

Costa-Guarisco, Letícia Pimenta; Fraga, Francisco José; Iório, Maria Cecília Martinelli
Índice Percentual de Identificação de Fonemas Fricativos: proposta para avaliação da
compressão de frequências

Audiology - Communication Research, vol. 21, 2016
Academia Brasileira de Audiologia
São Paulo, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=391544881027>

Índice Percentual de Identificação de Fonemas Fricativos: proposta para avaliação da compressão de frequências

Fricative Phonemes Identification Index: proposal for frequency compression evaluation

Letícia Pimenta Costa-Guarisco¹, Francisco José Fraga², Maria Cecília Martinelli Iório³

RESUMO

Objetivo: Desenvolver um teste para avaliar a compressão de frequências por meio da identificação de fonemas fricativos e aplicá-lo. **Métodos:** Organizou-se um material de fala composto por 24 palavras monossilábicas, contendo os fonemas /s, z, f, v, ſ, ʒ/. As palavras foram gravadas por oito sujeitos, totalizando uma amostra de 192 gravações. Aplicaram-se filtros passa-baixas nas palavras com frequências de cortes em 1,5 kHz; 2 kHz e 3 kHz, a fim de impedir a detecção sonora nas faixas de frequências filtradas, simulando perdas auditivas em frequências altas. Aplicou-se o teste em dez indivíduos normo-ouvintes, que deveriam ouvir a palavra e identificá-la na tela de um computador. Foram oferecidas 12 opções de resposta, fixando a vogal e variando o fonema fricativo inicial e a presença da fricativa final /s/. O teste foi composto por 384 repetições de palavras, sendo 192 com compressão de frequências. As respostas foram computadas e comparadas por meio de testes estatísticos. **Resultados:** A identificação dos fonemas foi favorecida pelo aumento da frequência de corte, exceto em 3 kHz. Houve melhora na identificação dos fonemas fricativos /ʃ, ʒ/ com o uso da compressão de frequências, porém, para os demais fonemas avaliados (/f, v/ e /s, z/) não houve diferenças. **Conclusão:** O teste de identificação de fonemas fricativos se mostrou eficiente e confiável. Para as frequências de corte em 1,5 kHz e 2 kHz, a compressão de frequências foi eficaz para a identificação dos fonemas fricativos /ʃ, ʒ/, porém, a identificação do /s/ final mostrou-se comprometida pelo algoritmo.

Descritores: Perda auditiva de alta frequência; Testes de discriminação da fala; Auxiliares de audição; Perda auditiva neurosensorial; Percepção da fala

ABSTRACT

Purpose: Develop a test to evaluate frequency compression by identifying and applying fricative phonemes. **Methods:** A speech material was composed of 24 monosyllabic words, containing the phonemes /s, z, f, v, ſ, ʒ/. Eight subjects recorded words, for a total sample of 192 recordings. Low-pass filters were applied to words with cutoff points values of 1,5 kHz; 2 kHz and 3 kHz, in order to prevent sound detection in the filtered frequency bands, simulating hearing loss at high frequencies. The test was applied to ten normal hearing individuals, who had to listen to the word and identify it on a computer screen. Twelve response options were offered, fixing the vowel and varying the initial fricative phoneme and the presence of the final fricative /s/. The test consisted of 384 word repetitions, 192 with frequency compression. Responses were computed and compared using statistical tests. **Results:** The identification of the phonemes was favored by the increase of the cutoff frequency, except in 3 kHz. There was improvement in the identification of fricative phonemes /ʃ, ʒ/ with the use of frequency compression, however, for the other assessed phonemes (/f, v/ e /s, z/) no differences were found. **Conclusion:** The identification test of fricative phonemes was efficient and reliable. For the cutoff frequencies in 1,5 kHz and 2 kHz, the frequency compression was effective for the identification of the fricative phonemes /ʃ, ʒ/, however the identification of /s/ at the end was affected by the algorithm.

Keywords: Hearing loss, High-frequency; Speech discrimination tests; Hearing aids; Hearing loss, Sensorineural; Speech perception

Trabalho realizado no Curso de Pós-Graduação em Fonoaudiologia (Doutorado), Universidade Federal de São Paulo – UNIFESP – São Paulo (SP), Brasil.
(1) Curso de Fonoaudiologia, Faculdade de Medicina, Universidade Federal de Minas Gerais – UFMG – Belo Horizonte (MG), Brasil.

(2) Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas (CECS) Universidade Federal do ABC – UFABC – Santo André (SP), Brasil.

(3) Curso de Fonoaudiologia, Universidade Federal de São Paulo – UNIFESP – São Paulo (SP), Brasil.

Conflito de interesses: Não

Contribuição dos autores: *LPCG* concepção e delineamento do estudo, coleta, análise e interpretação dos dados e redação do artigo; *FJF* concepção e delineamento do estudo, análise e interpretação dos dados e revisão do artigo; *MCMI* concepção e delineamento do estudo, análise e interpretação dos dados e revisão do artigo.

Autor correspondente: Letícia Pimenta Costa-Guarisco. E-mail: lepcosta@hotmail.com

Recebido em: 9/6/2016; Aceito em: 5/10/2016

INTRODUÇÃO

A deficiência auditiva neurosensorial descendente frequentemente está associada a dificuldades acentuadas no reconhecimento de palavras, principalmente na detecção e discriminação de sons fricativos, mesmo com a utilização de próteses auditivas. Nos últimos anos, pesquisadores têm atribuído essa dificuldade à presença de zonas mortas na cóclea, o que torna determinadas regiões cocleares não responsivas à estimulação sonora, devido à ausência de células ciliadas internas ou neurônios adjacentes funcionais⁽¹⁾. Com a descrição das zonas mortas na cóclea, o estudo dos algoritmos de rebaixamento de frequências, como compressão, transposição e translação de, voltou à tona, na tentativa de oferecer importantes pistas de fala contidas nas altas frequências. Por meio destas técnicas, os componentes de frequências altas são modificados para frequências mais baixas, alterando o espectro sonoro dos sons de altas frequências⁽²⁾.

