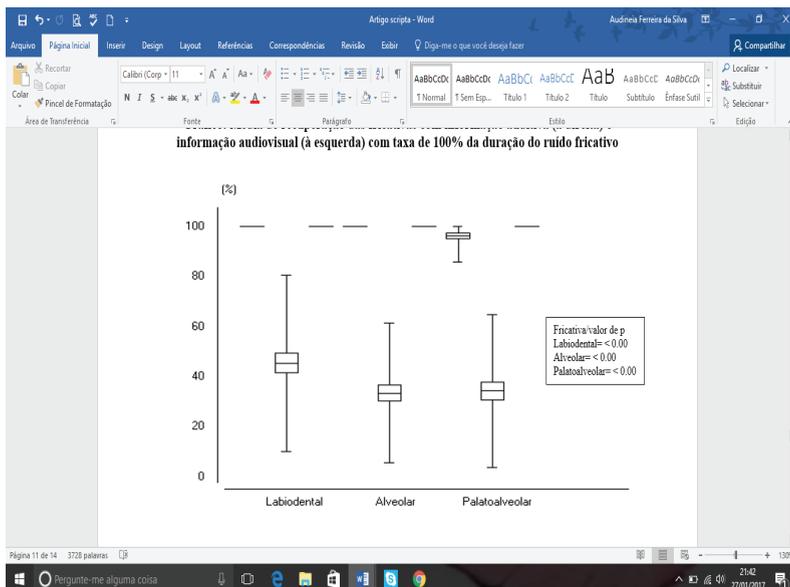
Como se observa nos gráficos 3, as médias de recuperação para as fricativas labiodentais, e palatoalveolares são sempre maiores quando elas foram apresentadas com informação audiovisual. Podemos inferir que a manipulação da duração do ruído foi menos prejudicial para a correta identificação das fricativas quando elas eram apresentadas com estímulo audiovisual.

No gráfico 4, temos a comparação das fricativas com 100% de duração, ou seja, sem manipulação do sinal fricativo. Nessa situação, as fricativas foram

apresentadas em três tipos de informação, auditiva, visual e audiovisual.

Podemos observar que as fricativas apresentaram suas menores médias de recuperação quando apresentadas com a informação exclusiva da visão. Com essa informação, as labiodentais foram as que apresentaram maior média de recuperação, 45.46%, enquanto que a alveolar apresentou 33.47% e a palatoalveolar apresentou 34.32%. Como se nota, com a informação auditiva, as fricativas labiodentais e palatoalveolares apresentam comportamento oposto, com relação ao tipo de informação. A fricativa labiodental apresenta maior média de recuperação, 99.56%, ao passo que a palatoalveolar apresenta a menor média de recuperação, 93.51%. Quando apresentadas com a informação audiovisual, a fricativa labiodental é a que apresenta a menor média de recuperação, 97.43%, enquanto a palatoalveolar apresenta a maior média, 99.12%. Com a informação audiovisual, as fricativas alveolares apresentaram valor médio intermediários, 98.25%, entre os valores para as labiodentais e palatoalveolares, como podemos observar, também, no gráfico 4.

**Gráfico 4: Média de recuperação das fricativas com informação auditiva (à direita), informação apenas visual (centro) e informação audiovisual (à esquerda) com taxa de 100% da duração do ruído fricativo.**

