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.

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.

Artigo Original de Pesquisa

Artigo Submetido a Avaliação por Pares

Conflito de Interesse: não

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. Frequency compression and its effects in speech recognition (original title: Compressão de freqüências e suas implicações no reconhecimento de fala). Pró-Fono Revista de Atualização Científica. 2009 abr-jun;21(2):149-54.

Introduction

There is a consensus that the main difficulty related to hearing loss refers to communication, with the loss in the ability of speech discrimination and recognition. However, the increase on acoustic information available through hearing aids does not always provide the complete restoration of these abilities. Some patients present little or no benefit with amplification, particularly those with severe high frequencies hearing loss (1).

Several studies demonstrate the contribution of high frequencies on speech intelligibility. Consequently, the sloping sensorineural hearing loss is related to the difficulty in understanding speech, even with the use of hearing aids. As the degree of the hearing loss increases, some frequencies do not contribute or even reduce the information available in other preserved frequencies - as it occurs in the presence of cochlea dead regions, described in 2000 (2). According to that study, the presence of dead regions in the cochlea - i.e. regions that have no inner hair cells and/or adjacent functional neurons - may explain the difficulties in hearing aids adaptation. The amplification of sounds in the frequency range that corresponds to the dead regions does not result in benefit and may even impair speech intelligibility. Therefore, some authors recommend caution in high frequencies amplification when hearing threshold is above 55 dB HL (3,4).

In such cases, a solution may be the use of hearing aids with frequency compression which modifies the components of high frequencies into low frequencies, where the recovery of auditory function may be more effective (5). Thus, the spectrum is reduced to a narrower range, being perceived as distorted, however, maintaining the distribution of sound waves and their inter-relationships in the message heard.

Speech reproduction onto a slower sampling rate, or reduced rate by zero crossings, are some methods of frequency lowering that have been employed in the recent decades (6). All these methods involve some kind of distortion of the speech signal, more or less noticeable, usually dependent on the degree of spectral change made. Many schemes of frequency lowering have perceptibly modified important speech characteristics such as rhythmic and temporal patterns, pitch and duration of segmental elements.

The use of frequency compression curves was suggested in an important research on frequency lowering (6). This technique involves monotonic compression of the short time spectrum, without changing the pitch while avoiding some of the problems seen in other methods.

The present study aims to develop and evaluate the frequency compression algorithm described in previous study (6), with some modifications. This is a pilot study where the modified algorithm was applied to a list of monosyllabic words to be recognized and replicated by normal hearing subjects, considering the compression ratio applied (3:1, 2:1, 1:1) for a subsequent study in deaf individuals.

The purpose of this study was to conduct a descriptive analysis of results in normal subjects, considering the compression ratio applied and the familiarity with the words of the test.

Method

This study was conducted at the Department of Integrated Care, Research and Teaching on Hearing (NIAPEA), Federal University of São Paulo - Paulista School of Medicine, after approval by the Research Ethics Committee of the Federal University of São Paulo / Hospital São Paulo, under the protocol 0150/07. All participants signed the free and informed consent form.

The study included 18 normal listeners of both genders, with ages between 21 and 42 years. Of the participants, eight were Speech-Language Pathologists and Audiologists and were familiar with the list of words contained in the applied test. The other ten participants were companions of patients of the clinic, without any prior knowledge of the words on the list. Thus, two groups were defined: F composed by Speech-Language Pathologists and Audiologists and P, composed by the remaining participants.

The participants had hearing thresholds better than 20 dB in the frequencies from 250 to 8000 Hz, measured before the beginning of the evaluation.

The speech material used in this study consisted of monosyllabic words applied through TDH 39 headphones at 60 dB NA intensity, in silence, in monotic task in both ears. The subjects were instructed to repeat, exactly, the monosyllables presented. The word recognition test (WRT) was established by counting the number of words repeated correctly.

For the WRT, the list of 25 monosyllabic words, phonetically balanced (7) and available on CD (8) was used. A new organization of the list of words was played in another CD in three different sequences of the same words, to reduce the listener learning effect.

