# Compressão de freqüências e suas implicações no reconhecimento de fala****

## Frequency compression and its effects in speech recognition

Letícia Pimenta Costa Spyer Prates*
Francisco José Fraga da Silva**
Maria Cecília Martinelli Iório***

*Fonoaudióloga. Doutoranda em Ciências pela Universidade Federal de São Paulo - Escola Paulista de Medicina. Fonoaudióloga do Hospital das Clínicas - Universidade Federal Minas Gerais. Endereço para correspondência: Av. André Cavalcanti, 381 - Apto. 204. Belo Horizonte - MG - CEP 30430-110 (lepcosta@hotmail.com).

**Engenheiro. Doutor em Engenharia Eletrônica e Computação pelo Instituto de Tecnologia da Aeronáutica. Professor Adjunto da Universidade Federal do ABC.

***Fonoaudióloga. Doutora em Distúrbios da Comunicação Humana pela Universidade Federal de São Paulo - Escola Paulista de Medicina. Professora Adjunta do Curso de Fonoaudiologia Universidade Federal de São Paulo - Escola Paulista de Medicina.

****Trabalho Realizado na Universidade Federal de São Paulo.

Artigo Original de Pesquisa

Artigo Submetido a Avaliação por Pares

Conflito de Interesse: não

**Abstract**
Background: frequency compression. Aim: to evaluate the index of speech recognition (IPRF) using frequency compression in three different ratios. Methods: monosyllabic words were recorded using an algorithm of frequency compression in three ratios: 1:1, 2:1, 3:1, generating three lists of words. Eighteen listeners accomplished the IPRF using the modified words. They were subdivided in two groups, considering familiarity with the speech material: group of audiologists (F) and group of patients (P). Results: a statistically significant decrease in accuray was observed when using frequency compression. Group F presented a better performance than Group P in all of the applied ratio frequency compression ratios. Conclusion: Frequency compression hinders speech recognition; as the compression ratio increases, so does the level of difficulty. Familiarity with the words facilitates recognition in any hearing condition.
**Key Words:** Hearing Aid; Hearing Loss; High-Frequency; Speech Discrimination Test.

**Resumo**
Tema: compressão de freqüências. Objetivo: avaliar o índice percentual de reconhecimento de fala (IPRF) utilizando compressão de freqüências em três razões diferentes. Métodos: palavras monossílabas foram gravadas utilizando um algoritmo de compressão de freqüências em três razões: 1:1, 2:1, 3:1, gerando três listas de palavras. Dezoito normo-ouvintes realizaram o IPRF utilizando as listas de palavras modificadas. Foram subdivididos em dois grupos, considerando a familiaridade com o material de fala gravado: grupo de fonoaudiólogas (F) e grupo de acompanhante de pacientes (P). Resultados: observou-se uma piora estatisticamente significante no IPRF quando se utilizou compressão de freqüências. O grupo F teve melhor desempenho que o grupo P em todas as razões de compressão aplicadas. Conclusão: a compressão de freqüências dificulta o reconhecimento da fala, sendo que, quanto maior a razão de compressão, maior é a dificuldade. A familiaridade com as palavras facilita o seu reconhecimento em qualquer condição de escuta.
**Palavras-Chave:** Auxiliares de Audição; Perda Auditiva de Alta Freqüência; Teste de Discriminação de Fala.

Recebido em 01.02.2008.
Revisado em 16.03.2008; 03.06.2008; 24.10.2008; 31.10.2008; 08.03.2009.
Aceito para Publicação em 04.05.2009.

Referenciar este material como:
Prates LPCS, Silva FJF, Iório MCM. Compressão de freqüências e suas implicações no reconhecimento de fala. Pró-Fono Revista de Atualização Científica. 2009 abr-jun;21(2):149-54.