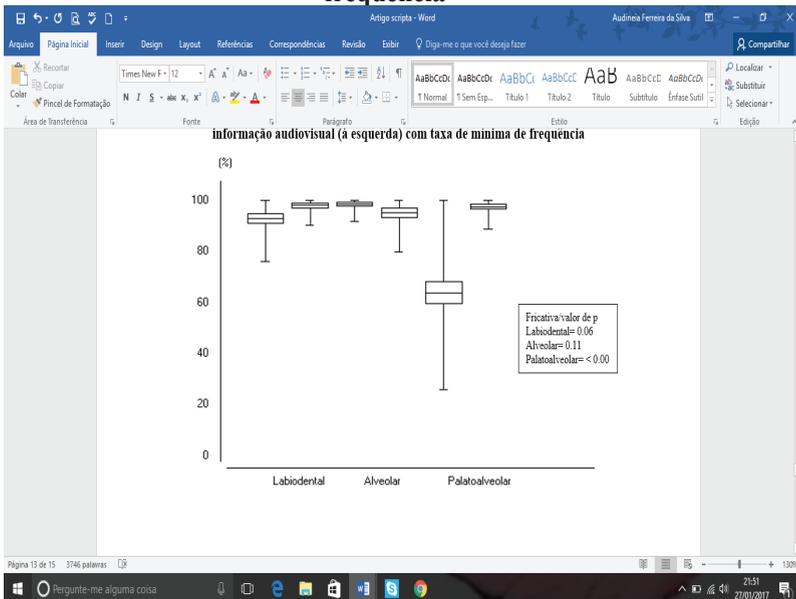


Fonte: Elaboração dos autores

Como podemos observar, quando as fricativas foram apresentadas sem manipulação do sinal, 100% de duração, as médias para a recuperação com as informações auditiva e audiovisual são parecidas. Verificamos que é com a informação visual que a identificação perceptual das fricativas foi mais prejudicada. Notamos que as médias de recuperação das fricativas com informação auditiva e audiovisual são estatisticamente diferentes das médias de recuperação das fricativas com informação somente visual.

Para a manipulação da frequência do espectro, observamos que os resultados se assemelham aos encontrados para a duração. É possível observar que as médias de recuperação das fricativas com informação audiovisual são, em geral, maiores que as médias de recuperação com a informação auditiva. No gráfico 5, notamos que, com a taxa mínima de frequência, as fricativas labiodentais e palatoalveolares registraram as maiores médias de recuperação quando apresentadas com a informação audiovisual, respectivamente 94.80% e 94.37%. Com essa taxa, a fricativa alveolar foi mais bem identificada perceptualmente com a informação auditiva, 95.66%.

**Gráfico 5: Média de recuperação das fricativas com informação auditiva (à direita) e informação audiovisual (à esquerda) com taxa de mínima de frequência**

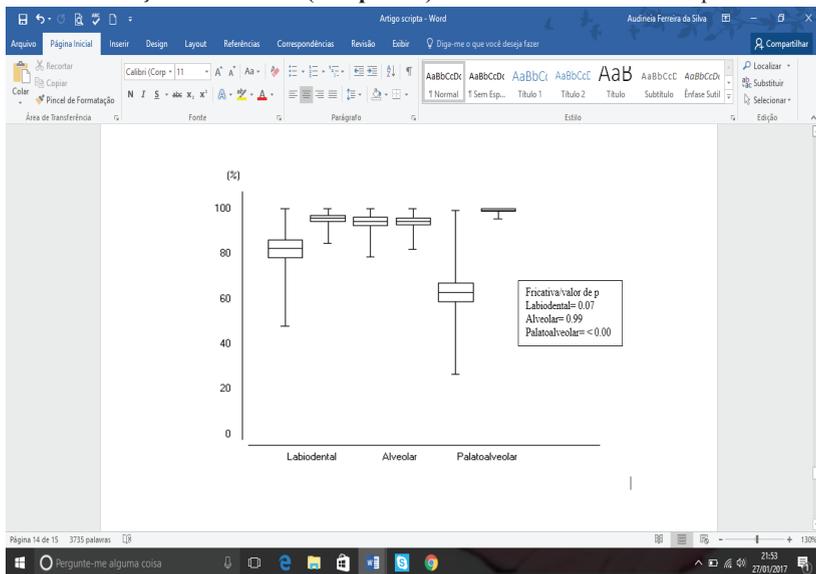


Fonte: Elaboração dos autores

Como vemos no gráfico acima, a diferença entre as médias de recuperação com a informação auditiva e audiovisual se mostrou significativa apenas para as fricativas palatoalveolares, com valor de  $p < 0.00$ .

Quando apresentavam taxa média de frequência, como ilustrado no gráfico 6, as fricativas foram, de forma categórica, mais bem identificadas perceptualmente quando apresentavam informação audiovisual.