To determine the pure tone and speech tests thresholds, the hardware of Aurical system from Madsen Electronics was used, coupled to a Pentium computer, from which the audiometer Aurical (Aurical Audiometer) was selected. The speech procedures

were applied in a sound proof booth using a portable compact disc player, model 4147 from Toshiba, coupled to the hardware system Aurical and TDH 39 headphones, besides the CD containing speech samples.

The lists of words had the spectrum of the speech signal modified by compression of frequencies. That is, an algorithm executed a lowering spectrum compression of short duration of the speech signal, causing a sound distortion but without significant loss of information of the spectrum of frequencies.

The processing of speech signals used in this study was implemented by the software Matlab at the Engineering and Modeling Center of the ABC Federal University by the Engineer responsible for this study. For such, the recording of samples of speech on CD and a computer to assemble the processed speech material were necessary.

The frequency compression was performed by non-linear method - i.e. performing smaller compression in low frequencies and further compression on high frequencies (6). The sampling rate used for the digitization of the speech signal was 16 kHz.

Three compression ratios were used (or compression factor - K) in the lists of words: 1:1 (K = 1), 2:1 (K = 2) and 3:1 (K = 3), thus composing three lists of words generated by frequency spectrum compression process of the digitized speech signal.

The compression ratio of 1:1 (or the compression factor K = 1) refers to the absence of compression, i.e. the words were presented in a natural form, providing the whole spectrum of speech in the signal sampled at a rate of 16 kHz.

The compression ratios of 2:1 and 3:1 (or compression factor K = 2 and K = 3) mean frequency of application of compression in different proportions. The higher the compression ratio is, the greater the degree of frequency lowering - which creates major change on the speech spectrum.

The frequency compression curves used in this study can be observed in Figure 1. These curves were implemented computationally using the equation shown in the lower right corner of the figure, where the variable a controls the degree of nonlinearity of the curves (a = 0 turns the curve into a straight line). The total lack of compression corresponds to K = 1 and a = 0. When a = 0 and K = 2 for example, the compression is linear (a = 0) in the ratio of 2:1 (K = 2). This means, in this example, that the output frequency (processed signal) correspond exactly to half the value of the input frequency (the original signal). That is, if the original signal has a frequency component at 2000 Hz, this will correspond to 1000 Hz in the processed signal.

On the algorithm originally proposed (6), the curves were approximately linear (no compression) in the range from 0 to 1 kHz. In this study, the approximate range of linearity was extended up to 1.5 kHz, aiming to change as least as possible the perceptual distortion of formants and the pitch of the original speech signal.

Figure 2 displays the spectrogram of the monosyllable "jaz" in three situations evaluated in this study: K = 1 and a = 0 (i); K = 2 and a = 0.3833 (ii); K = 3 and a = 0.6 (iii). A fourth situation - not evaluated in this study - which corresponds to the linear compression, with K = 2 and a = 0 (iv) is also presented. Comparing the Figures 2-ii to 2 - iv, we can clearly observe the difference between the spectrogram obtained with the non-linear and the linear compression.

The lists of words were heard in order of difficulty, starting the list with K = 3 and ending with K = 1, in order for not to provide clues that facilitate the recognition of words - once the lists are composed by the same words arranged in different forms.

The results were treated statistically through the Wilcoxon and Mann-Whitney non-parametric tests. To complement the descriptive analysis, confidence intervals of the means were calculated. The significance level adopted was 5%. We use an asterisk (*) to characterize statistical significance.

Results

In Table 1, the mean values obtained in the WRT in compression ratios of 3:1 (K = 3), 2:1 (K = 2) and 1:1 (K = 1) are analyzed for groups of Speech-Language Pathologists and Audiologists (F) and companions of patients (P). Results of right and left ears are compared.

As there were no statistically significant differences between the WRT obtained for the right and the left ear in both groups, we chose to perform the remaining analysis considering the values of both ears. Thus, the sampling rate is doubled, making the results more reliable.

In Figure 3 the average values of WRT obtained in groups P and F can be observed, considering the compression ratio or the compression factor (K).