## Introdução

É consenso que a maior dificuldade relacionada à deficiência auditiva se refere à comunicação, com a perda na habilidade de discriminação e reconhecimento de fala. Entretanto, nem sempre o aumento da informação acústica disponível por meio das próteses auditivas proporciona o completo restabelecimento destas habilidades. Alguns pacientes apresentam pouco ou nenhum benefício com a amplificação, particularmente os indivíduos com perda auditiva acentuada em altas freqüências[1].

Diversos estudos demonstram a contribuição das freqüências altas para a inteligibilidade de fala. Consequentemente, a deficiência auditiva neurossensorial descendente é relacionada à dificuldade para compreender fala, mesmo com o uso das próteses auditivas. Conforme aumenta o grau da perda auditiva, algumas freqüências não contribuem ou até mesmo reduzem a informação disponível em outras freqüências preservadas, como ocorre na presença de zonas mortas na cóclea[2]. De acordo com o estudo, a presença de zonas mortas na cóclea, isto é, regiões que não apresentam células ciliadas internas e/ou neurônios adjacentes funcionais, pode explicar as dificuldades observadas na adaptação de próteses auditivas. A amplificação de sons na faixa de freqüência correspondente às zonas mortas não resulta em benefício e pode até mesmo prejudicar a inteligibilidade da fala. Por isso, alguns autores recomendam cautela na amplificação de altas freqüências com limiar auditivo superior a 55dB N[3-4].

Nestes casos, uma saída pode ser a utilização de próteses auditivas com compressão de freqüências, que alteram os componentes de freqüências altas em freqüências baixas, nas quais o aproveitamento da função auditiva pode ser mais efetivo[5]. Dessa forma, o espectro sonoro é reduzido em uma faixa mais estreita, sendo percebido de maneira distorcida, porém preservando-se a distribuição das ondas sonoras e suas inter-relações na mensagem ouvida.

Reprodução da fala a uma taxa de amostragem mais lenta, ou redução da taxa de cruzamentos por zero são alguns dos métodos de rebaixamento de freqüências que têm sido empregados nas últimas décadas[6]. Todos esses métodos envolvem algum tipo de distorção do sinal de fala, mais ou menos perceptível, geralmente dependente do grau de alteração espectral realizada. Muitos dos esquemas de rebaixamento de freqüências têm alterado perceptivelmente importantes características da fala, como padrões rítmicos e temporais, *pitch* e duração de elementos segmentais.

O uso de curvas de compressão de freqüências foi sugerido em importantes investigações sobre rebaixamento de freqüências[6]. Esta técnica envolve a compressão monotônica do espectro de tempo curto, sem alteração do pitch e ao mesmo tempo evitando alguns dos problemas observados em outros métodos.

O presente estudo se propõe a desenvolver e avaliar o algoritmo de compressão de freqüências descrito em um estudo anterior[6], com algumas alterações. Trata-se de um estudo piloto, onde o algoritmo alterado foi aplicado em uma lista de palavras monossílabas para serem reconhecidas e repetidas por ouvintes normais, considerando a razão de compressão aplicada (3:1, 2:1, 1:1), para posterior estudo em indivíduos deficientes auditivos.

O objetivo deste trabalho foi fazer uma análise descritiva dos resultados encontrados em indivíduos normais, considerando a razão de compressão aplicado e a familiaridade com as palavras do teste.

## Método

Esta pesquisa foi realizada no Núcleo Integrado de Atendimento, Pesquisa e Ensino em Audição (Niapea) da Universidade Federal de São Paulo - Escola Paulista de Medicina após aprovação do Comitê de Ética em Pesquisa da Universidade Federal de São Paulo - Hospital São Paulo, sob o protocolo 0150/07 e assinatura do Termo de Consentimento Livre e Esclarecido, por todos os indivíduos da amostra.

Participaram deste estudo 18 normo-ouvintes de ambos os sexos e idades compreendidas entre 21 e 42 anos, sendo que destes, oito eram fonoaudiólogos familiarizados com a lista de palavras contidas no teste aplicado. Os outros dez ouvintes eram acompanhantes de pacientes atendidos no ambulatório, sem qualquer conhecimento prévio das palavras contidas na lista. Dessa forma, definiram-se dois grupos: F, formado pelos fonoaudiólogos e P, formado pelos demais ouvintes.