Por suas características espectrais, os sons fricativos são os mais prejudicados nas deficiências auditivas neurosensoriais descendentes, mas também os que mais se beneficiam das técnicas de rebaixamento de frequências^(3,4,5,6,7). No entanto, algumas vezes, o uso destes recursos gerou distorção, dificultando a identificação desses fonemas^(8,9). As informações acústicas para o reconhecimento de fala são várias, como, por exemplo, as mudanças na percepção da intensidade do som, contidas no “envelope temporal” da fala. Essas informações permitem diferenciar os sons vozeados dos não vozeados, além de fornecerem pistas quanto ao modo articulatório (por exemplo, fricativas X plosivas). No entanto, para melhor desempenho nas tarefas de fala, é importante a percepção do modo e ponto articulatório dos sons, sendo necessária uma boa audição em altas frequências⁽¹⁰⁾.

A análise acústica dos sons tem fornecido importantes informações sobre a percepção sonora e pistas de fala^(11,12,13,14). A grande concentração de energia acústica das vogais situa-se na faixa de baixas frequências (0,4 kHz a 0,5 kHz), localizando-se na melhor região da curva de audibilidade humana. Apesar das vogais serem naturalmente mais intensas do que as consoantes, a inteligibilidade da fala depende dos sons consonantais, cuja distribuição de energia é pequena e alcança frequências superiores a 2 kHz. A energia de fala concentrada nos sons agudos é de, aproximadamente, 20 dB a 35 dB mais fraca do que a energia de fala concentrada em 0,5 kHz, o que pode explicar a dificuldade que indivíduos com perda auditiva em altas frequências apresentam no reconhecimento de fala⁽¹⁴⁾.

As consoantes podem ser classificadas quanto ao ponto e modo de articulação, além de sonoridade. Os fonemas sonoros apresentam maior intensidade, são mais graves e mais curtos em duração, comparados a seus pares surdos, o que contribui para a discriminação do traço de sonoridade. Os sons de fricativos são caracterizados pela passagem do fluxo de ar por um estreitamento na boca, suficiente para produzir uma turbulência

áerea. No português falado no Brasil, as fricativas são o /f, v, s, z, ʃ, ʒ/, que se diferenciam quanto ao ponto articulatório e sonoridade, mantendo o modo de produção do som. As anteriores /f, v/ apresentam a intensidade mais reduzida de todas as consoantes do português e sua faixa de frequências é bastante ampla, indo de 1,2 kHz a 7 kHz. As médias /s, z/ são sons mais fortes, agudos, com faixa de frequências acima de 4,5 kHz, chegando a 8 kHz, no português. As posteriores /ʃ, ʒ/ também apresentam faixa de frequência ampla, semelhante a /s, z/, porém mais grave, com frequências entre 2,5 kHz e 6 kHz⁽¹⁴⁾.

Muitas variáveis podem interferir na percepção e discriminação da fala, como os fonemas envolvidos (espectro de frequências baixas e altas), a voz do locutor (graves X agudas) e a extensão da perda auditiva, preservando, mais ou menos, as frequências altas. De maneira geral, os sons fricativos são menos intensos e possuem importantes pistas espectrais de altas frequências, superiores a 2 kHz, que os colocam em situação desfavorável nas perdas auditivas descendentes.

Para controlar essas variáveis e avaliar a percepção das fricativas nas perdas auditivas, este estudo teve como objetivo desenvolver um teste de identificação de fonemas fricativos e aplicá-lo.

MÉTODOS

Esta pesquisa foi desenvolvida após ter sido aprovada pelo Comitê de Ética em Pesquisa da Universidade Federal de São Paulo, sob o protocolo número 0150/07. Os participantes assinaram o Termo de Consentimento Livre e Esclarecido, com todas as informações, antes de se submeterem à tarefa proposta. Apresenta-se aqui um instrumento de avaliação para verificar o Índice Percentual de Identificação de Fonemas Fricativos (IPIFF) e, para tanto, foi proposto um estudo comparativo experimental.

Casuística

O estudo foi realizado em sujeitos normo-ouvintes, com perdas auditivas simuladas, e a identificação das palavras foi analisada e comparada em duas diferentes situações de escuta: fala filtrada e fala comprimida.

A fim de se excluir viés cognitivo e intelectual para inclusão na amostra, os indivíduos foram selecionados, considerando:

- idade entre 18 e 40 anos;
- nível de escolaridade mínimo compatível com ensino médio completo;
- familiaridade com o uso do computador;
- ser brasileiro nato;
- não possuir experiência prévia com o algoritmo de compressão de frequências.

A casuística foi composta por dez voluntários normo-ouvintes, sendo cinco homens e cinco mulheres, com idades entre 23 e 30 anos, nível de escolaridade superior completo e profissões

variadas. Todos os indivíduos realizaram avaliação auditiva, composta por Audiometria Tonal, Logoaudiometria e Imitaciometria. Os procedimentos desta pesquisa foram realizados em sala acusticamente tratada. Para realizar a Audiometria, utilizou-se o audiômetro da marca Grason-Stadler®, modelo GSI-61. Para a Imitaciometria, o imitaciômetro da marca Grason-Stadler®, modelo GSI 38. Considerou-se normalidade auditiva a presença, na Imitaciometria, de curvas timpanométricas tipo A e reflexos acústicos contralaterais presentes bilateralmente; na Audiometria Tonal, limiares auditivos tonais por via aérea inferiores a 20 dB NA, nas frequências de 250 kHz a 8 kHz e na Logoaudiometria, limiares de reconhecimento de fala (LRF) iguais ou inferiores a 20 dB NA e Índices Percentuais de Reconhecimento de Fala (IPRF) iguais ou superiores a 92%, bilateralmente.