Discussion

The study of speech recognition using compression of frequencies has been proposed by many authors in studies dating from the 70's or earlier. What differs among these studies is how the algorithm is processed. However, despite the divergent and often disappointing results, even

today, many researchers focus on the same algorithm as an effective improvement of speech recognition, especially for the hearing impaired with losses at high frequencies. With the discovery of dead regions in the cochlea (2), and successive studies demonstrating its negative impact on the ability of recognition of words (4), the frequency compression is again investigated with a reinvigorated proposal that, having all technology for sound amplification available, seems to be an effective outcome in improving speech discrimination of the hearing impaired with presence of cochlea dead regions.

The purpose of this study was to develop an algorithm for compression of frequencies, and to assess, in normal individuals, the recognition of words using this algorithm. With the aim of conducting a pilot study, the compression of frequencies in three distinct ratios was used: 3:1 (K = 3), 2:1 (K = 2) and 1:1 (K = 1), changing the degree of distortion of the recorded words. Furthermore, it was also evaluated whether the familiarity with the words of the test facilitated their recognition.

FIGURE 1. Frequency compression curves used in the processing of the speech signal. In the horizontal axis depicts the range of frequencies of input (the original signal) and in the vertical axis, the range of frequencies of output (processed signal) for the compression factors K = 2 (full line) and K = 3 (dashed line).

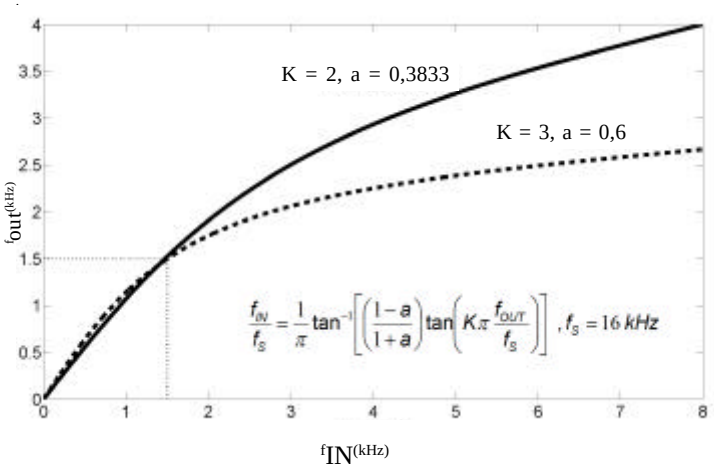


FIGURE 2: Spectrograms of utterance "jaz": (i) original (K = 1 and a = 0); (ii) non-linear compression with K = 2 and a = 0.3833; (iii) non-linear compression with K = 3 and a = 0.6; and (iv) linear compression, with K = 2 and a = 0 (situation not assessed in this study).