Os participantes apresentavam limiares auditivos melhores que 20dB NA nas freqüências de 250 a 8.000Hz, aferidos antes do início da avaliação.

O material de fala utilizado neste estudo foi constituído por palavras monossílabas aplicadas por meio de fones TDH 39, na intensidade de 60dB NA, no silêncio, em tarefa monótica, em ambas as orelhas. Os indivíduos foram orientados a repetir, com exatidão, os monossílabos apresentados. O índice percentual de reconhecimento de fala (IPRF) foi estabelecido contando-se as palavras repetidas corretamente.

Utilizou-se, para a pesquisa do IPRF, a lista de 25 monossílabos, foneticamente balanceados[7], disponíveis em CD[8]. Uma nova organização desta lista de palavras foi reproduzida em outro CD em três seqüências diferentes das mesmas palavras, para reduzir o aprendizado do ouvinte.

Para a pesquisa dos limiares tonais e dos testes de fala foi utilizado o *hardware* do sistema Aurical da marca *Madsen Eletronics*, acoplado a um computador de processador *Pentium*, onde foi selecionado o audiômetro Aurical (*Aurical Audiometer*). Os procedimentos de fala foram aplicados em uma cabina acústica utilizando um compact disc player portátil, modelo 4147 da marca Toshiba, acoplado ao hardware do sistema Aurical e fones TDH 39, além do CD contendo as amostras de fala.

As listas de palavras tiveram o espectro do sinal de fala modificado por compressão de freqüências, isto é, um rebaixamento executado por um algoritmo de compressão do espectro de tempo curto do sinal de fala, provocando uma distorção sonora, porém sem perda significativa de informações do espectro de freqüências.

O processamento dos sinais de fala utilizados neste trabalho foi implantado pelo *software Matlab*, no Centro de Engenharia e Modelagem da Universidade Federal do ABC, pelo engenheiro responsável participante deste estudo. Para isso foram necessários a gravação das amostras de fala em CD e um computador para montagem do material de fala processado.

A compressão de freqüências foi executada pelo método não-linear, ou seja, realizando menor compressão nas baixas freqüências e comprimindo mais as altas freqüências[6]. A taxa de amostragem utilizada para a digitalização do sinal de fala foi de 16kHz.

Foram utilizados três razões de compressão (ou fator de compressão K) nas listas de palavras: 1:1 (K=1), 2:1 (K=2) e 3:1 (K=3); compondo assim três listas de palavras geradas pelo processamento de compressão do espectro de freqüências do sinal de fala digitalizado.

A razão de compressão 1:1 (ou o fator de compressão K = 1) se refere à ausência de compressão, ou seja, as palavras foram apresentadas de forma natural, oferecendo todo o espectro da fala contido no sinal amostrado à taxa de 16kHz.

As razões de compressão 2:1 e 3:1 a fatores de compressão (K = 2 e K = 3) significam aplicação da compressão de freqüência em diferentes proporções. Quanto maior a razão de compressão, maior o grau do rebaixamento de freqüências, o que gera maior alteração no espectro da fala.

As curvas de compressão de freqüências utilizadas neste trabalho podem ser visualizadas na Figura 1. Estas curvas foram implantadas computacionalmente por meio da equação mostrada no canto inferior direito da figura, onde a variável a controla o grau de não-linearidade das curvas (a = 0 transforma a curva em uma reta). A ausência total de compressão corresponde a K = 1 e a = 0. Quando a = 0 e K = 2, por exemplo, a compressão é linear (a = 0) na razão de 2:1 (K = 2). Isso significa, neste exemplo, que as freqüências de saída (do sinal processado) correspondem exatamente à metade do valor das freqüências de entrada (do sinal original). Ou seja, se o sinal original possuir uma componente de freqüência em 2000Hz, esta corresponderá à 1000Hz no sinal processado.