**Gráfico 6: Média de recuperação das fricativas com informação auditiva (à direita) e informação audiovisual (à esquerda) com taxa de média de frequência**



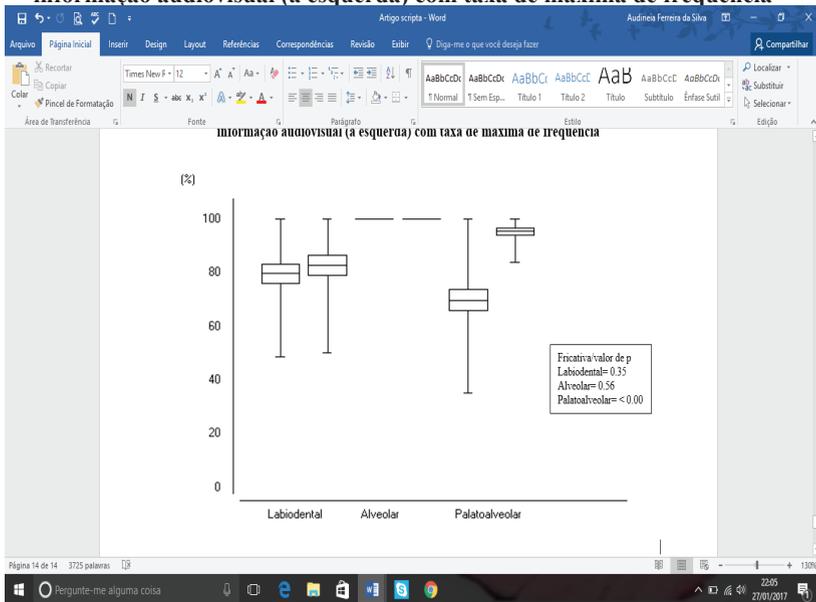
Fonte: Elaboração dos autores

Com a taxa média de frequência, as fricativas labiodentais apresentaram médias de recuperação de 82.39% e 92.65%, respectivamente, com as informações auditiva e audiovisual. As fricativas alveolares, por sua vez, apresentaram médias de 89.66% e 91.76% quando apresentadas com as informações auditiva e audiovisual, respectivamente. As palatoalveolares, no mesmo contexto, apresentaram médias de 63.02% e 95.69%, quando apresentadas com informação auditiva e audiovisual, respectivamente. No caso dessa fricativa, a comparação entre as médias de recuperação se mostrou significativa, com valor de  $p < 0.00$ .

Com a taxa máxima de frequência, observa-se que as fricativas labiodentais e palatoalveolares apresentaram as maiores médias de recuperação com a informação

audiovisual. A fricativa alveolar, por sua vez, apresentou a maior média com a informação auditiva, conforme gráfico 7.

**Gráfico 7: Média de recuperação das fricativas com informação auditiva (à direita) e informação audiovisual (à esquerda) com taxa de máxima de frequência**



Fonte: Elaboração dos autores

No gráfico 7, vemos que, com a taxa máxima de frequência, as fricativas labiodentais apresentaram médias de 79.76% e 82.80% com as informações auditiva e audiovisual, respectivamente. A fricativa alveolar apresentou 99.56% e 99.12% de média quando apresentadas com as informações auditiva e audiovisual. Já as palatoalveolares apresentaram médias de recuperação de 69.87% e 93.06% com as informações auditiva e audiovisual, respectivamente. Para essa fricativa, a comparação entre as médias de recuperação foi significativa, apresentando valor de  $p < 0.00$ .

#### 4 Considerações Finais

Nossos resultados evidenciam que os índices de identificação perceptual das fricativas foram, de maneira geral, maiores quando elas apresentavam a informação audiovisual. Nossos achados atestam que, quando as fricativas têm seu sinal manipulado, em termos de duração e frequência, seu desempenho

perceptual aumenta nos casos em que a informação visual é apresentada com a auditiva, ou seja, diante da manipulação do sinal, as fricativas apresentam médias de recuperação mais altas com o estímulo audiovisual do que com o estímulo, apenas, auditivo. Esses resultados corroboram os tradicionais estudos de Sumbly e Pollack (1954) e McGurk e MacDonald (1976), que afirmam que, na percepção, os ouvintes integram visão e audição. Assim, e considerando os pressupostos do FLMP de Massaro (1987), nossos resultados evidenciam que, na percepção da fala, cada fonte de informação é mais influenciável, na medida em que a outra fonte é mais ambígua. Dessa forma, em nossos resultados, quanto mais a informação auditiva se mostrou ambígua, em decorrência da manipulação da duração e frequência, mais influenciável foi a informação audiovisual (visual). Diante disso, podemos inferir, com base em nossos resultados, que as informações presentes no sinal visual das fricativas colaboraram para o aumento nos índices de identificação e discriminação perceptual desses sons, ou seja, a informação visual tem participação importante na percepção das fricativas do Português Brasileiro.