Material de fala e processamento do sinal

Para esta avaliação, combinaram-se os fonemas fricativos /s/, /z/, /f/, /v/, /ʃ/, /ʒ/, na posição inicial, com os fonemas vocálicos /a/ e /i/, seguidos ou não do fonema fricativo /s/, na posição final, compondo uma amostra de 24 monossílabos (Quadro 1).

As combinações fonéticas realizadas formaram 18 palavras existentes na língua portuguesa falada no Brasil e seis não palavras (Houaiss, *online*¹), marcadas com um asterisco no Quadro 1. No entanto, todas essas combinações fonéticas compuseram sílabas existentes no vocabulário português.

A gravação original do material de fala e a edição foram realizadas em um laboratório de engenharia de um centro universitário. Em uma sala silenciosa, utilizou-se um microfone omnidirecional, *leadership headphone* multimídia 3969, posicionado em frente ao locutor e no mesmo nível de sua boca, conectado ao computador portátil marca HP®, *special edition*, modelo L2205nr. As palavras foram gravadas por oito locutores, quatro mulheres e quatro homens, brasileiros natos, de diferentes regiões do Brasil, orientados a falar de maneira clara e a manter um nível constante de voz, durante toda a gravação. Todas as 24 palavras foram lidas por cada locutor, com um intervalo aproximado de dois segundos entre elas. As gravações consideradas de qualidade insatisfatória, como

alterações na velocidade da fala ou distorções, foram repetidas. Constituiu-se uma amostra original de 192 palavras (24 palavras distintas repetidas por oito falantes). O material de fala gravado foi digitalizado pelo programa Matlab, a uma taxa de 16.000 amostras por segundo. Para cada palavra, foram identificados os pontos de corte de início e fim, removendo-se as pausas. Após a digitalização dos sinais de fala, foi feita a normalização da intensidade, a fim de padronizar toda a amostra. Em seguida, os sinais de fala foram divididos em quadros com sobreposição de 75% entre um quadro e o seguinte, para captar a dinâmica de transição entre os diferentes fonemas, garantindo-se maior fidedignidade dos sinais digitalizados. Para a análise e processamento, os sinais de fala foram passados do domínio do tempo para o domínio da frequência.

Foram aplicados filtros passa-baixas em palavras monossílabas contendo fonemas fricativos, nas amostras de fala processada pelo algoritmo de compressão de frequências e de fala original, em três frequências de corte diferentes, 1,5 kHz, 2 kHz e 3 kHz, para impedir a detecção sonora nas faixas de frequências filtradas, como ocorre nas perdas auditivas descendentes de grau severo a profundo, sugestivas de zonas mortas na cóclea. Para a aplicação dos filtros, as amostras de fala foram processadas no domínio da frequência, zerando-se a faixa de frequências superior ao corte estabelecido. Após a aplicação dos filtros de frequências, foi necessária uma nova normalização dos sinais de fala sintetizados, por causa das perdas de informações acústicas. Para descobrir como a correção de amplitude deveria ser aplicada aos sinais filtrados, a fim de restaurar o nível de pressão sonora do sinal de fala antes da filtragem, utilizou-se um medidor de nível de pressão sonora do fabricante Radio Shack, modelo 33-2055. O medidor foi colocado na posição equivalente ao ponto médio entre as duas orelhas, a uma distância de um metro do alto-falante e a zero grau azimute. Utilizou-se a escala A, com respostas rápidas para as medições. Para a medição dos níveis de pressão sonora das palavras, foi necessário determinar um valor médio de referência, obtido a partir da média dos valores de pico de maior amplitude das palavras. Selecionou-se, ao acaso, uma amostra de 30 palavras utilizadas no teste, repetidas por três vezes, para registro do pico de pressão sonora. A partir da

Quadro 1. Palavras monossílabas formadas a partir da combinação dos fonemas fricativos /s/, /z/, /f/, /v/, /ʃ/, /ʒ/, na posição inicial, com os fonemas vocálicos /a/, /i/, seguidos ou não do fonema fricativo /s/, na posição final

Ponto articulatório	Labiodentais		Palatais		Alveolares	
	Grupo /f, v/		Grupo /ʃ, ʒ/		Grupo /s, z/	
COM /s/ final	Faz	Vás	Chás	Jaz	Sás*	Zás
	Fiz	Vis	Xis	Giz	Sis*	Zis*
SEM /s/ final	Fá	Vá	Chá	Já	Sá	Zá*
	Fi	Vi	Xi	Gi*	Si	Zi*

Legenda: * não palavras

¹Dicionário Eletrônico – dicionariohouaiss.com.br, 20 fev 2009.

média das 90 palavras repetidas, determinou-se o valor médio de referência do nível de pressão sonora. Este procedimento foi aplicado em todas as situações de escuta. Verificou-se que os sinais sintetizados precisavam de uma amplificação adicional de 5 dB e 4 dB, para a aplicação de filtros com frequências de cortes em 1,5 kHz e 2 kHz, respectivamente.

Foi elaborado um programa específico de computador (*software*) para aplicar o teste Índice Percentual de Identificação de Fonemas Fricativos (IPIFF), que continha todo o material de fala processado. O teste foi composto por 384 palavras, sendo os 24 monossílabos falados por oito locutores. Como o estudo foi realizado em normo-ouvintes, foi necessário aplicar os filtros passa-baixas em todo o material de fala, para simular a perda auditiva. Assim, a amostra de fala conteve 192 gravações de fala modificadas apenas por filtros passa-baixas e 192 gravações de fala com compressão de frequências associada aos filtros.

Por meio do *software*, foi possível arquivar os dados dos sujeitos e os resultados do teste obtidos por sessão, contendo o número de palavras ouvidas, os acertos e o tempo gasto para se completar a sessão. Os índices percentuais de identificação de fonemas fricativos foram disponibilizados pelo programa em tabelas, para a análise dos dados, considerando a extensão da perda auditiva (aqui simulada por filtros passa-baixas) e o processamento do sinal (uso de um algoritmo de compressão de frequências).