No algoritmo originalmente proposto[6], as curvas eram aproximadamente lineares (ausência de compressão) na faixa de 0 a 1kHz. Neste estudo, a faixa de linearidade aproximada foi estendida até 1,5kHz, visando alterar o menos possível a distorção perceptual das formantes e do pitch do sinal de fala original.

A Figura 2 ilustra os espectrogramas do monossílabo "jaz" nas três situações avaliadas nesta pesquisa: K = 1 e a = 0 (i); K = 2 e a = 0,3833 (ii); K = 3 e a = 0,6 (iii). Também é apresentada uma quarta situação, não avaliada neste estudo, que corresponde à compressão linear, com K = 2 e a = 0 (iv). Comparando a Figura 2 - ii e 2 - iv, pode-se visualizar claramente a diferença entre os espectrogramas obtidos com a compressão não-linear e linear .

As listas de palavras foram ouvidas em ordem decrescente de dificuldade, iniciando pela lista com K = 3 e terminado com K = 1, para não oferecer pistas facilitadoras para o reconhecimento das palavras ouvidas, já que as listas são compostas pelas mesmas palavras ordenadas de forma diferente.

Os resultados foram tratados estatisticamente, por meio dos testes não paramétricos de Wilcoxon e Mann-Whitney. Para complementação da análise descritiva, calculou-se o Intervalo de Confiança para média. Estabeleceu-se o nível de significância em 5%. Quando a análise estatística calculada apresentou significância, usamos um asterisco (*) para caracterizá-la.

## Resultados

Na Tabela 1 analisaram-se os valores médios do IPRF obtidos nas razões de compressão 3:1 (K=3), 2:1 (K=2) e 1:1(k=1), nos grupos de fonoaudiólogas (F) e de acompanhantes de pacientes (P) e compararam-se os resultados entre as orelhas direita e esquerda.

Como não foram encontradas diferenças estatisticamente significantes no IPRF obtido entre a orelha direita e esquerda nos dois grupos estudados, optou-se por realizar as demais análises considerando os valores de ambas as orelhas. Dessa forma, a taxa de amostragem duplicou-se, tornando os resultados mais fidedignos.

Na Figura 3 encontram-se os valores médios do IPRF obtidos nos grupos P e F, considerando a razão de compressão ou fator de compressão (K).

## Discussão

O estudo do reconhecimento da fala utilizando a compressão de freqüências já foi proposto por muitos autores em trabalhos que datam desde a década de 70, ou antes. O que difere estes estudos é a forma como o algoritmo é processado. No entanto, apesar dos resultados divergentes e muitas vezes desanimadores, ainda nos dias de hoje, muitos pesquisadores apostam neste algoritmo como uma saída na melhora efetiva do reconhecimento de fala, principalmente para os deficientes auditivos com perdas em altas freqüências. Com a descoberta das zonas mortas na cóclea[2], e os sucessivos estudos demonstrando seu impacto negativo na habilidade de reconhecimento de palavras[4], o estudo da compressão de freqüências volta nos dias atuais com uma proposta revigorada que, ao dispor de toda tecnologia de amplificação sonora, promete ser uma saída na melhora efetiva da discriminação de fala dos deficientes auditivos com presença de zonas mortas na cóclea.

FIGURA 1. Curvas de compressão de freqüências utilizadas no processamento do sinal de fala. No eixo horizontal está representada a faixa de freqüências de entrada (do sinal original) e no eixo vertical, a faixa de freqüências de saída (do sinal processado) para os fatores de compressão K = 2 (linha cheia) e K = 3 (linha tracejada).