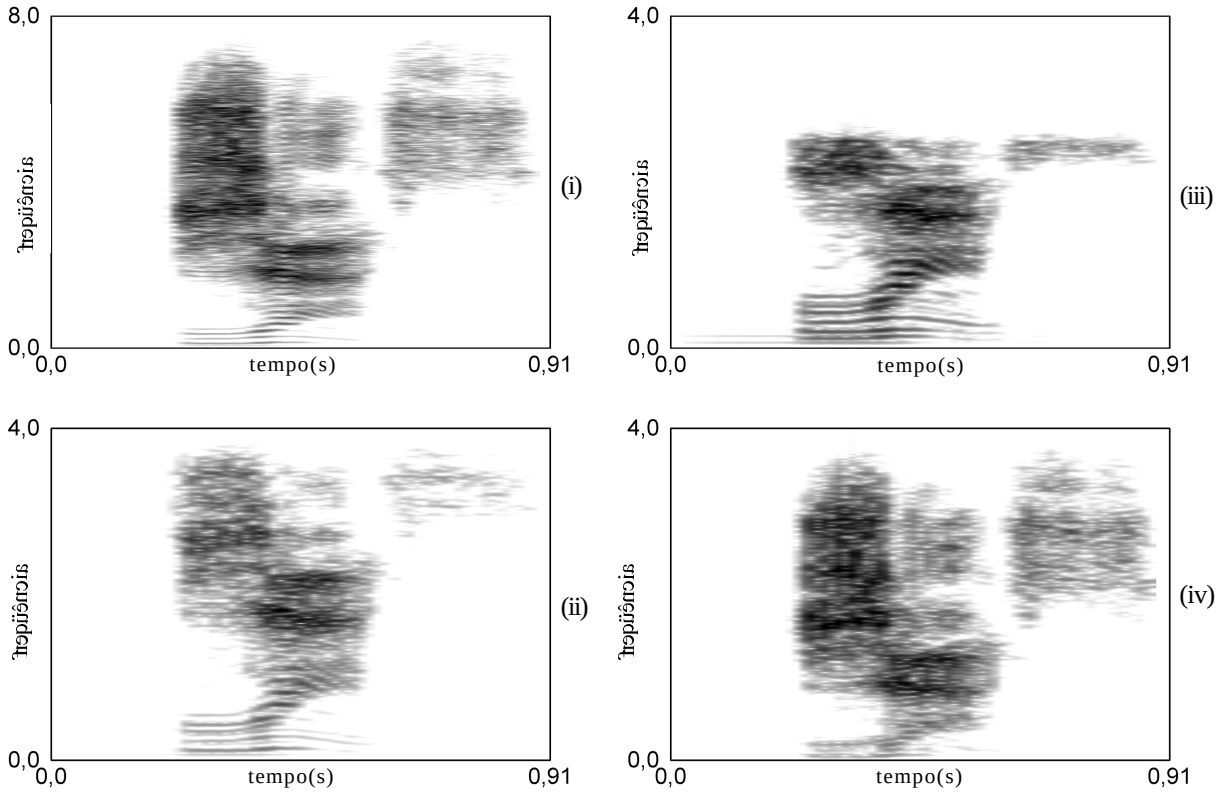
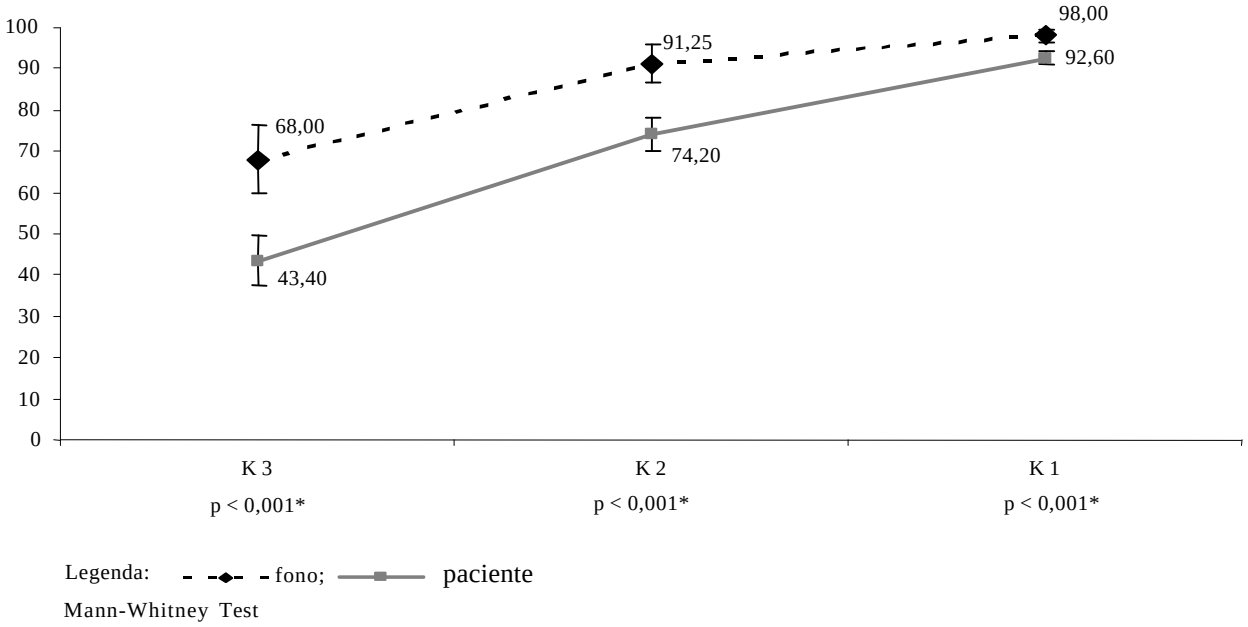


TABLE 1. Descriptive analysis of WRT results of the group of Speech-Language Pathologists and Audiologists (F) and companions of patients (P) with compression factors (K) 3, 2 and 1 for right and left ear.