Ambiente de teste, equipamentos e calibração

Os procedimentos desta pesquisa foram realizados em sala acusticamente tratada, contendo uma mesa para aparar o computador e uma cadeira confortável, localizada na frente do computador, para acomodar o ouvinte. Para calibrar e controlar a intensidade dos estímulos de fala, utilizou-se o audiômetro da marca Grason-Statler®, modelo GSI-61, acoplado a um computador do tipo portátil, da marca Toshiba®, modelo Pentium 4, Windows XP, com 512 de memória *ram*, contendo o teste, e um *mouse* acoplado para o comando do teste.

A avaliação foi feita em campo livre, sendo necessária uma caixa de som da marca Ventura® 400 W, modelo TNS. O audiômetro foi ajustado para saída em um canal, configurado e calibrado para alto-falante. Utilizou-se a intensidade de 70 dB (A), verificando-se o conforto auditivo para os sujeitos. O procedimento de calibração foi realizado no ambiente em que o teste foi aplicado. Para a calibração dos sinais de fala, utilizou-se o medidor de nível de pressão sonora da marca Radio Shack, modelo 33-2055, posicionado em um ponto médio entre as duas orelhas, a uma distância de um metro do alto-falante e a zero grau azimute, utilizando-se a escala A, com respostas rápidas para as medições. Variou-se o *dial* do audiômetro em intervalos de 1 dB, para se estabelecer 65 dB (A) como valor médio de referência do nível de pressão sonora, por ser um nível de escuta considerado confortável para ouvintes normais. Nas duas situações de escuta, houve uma variação de até, mais

ou menos, 6 dB do valor médio. Esta variação foi considerada adequada, uma vez que, na fala, pode-se observar uma diferença de 30 dB entre o som mais intenso e o menos intenso⁽¹⁵⁾. Houve semelhança entre o valor médio do nível de pressão sonora obtido nas duas situações de escuta. Para calibração do audiômetro, utilizou-se o tom de calibração disponível no programa, ajustando-se a unidade de volume (VU) na posição zero, de forma a garantir a reprodutibilidade das condições de apresentação.

Procedimentos - Teste de Identificação de Fonemas Fricativos

Após a avaliação audiológica, os indivíduos realizaram o teste IPIFF nas três condições de escuta (em relação à frequência de corte), na seguinte ordem de execução: 2 kHz, 1,5 kHz e 3 kHz. Esta ordem foi definida, considerando-se o nível de dificuldade na compreensão das palavras e oferecendo, na primeira sessão, uma condição intermediária de inteligibilidade. A aplicação do teste foi feita em três sessões distintas, em dias diferentes, para evitar a fadiga.

Antes de se iniciar o teste, foi oferecido um treinamento para que os ouvintes pudessem se familiarizar com o programa e também se preparar, assegurando a compreensão do procedimento. O treinamento foi composto por dez palavras, selecionadas aleatoriamente pelo programa e funcionou exatamente como o teste, sendo realizado antes de se iniciar qualquer sessão. As palavras foram oferecidas de forma randomizada pelo programa, sem qualquer ordem lógica ou previsível, de forma a não oferecer pistas, caracterizando um estudo duplo-cego, uma vez que nem os sujeitos avaliados, nem o pesquisador sabiam qual o tipo de processamento de sinal aplicado em cada palavra.

O teste foi elaborado com respostas do tipo múltipla escolha, em que os ouvintes deveriam identificar, na tela, a palavra ouvida. Tal estratégia foi aplicada como forma de garantir a fidedignidade na análise das respostas, uma vez que, devido à presença de algumas palavras sem sentido e outras pouco usuais, a reprodução oral da palavra ouvida poderia se dar por aproximação a uma palavra conhecida e não por reconhecimento dos sons. Além disso, a estratégia de repetição de palavras poderia gerar confusão na resposta, por distorção ou imprecisão articulatória, comprometendo a análise das respostas. Dessa forma, optou-se pela tarefa de leitura e identificação visual de itens como resposta, sendo que a cada palavra pronunciada, 12 opções de resposta eram oferecidas. Como os sons fricativos foram os objetos de análise neste trabalho, os monossílabos oferecidos na tela foram formados pela mesma vogal ("a" ou "i"), para que os ouvintes decidessem qual a fricativa inicial estavam escutando e se havia presença ou ausência da fricativa /s/ final. A distribuição das palavras escritas na tela também se deu de forma aleatória, para que os ouvintes não fossem direcionados ao fazerem suas escolhas. Para esta tarefa, os ouvintes teriam que ser hábeis para detectar, reconhecer auditivamente o

som e identificar a forma escrita da palavra ouvida para, assim, marcarem as suas escolhas.

Os sujeitos foram habilitados a conduzirem o teste sozinhos, marcando com o *mouse* a palavra ouvida. Para isso, escutaram, necessariamente, duas vezes cada palavra, antes de marcarem sua opção e, mais uma vez, a palavra foi repetida para confirmar a escolha. Neste momento, os ouvintes poderiam confirmar ou cancelar sua resposta. Caso cancelassem, tinham uma nova chance de escolha. A opção marcada poderia ser cancelada quantas vezes fossem necessárias, a fim de que os ouvintes se certificassem de sua escolha. Ao confirmar a palavra ouvida, esta era registrada e uma nova palavra era apresentada, dando continuidade ao teste. Ao marcar a última palavra do teste, a sessão se concluía. O pesquisador só teve acesso aos resultados após a conclusão do teste, não atribuindo opinião ou reforço às respostas dadas pelos ouvintes.

Os sujeitos tiveram liberdade de interromper o teste quantas vezes quisessem, em qualquer momento, para descanso, por quanto tempo julgassem necessário. Foi sugerido um intervalo a cada 20 minutos.

Todas as palavras identificadas foram arquivadas automaticamente pelo programa, no momento do teste. Dessa forma, após a conclusão de cada sessão, todas as 384 palavras foram computadas, possibilitando-se a visualização dos fonemas identificados corretamente e das trocas e omissões ou deleções ocorridas (no caso da fricativa final /s/).