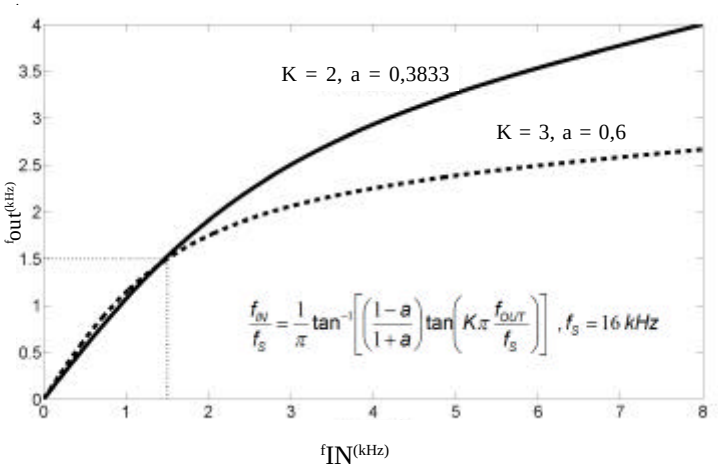



FIGURA 2. Espectrogramas da elocução "jaz": (i) original (K = 1 e *a* = 0); (ii) compressão não-linear com K = 2 e *a* = 0,3833; (iii) compressão não-linear com K = 2 e *a* = 0,6; e (iv) compressão linear, com K = 2 e *a* = 0 (situação não avaliada neste estudo).