## Auditory and Visual Perception of Fricatives in Brazilian Portuguese with Acoustic Signal Manipulation

### Abstract

This article aims to evaluate the role of auditory and visual information in the perception of Brazilian Portuguese fricatives, when they are presented with ambiguities/manipulation in the acoustic signal. To do this, we have gathered a disyllable word corpus with the  $C_1V_1.C_2V_2$  syllable structure, where  $C_1$  is a fricative,  $C_2$  is a plosive, and  $V_1$  e  $V_2$  are /a/ or /i/ or/u/. After recording the corpus, the acoustic signal for the fricatives has been manipulated regarding both their noise duration and their spectral frequency. Our results show that fricative identification indexes were generally bigger when they showed audiovisual information. When fricatives have their signal manipulated in terms of duration and frequency, our findings show that their perceptual performance increases in the cases when visual information is presented together with auditory information. This means that, in regard to signal manipulation, fricatives present higher recovery averages with audiovisual information rather than with auditory information alone. These findings comply with the traditional studies by Sumbly & Pollack (1954), as well as those by McGurk & MacDonald (1976). They stated that hearers integrate eyesight and hearing in speech perception. Taking Massaro's (1987) FLMP presuppositions into consideration, our findings bring evidence that the more ambiguous the other source, the more influential each information source becomes in speech perception.

Keywords: Perception. Fricatives. Auditory and Visual information.

## Referências

- AYRES, M.; AYRES JR., M.; AYRES, D. L., SANTOS, A.S. **BioEstat**. Versão 5.0, Sociedade Civil Mamirauá, MCT – CNPq, Belém, Pará, Brasil. 2007.
- BOERSMA, P.; WEENINK, D. **Praat**. [Computer software]. Amsterdam, The Netherlands: Institute of Phonetic Sciences, University of Amsterdam. 2002.
- FERREIRA-SILVA, A.; PACHECO, V.; CAGLIARI, L. C. Descritores estatísticos na caracterização das fricativas do Português Brasileiro: Características espectrais das fricativas. **Acta, Scientiarum. Language and Culture**. Maringá, v. 37, n. 4, p. 371-379, Oct.-Dec., 2015.
- HARRIS, K. S. Cues for the Discrimination of American English Fricatives in spoken syllables. **Lang. Speech** 1, 1-7, 1958.
- HEINZ, J. M.; STEVENS, K. N. On the Properties of Voiceless Fricative consonants. **J. Acoust. Soc. Am.** V. 33, 589-596. 1961.
- MASSARO, D. W. **Speech Perception by Ear and Eye: A paradigm for psychological inquiry**. Hillsdale, NJ: Lawrence Erlbaum Assoc, Inc. 1987.
- MASSARO, D. W. Illusions and Issues in Bimodal Speech Perception. In: **Auditory Visual Speech Perception Conference - AVSP'98**. Sydney. p. 21-26, 1998.
- MCGURK, H.; MACDONALD, J. Hearing lips and seeing voices. **Nature**. V. 264, n. 23, p. 746-748, dec. 1976.
- ODEN, G. C.; MASSARO, D. W. Integration of featural information in speech perception. **Psychological Review**, 85, 172-191. 1978
- PACHECO, V. **O efeito dos estímulos auditivo e visual na percepção de marcadores prosódicos lexicais e gráficos usados na escrita do Português do Brasil**. (Tese de Doutorado). Campinas, SP: [s.n.], 2006.
- RAUBER, A. S.; RATO, A.; SANTOS, G. R.; KLUGE, D. C.; FIGUEIREDO, M. **TP Perception Tests and Perceptual Training With Immediate Feedback**, versão 2.0. Disponível em: [http://www.worken.com.br/tp\\_regfree.php](http://www.worken.com.br/tp_regfree.php), acesso em 12 agosto de 2012.
- SCHIFFMAN, H. R. **Sensação e percepção. Tradução** L. A. F. Pontes & S. Machado. Rio de Janeiro: LTC, 2005.

SILVA, A. F. **Estudo das características acústicas das fricativas do Português Brasileiro**. 108 f. (Dissertação de mestrado) Universidade Estadual Paulista “Júlio de Mesquita Filho”, Araraquara. 2012.

STREVENS, P. Spectra of Fricative Noise in Human Speech. **Language and Speech**, v. 3, 32-49. 1960.

SUMBY W. H, POLLACK I. Visual Contribution to Speech Intelligibility in noise. **Journal of Acoustical Society of America**, v. 26, n. 2, 212-215, 1954.