Cada fonema foi repetido 32 vezes, em cada condição de escuta, ou seja, os grupos fonêmicos /f, v/, /ʃ, ʒ/ e /s, z/ foram apresentados 64 vezes na fala filtrada e 64 vezes na fala comprimida. Estes números foram calculados de forma a garantir significância estatística. A fricativa /s/ final apresentou uma taxa de 96 ocorrências, em cada condição de escuta.

A fim de neutralizar o efeito de acertos ao acaso, com chance de 1/6 (16,66%) para as fricativas iniciais e de 1/2 (50%) para a fricativa /s/ final, as porcentagens de acerto foram corrigidas, por meio da seguinte fórmula:

$$PA_{corrigida} = \frac{(PA - A_{acaso})}{(100 - A_{acaso})} \times 100$$

Sendo, PA = porcentagem de acertos; PA corrigida = porcentagem de acertos corrigida; A acaso = acertos ao acaso

Método estatístico

Os resultados foram descritos quanto à habilidade de se identificar corretamente os fonemas fricativos, considerando-se as seguintes variáveis: processamento do sinal (filtro passa-baixas X compressão de frequências), frequência de corte (1,5 kHz X 2 kHz X 3 kHz), locutor (masculino X feminino) e grupo fonêmico (/f-v/, /ʃ-ʒ/, /s-z/ e /s/ final).

As PA corrigidas foram utilizadas em todas as análises. Para avaliar o efeito e a interação das variáveis “processamento do

sinal” e “frequência de corte”, utilizou-se o teste estatístico Anova Two-way e adotou-se nível de 95% de confiança.

RESULTADOS

O tempo médio de duração das sessões foi de 75 minutos em 2 kHz, 60 minutos em 1,5 kHz e 40 minutos em 3 kHz. Acredita-se que, pelo fato dos testes terem começado em 2 kHz, gastou-se mais tempo nesta sessão, em razão da não familiaridade com o procedimento.

A visualização gráfica dos IPIFFs médios e erros padrão, para cada frequência de corte, considerando o sexo dos locutores, na fala comprimida e fala filtrada, está apresentada na Figura 1.

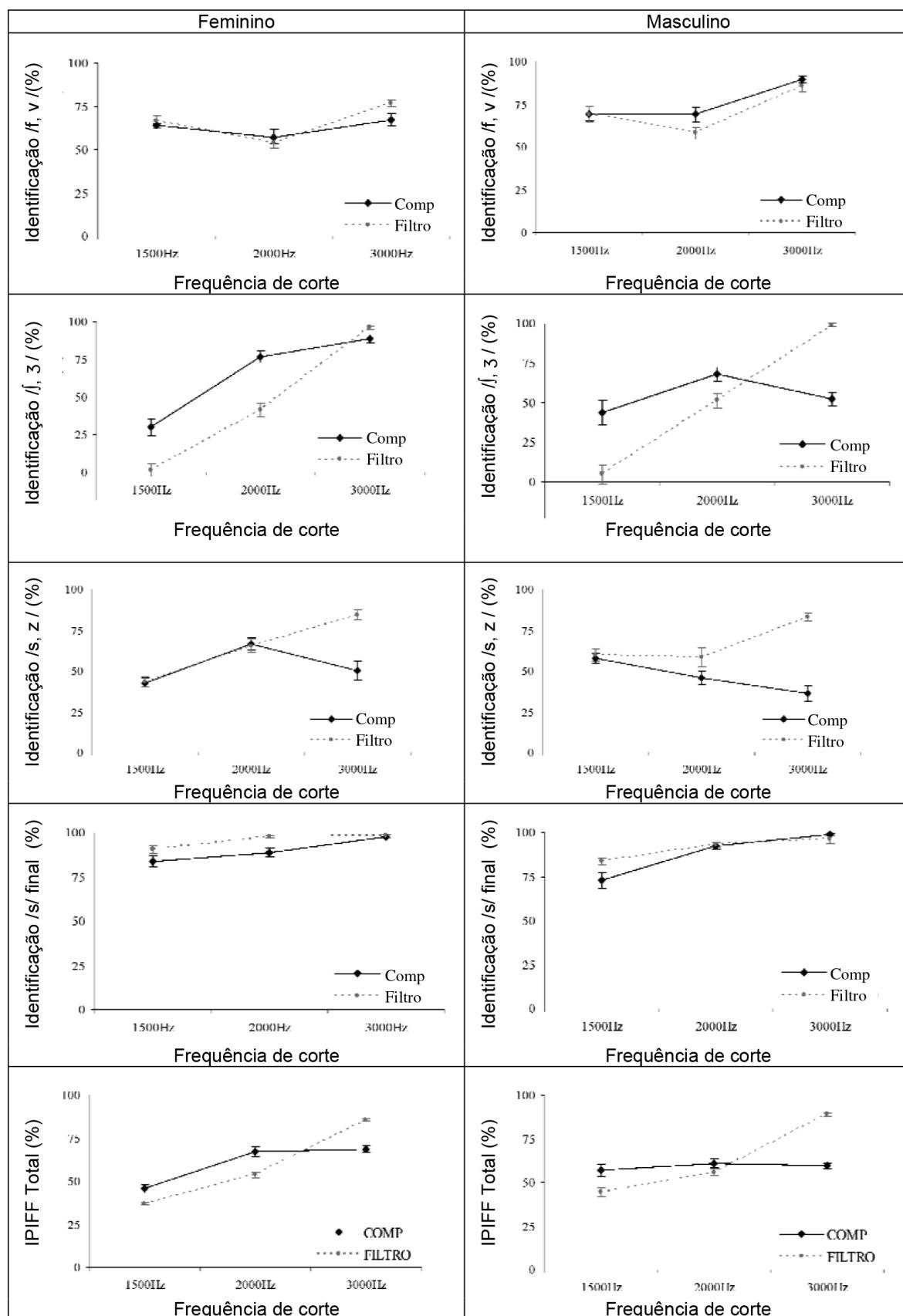
Aplicando-se o teste estatístico Anova Two-way, foi possível analisar o efeito isolado da compressão de frequências e dos filtros passa-baixas na identificação dos fonemas fricativos. Também foi possível avaliar se essas duas variáveis interagem, ou seja, se são dependentes entre si. Para estudar o IPIFF, foram utilizados os valores corrigidos individuais (n=10). Face ao bom desempenho na identificação dos fonemas fricativos na fala filtrada, em 3 kHz, e a piora sistemática em todas as análises com a compressão de frequências neste mesmo corte, optou-se por excluir essa frequência do estudo de efeito e interação entre as variáveis. Os valores médios, desvios padrão e valor de p dos IPIFFs, em cada grupo fonêmico, segundo as variáveis “processamento do sinal” (filtro X compressão) e “frequência de corte” (1,5 kHz e 2 kHz) estão apresentados na Tabela 1 (Tabela 1).