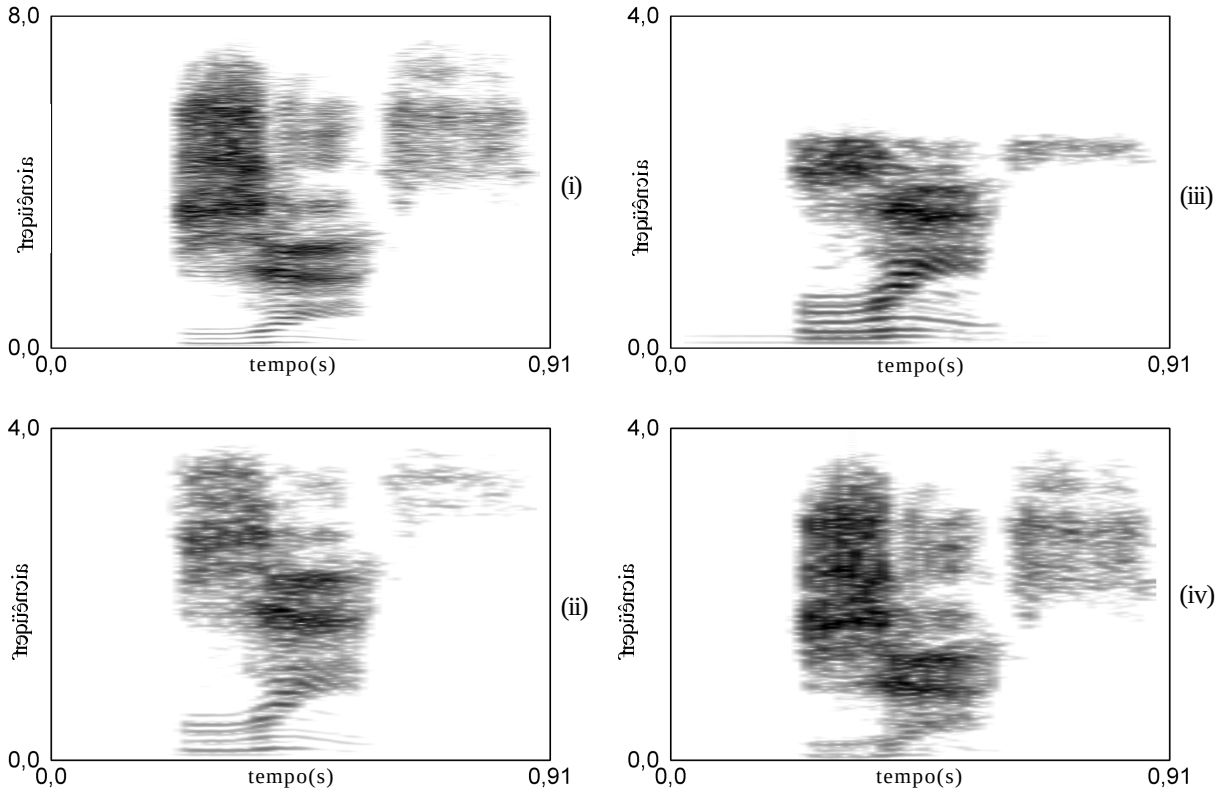
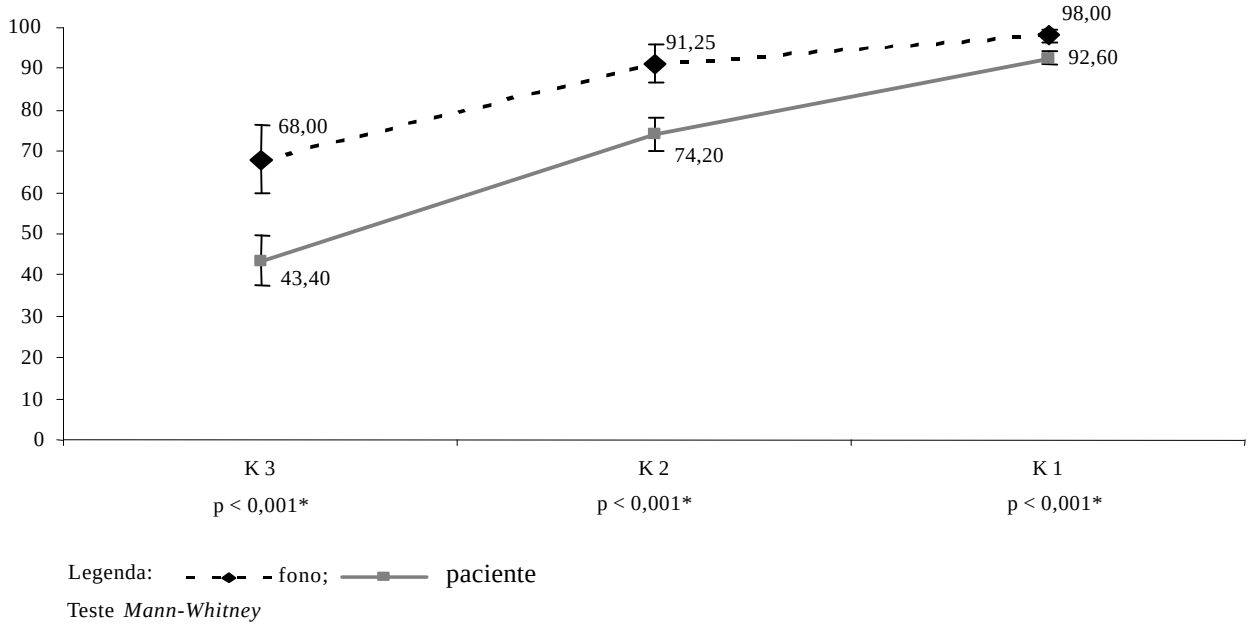


TABELA 1. Análise descritiva do IPRF encontrado no grupo de fonoaudiólogas (F) e acompanhantes de pacientes (P), com fatores de compressão (K) 3, 2 e 1, para orelha direita e esquerda.