	Group P						Group F					
	K3		K2		K1		K 3		K 2		K 1	
	RE	LE	RE	LE	RE	LE	RE	LE	RE	LE	RE	LE
Mean	42,40	44,40	74,00	74,40	92,00	93,20	68,50	67,50	91,50	91,00	98,00	98,00
Median	42	42	74	76	92	94	78	74	94	96	100	100
SD	15,23	12,57	8,27	10,70	3,27	4,24	19,18	16,06	9,90	9,50	3,02	3,02
VC	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
CI	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-value	0,509		0,862		0,180		0,480		0,655		1,000	

Wilcoxon Test Note: SD: Standard Deviation; VC: variation coefficient; Q1: first quartile; Q2: second quartile; N: sample size; CI: confidence interval

FIGURE 3. Comparative graphic of average WRT values obtained in groups P and F, considering the compression factor (K) 3, 2 and 1.



As result, we found poorer performance on tests of word recognition the higher the compression ratio was in all groups evaluated. Figure 3 shows that the group F presented better performance in the recognition of words in all the compression ratios evaluated (p <0001).

For K = 2, it was possible to achieve a mean rate of word recognition of 91.25% in group F, which can be considered an excellent performance. However, in group P, at the same compression ratio, the percentage rate of word recognition was 74.2%, statistically lower than the F group (p <0001). By this result, we can state that familiarity with the words of the test facilitated their recognition at all compression ratios studied. This leads us to believe that the prior training using this algorithm can be a way to improve the recognition of words.

Still in Figure 3, it can be noticed by the crescent lines a gradual improvement in the recognition of words as the compression ratio decreases. This trend could be observed for both groups.

A study (9) conducted with normal hearing participants using frequency compression algorithm showed that compression ratios equal or greater than 1:43: 1 (i.e. K< 1.43) did not alter the performance in speech recognition. However, the authors investigated only the compression ratios of 2:1 (K = 2), 1.66:1 (K = 1.66), 1.43:1 (K = 1.43),

1.25:1 (K = 1.25) and 1:11: 1 (K = 1.11), which are much smaller than those used in this study, indicating less distortion of the speech signal. Moreover, in the present study, the compression used was non-linear, while this study (9) used only the linear compression.

Other authors (9,10,11) concluded in their studies that the algorithms of frequency lowering should be implemented cautiously for the non degradation of the speech signal. The authors believe that prior training with the algorithm facilitates the recognition of words because the patient learns to listen to new speech clues. In contrast, the effects of instantaneous distortion of the spectrum of speech caused by the frequency lowering are poorer in normal hearing individuals as compared to real patients, once normal hearing individuals are not accustomed to a degraded speech signal.

The idea of conducting a pilot study allowed evaluating the variables that could influence the test applied in hearing impaired ones. It is intended, in future, to continue this study using the compression of frequencies in the hearing impaired ones with the presence of dead regions in the cochlea. For being a pilot study, we can and should be questioning the methodology applied. It is believed that the ratios of frequency compression used were too high and, therefore, it would be important to study lower compression ratios to promote less distortion in the speech signal, as other authors suggest(9).

Moreover, it is believed to be necessary a speech sample more appropriate to the proposal of this study with a larger sample, using recordings with male and female speakers (10). Also, it would be important to obtain a frequency of presentation of phonemes large enough to analyze the phonemic recognition of each group separately (11). This would allow the study of frequency compression behavior for each sound in particular and the precise benefits and harms of this algorithm for the recognition of words, according to each phonemic group separately analyzed.

Conclusions

1. The frequency compression ratios of 2:1 and 3:1 difficult speech recognition in normal hearing subjects.
2. The higher the frequency compression ratio is the worse the speech recognition is.
3. Familiarity with listened words facilitates their recognition even when these words are distorted by frequency compression.

References

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.