DISCUSSÃO

A simulação de perdas auditivas em normo-ouvintes é uma prática metodológica comum na literatura pesquisada, seja por meio de filtros de frequências^(3,6,16,17,18), seja por outros recursos, como, por exemplo, codificador de voz por bandas de ruído ou *noiseband vocoder*, para simular implantes cocleares^(19,20). A realização deste tipo de experimento com indivíduos normo-ouvintes é vantajosa, pois permite uma avaliação específica do algoritmo, sem influências de quaisquer outros fatores associados à deficiência auditiva neurosensorial.

Foi elaborado um programa específico de computador (*software*), para aplicar o teste Índice Percentual de Identificação de Fonemas Fricativos (IPIFF). Em razão da presença de algumas palavras sem sentido e outras pouco usuais, o teste foi elaborado com respostas do tipo múltipla escolha, em que o ouvinte deveria identificar, na tela do computador, a palavra ouvida. Este tipo de tarefa também foi utilizado em outros estudos^(9,20).

No português brasileiro (PB), os sons africados aparecem apenas como alofones (ex. tia, dia), isto é, não geram oposições fonológicas que levem à distinção de significados⁽¹⁴⁾. Assim, apenas os sons fricativos do PB foram incluídos na amostra de palavras aplicadas neste estudo.



Legenda: IPIFF = Índice Percentual de Identificação de Fonemas Fricativos

Figura 1. IPIFF médios na fala comprimida e na fala filtrada, segundo a frequência de corte para locutores femininos e masculinos, em cada grupo fonêmico, apresentados de cima para baixo, na seguinte sequência: /f, v/, /ʃ, ʒ/, /s, z/, /s/ final e identificação total

Tabela 1. Interação e efeito das variáveis “processamento do sinal” (filtro x compressão) e “frequência de corte” (1,5 kHz x 2 kHz)

Grupo fonêmico	Variáveis de análise	Processamento do sinal	Média de reconhecimento (%)	DP	Valor de p
f-v	Frequência de corte	1,5 kHz	67,10	9,24	0,8414
		2 kHz	59,50	11,52	
	Processamento do sinal	Filtro	61,75	11,99	0,4479
		Compressão	64,85	9,98	
	Interação				0,2277
	Frequência de corte	1,5 kHz	20,03	24,95	0,0000*
ʃ-z		2 kHz	59,31	18,26	
Processamento do sinal	Filtro	24,72	26,42	0,0000*	
	Compressão	54,63	24,49		
Interação				0,4269	
s-z	Frequência de corte	1,5 kHz	50,69	7,23	0,8004
		2 kHz	58,94	11,89	
	Processamento do sinal	Filtro	56,50	12,42	0,1829
		Compressão	53,13	8,32	
	Interação				0,4758
	Frequência de corte	1,5 kHz	82,55	10,07	0,0108*
S final		2 kHz	92,86	4,75	
Processamento do sinal	Filtro	91,09	6,55	0,0006*	
	Compressão	84,32	10,62		
Interação				0,4145	
Total	Frequência de corte	1,5 kHz	45,94	8,52	0,0010*
		2 kHz	59,25	7,89	
	Processamento do sinal	Filtro	47,66	8,57	0,0001*
		Compressão	57,53	10,16	
	Interação				0,7223

* Valores significativos ($p<0,05$) – Teste Anova Two-way**Legenda:** DP = desvio padrão

Considerou-se que a metodologia desenvolvida para a avaliação da identificação de fonemas fricativos foi eficiente, fornecendo respostas confiáveis aos objetivos propostos nesta pesquisa. No entanto, o uso do computador e a necessidade de boa alfabetização podem limitar a sua utilização na prática clínica e pesquisas, uma vez que o Brasil ainda lida com problemas de escolaridade e de pouco acesso à computação, principalmente na população adulta.

Na prática audiológica, o uso do computador pode não ser uma boa estratégia, pois torna a avaliação longa e complicada para pacientes não familiarizados com essa tecnologia. No entanto, para fins de pesquisa, acredita-se que a tarefa de identificação de palavras, ao invés do reconhecimento, em que o indivíduo deve repetir as palavras ouvidas, evita viés na captação das respostas decorrentes de fala distorcida, ou na marcação adequada dos resultados, por parte do examinador.

Nos gráficos dos IPIFF médios para fala filtrada nas frequências de corte em 1,5 kHz, 2 kHz e 3 kHz (Figura 1) constatou-se que o melhor desempenho na tarefa proposta foi

dependente da maior faixa de frequências audível, ou seja, quanto mais alta a frequência de corte, melhor a identificação da fala, tanto para locutores do gênero masculino, quanto feminino. Porém, como previsto, apesar da frequência de corte em 3 kHz permitir um bom desempenho na identificação fonêmica ($>80\%$), houve piora acentuada na identificação de alguns fonemas, quando a compressão de frequências foi utilizada, pois, para perdas auditivas a partir de 3,5 kHz, a compressão de frequências não é indicada, uma vez que as frequências acima de 3,5 kHz não são absolutamente necessárias para a boa inteligibilidade da fala⁽¹⁶⁾.