| | Grupo P | | | | | | Grupo F | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | K3 | | K2 | | K1 | | K 3 | | K 2 | | K 1 | |
| | OD | OE | OD | OE | OD | OE | OD | OE | OD | OE | OD | OE |
| média | 42,40 | 44,40 | 74,00 | 74,40 | 92,00 | 93,20 | 68,50 | 67,50 | 91,50 | 91,00 | 98,00 | 98,00 |
| mediana | 42 | 42 | 74 | 76 | 92 | 94 | 78 | 74 | 94 | 96 | 100 | 100 |
| desvio padrão | 15,23 | 12,57 | 8,27 | 10,70 | 3,27 | 4,24 | 19,18 | 16,06 | 9,90 | 9,50 | 3,02 | 3,02 |
| CV | 35,9% | 28,3% | 11,2% | 14,4% | 3,5% | 4,5% | 28,0% | 23,8% | 10,8% | 10,4% | 3,1% | 3,1% |
| Q1 | 32 | 37 | 69 | 65 | 89 | 89 | 53 | 54 | 87 | 83 | 96 | 96 |
| Q3 | 51 | 44 | 76 | 80 | 95 | 96 | 80 | 77 | 100 | 97 | 100 | 100 |
| N | 10 | 10 | 10 | 10 | 10 | 10 | 8 | 8 | 8 | 8 | 8 | 8 |
| IC | 9,44 | 7,79 | 5,13 | 6,63 | 2,02 | 2,63 | 13,29 | 11,13 | 6,86 | 6,58 | 2,10 | 2,10 |
| p-valor | 0,509 | | 0,862 | | 0,180 | | 0,480 | | 0,655 | | 1,000 | |

Teste de Wilcoxon. Legenda: CV = coeficiente de variação; Q1 = primeiro quartil; Q3 = terceiro quartil; N = número da amostra; IC = intervalo de confiança.

FIGURA 3. Gráfico comparativo dos valores médios de IPRF obtidos no grupo P e F, considerando o fator de compressão (K) 3, 2 e 1.




A proposta deste estudo foi desenvolver um algoritmo de compressão de freqüências, e avaliar, em indivíduos normais, o reconhecimento de palavras utilizando este algoritmo. Com o objetivo de se fazer um estudo piloto, utilizou-se a compressão de freqüências em três razões distintas: 3:1 (K = 3), 2:1 (K = 2) e 1:1 (K = 1), alterando o grau de distorção das palavras gravadas. Além disso, avaliou-se se a familiaridade com as palavras do teste facilitaria o seu reconhecimento.

Como resultado, encontrou-se pior desempenho no teste de reconhecimento de palavras quanto maior a razão de compressão, em todos os grupos avaliados. A Figura 3 demonstrou que o grupo F teve maior facilidade no reconhecimento das palavras em todas as razões de compressão avaliadas ($p < 0,001$).

Para K = 2, foi possível atingir um índice de reconhecimento de palavras médio de 91,25% no grupo F, o que pode ser considerado um resultado excelente. Já no grupo P, nesta mesma razão de compressão, o índice percentual de reconhecimento de palavras foi de 74,2%, estatisticamente inferior ao grupo F ($p < 0,001$). Por este resultado, pode-se dizer que a familiaridade com as palavras do teste facilitou o seu reconhecimento em todas as razões de compressão estudadas. Isso nos leva a crer que o treino prévio utilizando este algoritmo pode ser uma saída para melhorar o reconhecimento de palavras.

Ainda na Figura 3, pode-se notar pelas linhas ascendentes uma melhora gradual no reconhecimento das palavras à medida que se diminui o fator de compressão. Essa tendência pôde ser observada nos dois grupos.

Um estudo[9] realizado em ouvintes normais utilizando um algoritmo de compressão de freqüências demonstrou que razões de compressão superiores ou iguais a 1.43:1 (ou seja, K < 1.43) não alteraram o desempenho no reconhecimento da fala. No entanto, os autores pesquisaram somente as razões de compressão de 2:1 (K = 2.0), 1.66:1 (K = 1.66), 1.43:1 (K = 1.43), 1.25:1 (K = 1.25) e 1.11:1 (K = 1.11), que na sua maioria são bem menores que as utilizadas neste estudo, gerando menor distorção do sinal de fala. Além disso, na presente pesquisa, a compressão utilizada foi do tipo não-linear, ao passo que, no estudo referido[9], usou-se apenas a compressão linear.