Sendo assim, para se estudar a frequência de corte, a compressão de frequências e a interação destas duas variáveis (Tabela 1), desprezou-se a frequência de corte em 3 kHz, pois, quando o deficiente auditivo apresenta bons índices de reconhecimento da fala, provavelmente a compressão de frequências não melhorará o seu desempenho⁽¹⁶⁾.

A análise da variável “filtro” demonstrou que a frequência de corte em 2 kHz favoreceu a identificação fonêmica, exceto

a dos fricativos /f, v/ e /s, z/, que não apresentaram diferenças. Pode-se dizer que a variável produz efeito significativo no IPIFF médio, sendo que, de maneira geral, quanto maior a extensão da zona morta, pior os resultados na identificação dos fonemas fricativos. Isso reforça a ideia de que, quanto mais informações freqüenciais forem disponibilizadas, melhor será o IPIFF, mesmo utilizando compressão de freqüências (esta análise se refere apenas às freqüências de corte de 1,5 e 2 kHz).

Esses resultados já foram obtidos em estudos anteriores, que avaliaram a identificação de consoantes em deficientes auditivos com perdas de grau moderado a severo em freqüências altas, utilizando filtros passa-baixas, demonstrando, assim, que o reconhecimento de fala melhora com o aumento da faixa de freqüências disponível⁽²¹⁾. Em estudo que simulou, em normo-ouvintes, zonas mortas na cóclea, de diferentes tamanhos e regiões, quanto maior a extensão da zona morta, pior foi o reconhecimento da fala, principalmente se ocorria na faixa de freqüências de 1 kHz a 2 kHz⁽²⁰⁾.

Em relação à variável “processamento do sinal”, os resultados demonstraram melhora na identificação dos fonemas fricativos /ʃ, ʒ/ com o uso da compressão de freqüências. Isto quer dizer que o uso desse algoritmo produziu efeito positivo na identificação desse grupo fonêmico, quando se aplicou filtros em 1,5 kHz e 2 kHz. Para os demais fonemas avaliados (/f, v/ e /s, z/), não houve diferenças entre as duas variáveis. No entanto, devido à evidente melhora dos fonemas /ʃ, ʒ/, o IPIFF total foi influenciado, apresentando melhores resultados com o uso da compressão de freqüências.

Os resultados demonstraram que a compressão de freqüências auxiliou a percepção das fricativas /ʃ, ʒ/, pois, ao se aplicar filtros passa-baixas com freqüências de cortes em 1,5 kHz e 2 kHz, pistas necessárias à correta identificação destes sons foram perdidas. Dessa forma, como a energia destes fonemas está concentrada em freqüências mais altas (3 kHz a 4,5 kHz, aproximadamente), o algoritmo de compressão de freqüências trouxe essas informações para uma faixa de freqüências inferior, contribuindo para a identificação destes sons⁽⁶⁾.

Devido à distribuição espectral mais plana das fricativas /f, v/, o deslocamento do conteúdo espectral dessa região não afetou a percepção destas consoantes. Apesar da distribuição espectral média dos fonemas /s, z/ mostrar que já existem algumas pistas destas fricativas na faixa de 3 kHz a 4,5 kHz, elas ainda são insuficientes para auxiliar a identificação destes fonemas, quando há o deslocamento dessa faixa de freqüências para a área audível do espectro sonoro.

Esses resultados concordam com a literatura pesquisada^(9,22,23,24,25). Em um estudo, o uso de um algoritmo de transposição de freqüências em indivíduos deficientes auditivos com presença de zonas mortas na cóclea, a partir de 1,5 kHz, foi eficiente na identificação do fonema /ʃ/ e, de maneira geral, não houve prejuízo na identificação dos demais fonemas da língua inglesa. Nesse mesmo estudo, os autores demonstraram que o algoritmo aplicado foi eficaz para auxiliar a detecção

de fricativas finais /s, z/, ao comparar os resultados obtidos com a fala filtrada (controle)⁽⁹⁾. No Brasil, um estudo recente demonstrou que a compressão de freqüências em adultos propiciou melhora da audibilidade, da detecção dos fonemas /s, ʃ/ e do reconhecimento de fala, sendo a melhora progressiva ao longo de 12 meses⁽²²⁾.

O presente trabalho encontrou resultados que divergem da literatura supracitada. A identificação do /s/ final demonstrou piora na fala comprimida, mostrando a confusão do ouvinte quanto à detecção das fricativas finais. No entanto, os índices percentuais médios de acerto se mantiveram superiores a 84%.

Não houve interação entre as variáveis “processamento do sinal” (uso da compressão de freqüências) e “freqüências de corte” (extensão da zona morta), ou seja, ambas são independentes. Por isso, pode-se dizer que o efeito da compressão de freqüências é o mesmo nas duas zonas mortas simuladas, 1,5 kHz e 2 kHz.

CONCLUSÃO

O teste Índice Percentual de Identificação de Fonemas Fricativos (IPIFF), mostrou-se eficiente, confiável e prático para ser utilizado em pesquisas que pretendem avaliar a identificação de fonemas fricativos, por permitir uma avaliação duplocega, com taxa de amostragem significante e resultados fáceis de serem visualizados, devido ao processamento dos dados pelo próprio *software*. Além disso, permitiu comparar os resultados obtidos em diferentes situações de escuta, como no caso do uso de compressão de freqüências. Para a prática clínica, o teste se mostrou limitado, por exigir bom nível de alfabetização e familiaridade com a informática por parte dos pacientes. Ademais, o tempo médio de aplicação é de 60 minutos, o que pode dificultar seu uso clínico.

O teste, nas condições em que foi aplicado, demonstrou que a identificação dos fonemas, em geral, foi favorecida pelo aumento da freqüência de corte, tanto na fala filtrada, como na fala comprimida, exceto em 3 kHz, em que houve piora acentuada na identificação dos fonemas por compressão de freqüências. Para as freqüências de corte em 1,5 kHz e 2 kHz, a compressão de freqüência se mostrou eficiente para a identificação dos fonemas fricativos, principalmente do /ʃ, ʒ/. No entanto, a identificação do /s/ final foi comprometida pelo algoritmo.

REFERÊNCIAS

1. Moore BCJ. Dead regions in the cochlea: diagnosis, perceptual consequences and implications for the fitting of hearing aids. *Trends Amplif.* 2001;5(1):1-34. <http://dx.doi.org/10.1177/108471380100500102>
2. Vickers DA, Moore BCJ, Baer T. Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies. *J Acoust Soc Am.* 2001;110(2):1164-75. <http://dx.doi.org/10.1121/1.1381534>

3. Hicks BL, Braida LD, Durlach NI. Pitch invariant frequency lowering with non-uniform spectral compression. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing - ICASSP 81; 1981 Mar 30-Apr 1; Atlanta, Georgia. p. 121-4.
4. Muñoz CMA, Peggy BN, Rutledge JC, Gago A. Frequency lowering processing for listeners with significant hearing loss. Proceedings of the 6th IEEE International Conference on Electronics, Circuits and Systems - ICECS'99. 1999 Sep 5-8; Pafos, Cyprus. p. 741-4.
5. Wendy DE. Proportional frequency compression in hearing instruments. *Hear Rev*. 2001 Feb 4 [citado 2 mar 2008]. Disponível em: http://www.hearingreview.com/issues/articles/2001-02_04.asp
6. Silva FJF, Marotta AM. Rebaixamento de frequências para portadores de deficiência auditiva com perdas profundas em altas frequências. *Rev Fonoaudiol Brasil*. 2005;3:1-3.
7. Simpson A, Hersbach AA, McDermott HJ. Improvements in speech perception with an experimental nonlinear frequency compression hearing device. *Int J Audiol*. 2005;44(5):281-92. <http://dx.doi.org/10.1080/14992020500060636>
8. Simpson A, Hersbach AA, McDermott HJ. Frequency-compression outcomes in listeners with steeply sloping audiograms. *Int J Audiol*. 2006;45(11):619-29. <http://dx.doi.org/10.1080/14992020600825508>
9. Robinson JD, Baer T, Moore BCJ. Using transposition to improve consonant discrimination and detection for listeners with severe high-frequency hearing loss. *Int J Audiol*. 2007;46(6):293-308. <http://dx.doi.org/10.1080/14992020601188591>
10. Turner CW. The limits of high-frequency amplification. *Hear J*. 1999;52(2):10-4.
11. Manrique AM, Massone MI. Acoustic analysis and perception of Spanish fricative consonants. *J Acoust Soc Am*. 1981;69(4):1145-53.
12. Behlau M. Uma análise das vogais do português brasileiro falado em São Paulo: perceptual, espectrográfica de formantes e computadorizada de frequência fundamental [tese]. São Paulo: Universidade Federal de São Paulo; 1984.
13. Santos MTM. Uma análise espectrográfica dos sons fricativos surdos e sonoros do português brasileiro [monografia]. São Paulo: Universidade Federal de São Paulo; 1987.
14. Russo I, Behlau M. Percepção da fala: análise acústica do português brasileiro. São Paulo: Lovise; 1993.
15. Costa MJ. Desenvolvimento de sentenças em português: apresentação e estratégias de aplicação na audiolgia. Santa Maria: Pallotti; 1998.
16. Turner CW, Hurtig RR. Proportional frequency compression of speech for listeners with sensorineural hearing loss. *J Acoust Soc Am*. 1999;106(2):877-86. <http://dx.doi.org/10.1121/1.427103>
17. McDermott HJ, Dean MR. Speech perception with steeply sloping hearing loss: effects of frequency transposition. *Br J Audiol*. 2000;34(6):353-61. <http://dx.doi.org/10.3109/0300536400000151>
18. Silva FJF, Marotta AM. Frequency compression and frequency shifting for the hearing impaired. *Biomed Engineer*. 2004;2. <http://dx.doi.org/10.2316/Journal.216.2004.2.417-808>
19. Baskent D, Shannon RV. Speech recognition under conditions of frequency-place compression and expansion. *J Acoust Soc Am*. 2003;113(4):2064-76. <http://dx.doi.org/10.1121/1.1558357>
20. Baskent D, Shannon RV. Frequency transposition around dead regions simulated with a noiseband vocoder. *J Acoust Soc Am*. 2006;119(2):1156-63. <http://dx.doi.org/10.1121/1.2151825>
21. Simpson A, McDermott HJ, Dowell RC. Benefits of audibility for listeners with severe high-frequency hearing loss. *Hearing Res*. 2005;210(1-2):42-52. <http://dx.doi.org/10.1016/j.heares.2005.07.001>
22. Marchesin VC, Iório MCM. Estudo dos efeitos de longo prazo da compressão de frequências por meio de testes comportamentais verbais em adultos. *CoDAS*. 2015;27(1):37-43. <http://dx.doi.org/10.1590/2317-1782/20152014165>
23. Gresele ADP, Costa MJ, Garcia MV. Compressão de frequências no reconhecimento de fala de idosos com possíveis zonas mortas na cóclea. *Rev CEFAC*. 2015;17(1):223-37. <http://dx.doi.org/10.1590/1982-021620155414>
24. Robinson JD, Stainsby TH, Baer T, Moore BCJ. Evaluation of a frequency transposition algorithm using wearable hearing aids. *Int J Audiol*. 2009;48(6):384-93. <http://dx.doi.org/10.1080/14992020902803138>
25. Glista D, Scollie S, Bagatto M, Seewald R, Parsa V, Johnson A. Evaluation of nonlinear frequency compression: clinical outcomes. *Int J Audiol*. 2009;48(1):632-44. <http://dx.doi.org/10.1080/14992020902971349>