Outros autores[9-11] concluíram em seus estudos que os algoritmos de rebaixamento de freqüências devem ser implementados com cautela para não haver degradação do sinal de fala. Os autores acreditam que o treinamento prévio com o algoritmo facilitaria o reconhecimento das palavras, pois os pacientes aprendem a escutar as novas pistas de fala. Ao contrário, os efeitos instantâneos de distorção do espectro da fala provocados pelo rebaixamento de freqüência são mais maléficos aos ouvintes normais comparados aos pacientes reais, já que estes não estão acostumados ao sinal de fala degradado.

A idéia de se fazer um estudo piloto permitiu avaliar as variáveis que poderiam influenciar o teste aplicado em deficientes auditivos. Pretende-se, futuramente, continuar este estudo aplicando a compressão de freqüências em deficientes auditivos com presença de zonas mortas na cóclea. Por se tratar de um estudo piloto, pode-se e deve-se questionar a metodologia aplicada. Acredita-se que as razões de compressão de freqüências aplicadas podem ter sido muito altas e, portando, seria importante estudar razões de compressão menores, que promovam menos distorções no sinal de fala, como sugerem outros autores[9].

Além disso, acredita-se ser necessário criar um material de fala mais adequado à proposta deste estudo, com uma amostra de fala maior, utilizando gravações com locutores do sexo masculino e feminino[10]. Também seria importante obter uma freqüência de apresentação dos fonemas suficientemente grande para analisar o reconhecimento de cada grupo fonêmico isoladamente[11]. Isso permitiria estudar o comportamento da compressão de freqüências para cada som em particular e precisar os benefícios e malefícios deste algoritmo para o reconhecimento das palavras, em função de cada grupo fonêmico analisado separadamente.

## Conclusão

1. As razões de compressão de freqüências 2:1 e 3:1 dificultam o reconhecimento de fala em ouvintes normais.
2. Quanto maior a razão de compressão de freqüência pior o reconhecimento da fala.
3. A familiaridade com as palavras ouvidas facilita o seu reconhecimento, mesmo quando estas palavras estão distorcidas por compressão de freqüências.

## Referências Bibliográficas


1. Ching TYC, Dillon H, Katsh R, Byrne D. Maximizing effective audibility in hearing aid fitting. Ear Hear 2001;22(3):212-24.

2. Moore BCJ, Huss M, Vickers DA, Glasberg BR, Alcantara JI. A test for the diagnosis of dead regions in the cochlea. Br J Audiol. 2000;34:205-24.

3. Baer T, Moore BC, Kluk K. Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies. J Acoust Soc Am. 2002;112:1133-44.

4. Gordo A, Iorio MCM. Zonas mortas na cóclea em freqüências altas: implicações no processo de adaptação de prótese auditivas. Rev. Bras. Otorrinolaringol. 2007 May-June 73(3):299-307.

5. Vickers DA, Moore BCJ, Baer T. Effects of low-pass filtering on the intelligibility of speech in quiet for people with dead regions at high frequencies. J Acoust Soc Am. 2001;110(2):1164-75.

6. 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;6:121-4.

7. Pen M, Mangabeira-Albernaz PL Desenvolvimento de testes para logoaudiometria - discriminação vocal. In: Congresso Pan-americano de Otorrinolaringologia y Broncoesofagia. Anales. Lima - Peru; 1973. p. 223-6.

8. Pereira LD, Schochat E. Manual de avaliação do processamento auditivo central. São Paulo: Lovise; 1997.

9. Turner CW, Hurtig RR. Proportional frequency compression of speech for listeners with sensorineural hearing loss.J Acoust Soc Am 1999;106(2):877-86.

10. Baskent D, Shannon RV. Frequency transposition around dead regions simulated with a noise band vocoder. J Acoust Soc Am. 2006;119(2):1156-63.

11. Simpson A, Hersbach AA, McDermott HJ. Improvements in speech perception with na experimental nonlinear frequency compression hearing device. Int J Audiol. 2005;44(5):281-92.