18

Frequency Lowering Algorithms for the Hearing Impaired

Francisco J. Fraga[1], Leticia Pimenta C. S. Prates[2],
Alan M. Marotta[3] and Maria Cecilia Martinelli Iorio[4]
[1]*Universidade Federal do ABC (UFABC),*
[2]*Universidade Federal de Minas Gerais (UFMG),*
[3]*Centro Federal de Educação Tecnológica de Minas Gerais (CEFET/MG),*
[4]*Universidade Federal de São Paulo (UNIFESP)*
Brazil

1. Introduction

Hearing, as a complex function, requires perfect cochlea and auditory pathways and any interference in this system may compromise its performance. Hearing aids have made great advances in hearing rehabilitation. However, audiology professionals still face failures, usually related to speech discrimination. This is a common situation in sloping sensorineural hearing loss, which is often associated with marked difficulties in word recognition, especially in detection and discrimination of fricative phonemes, even using hearing aids.

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. 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.

Hearing loss is more common for high–frequency and mid–frequency sounds (1 to 3 kHz) than for low–frequency. Frequently, there are only small losses at low frequencies (below 1 kHz) but almost absolute deafness above 1.5 or 2 kHz. A considerable percentage of the hearing impaired with moderate/severe hearing loss has audiograms where the losses are profound for high frequencies, severe for medium frequencies and mild or moderate for low frequencies.

Such problems lead researchers to lower the spectrum of speech in order to match the residual low–frequency hearing of listeners with high–frequency impairments. For these patients, lowering the high–frequency speech spectrum to the frequencies where the losses are mild or moderate could be a good processing tool to be added in the implementation of a digital hearing aid device. In this section we will provide a brief review of frequency

transposition and frequency compression algorithms developed for hearing aid users along the past three decades.

Given the difficulties encountered on amplification of high frequencies, frequency lowering has been suggested by some authors in an attempt to provide speech cues contained in high frequencies (Hicks et al., 1981; Reed et al., 1983; Reed et al., 1985).

Speech playback at a slower sampling rate and reduction of zero-crossing rate are some of the frequency lowering methods that were used in surveys conducted before the eighties, as reported by Hicks et al. (1981). All these methods involve signal distortion, more or less noticeable, usually dependent on the degree of spectral change. Such schemes perceptually modify important speech characteristics such as rhythmic and time patterns, psychophysical frequency (pitch) sensation, and duration of segmental elements. In their paper (Hicks et al., 1981), authors presented a remarkable investigation on frequency lowering. Their technique involves monotonic compression of short-time spectrum, without pitch alteration, while avoiding some problems observed in other methods.

Reed et al. (1983) conducted an experimental study to evaluate human speech recognition, using linear and nonlinear frequency compression, according to the method proposed by Hicks et al. (1981). Initially, the study was conducted with six normal hearing individuals, performing consonant discrimination experiments on these normal listeners and results were compared with the control condition, using low-pass filtering. They have observed that Hick's frequency lowering scheme presented better performance for fricative and affricate sounds if compared with low pass filtering to an equivalent bandwidth. On the other hand, the performance of the low pass filtering was better for vowels, semivowels and nasal sounds. For plosive sounds, both methods have shown similar results. In general, the performance on the best frequency–lowering conditions was almost the same to that obtained on low pass filtering to an equivalent bandwidth.

Subsequently, in another study (Reed et al., 1985) authors applied the frequency compression test in individuals with mild to severe hearing loss and sloping audiograms. Non-linear frequency compression was used and results were compared with conventional sound amplification. They reported that frequency compression was not beneficial in any condition.

Turner and Hurtig (1999) commented that the frequency region where the hearing loss occurs, as well as its extent, are determining factors in word discrimination. In general, hearing loss exceeding 60 dB HL at frequencies above 2-3 kHz causes a reduction in speech discrimination. Therefore, they believe that the perception of high frequencies speech cues can be improved by changing the high frequency components to low-frequency regions, since the hearing sensitivity in these regions is greater than 60 dB HL.

The authors hypothesized that, in hearing loss above 60 dB HL for high frequency sounds, the auditory system loses the ability to discriminate the articulation point, as this information is contained in higher frequencies (above 1 kHz) of the speech spectrum. Thus, they suggest that only patients with hearing loss at frequencies above 2 kHz and hearing thresholds better than 60 dB HL in the frequencies below 2 kHz have the potential to benefit from frequency compression.

McDermott and Dean (2000) evaluated speech recognition in individuals with sloping hearing loss whose tone thresholds at low frequencies were better than 30 dB HL, while in medium and high frequencies presented profound hearing loss. Twenty six subjects were

evaluated through a task of speech recognition in noise, in free-field at 65 dB A, with the signal/noise ratio of 6 dB. Upon testing, subjects were not using their hearing aids. The results were similar to those obtained in a previous experiment - which used the same speech material and procedures - performed with normal hearing individuals with a hearing loss simulation using similar low-pass filters and cutoff frequencies. The authors affirmed that using normal hearing subjects on this type of experiment is advantageous because it ensures that evaluation of the algorithm is not influenced by other factors associated with sensorineural hearing loss.

Simpson et al. (2005) developed a frequency compression algorithm for hearing aids. They used a nonlinear compression method, increasing progressively the compression ratio for high frequency sounds. To evaluate the algorithm, they conducted a study on recognition of monosyllabic words in quiet, with 17 hearing aids users with moderate to severe sloping sensorineural hearing loss. Their objective was to compare phoneme recognition using conventional hearing aids and those with the frequency compression algorithm embedded. Of the 17 participating subjects, eight had improvement in recognition of phonemes using frequency compression; eight showed no differences regarding the amplification used and one subject presented significantly lower performance with the algorithm. Fricative sounds were the most favoured by frequency compression.

Using the same algorithm developed in their earlier studies, authors (Simpson et al., 2006) studied frequency compression in seven individuals with hearing loss suggestive of cochlear dead regions, defined as regions in the cochlea that have no inner hair cells or adjacent functional neurons (Moore et al., 2000). In general, performance in the task of speech in quiet with conventional amplification was similar to performance with frequency compression. The authors commented that it is possible that frequency compression has brought benefits in the discrimination of some phonemes and detriment of others, so the final score did not change. For example, fricative phonemes /ʃ/ and /ʒ/ were more correctly identified by frequency compression than by conventional amplification. On the other hand, recognition of phoneme /s/ was reduced. Fricatives /ʃ/, /z/ and /v/ were the most selected when frequency compression was used.

Kuk (2007) gave a brief discussion about benefits and applicability of frequency transposition. In that article, he presents the frequency transposition algorithm developed by Widex, available on the Inteo hearing aid. Continuing the previous study, Kuk et al. (2007) discussed the importance of experience with the frequency transposition algorithm to fitting success. According to authors, on the first experience, some users do not like the sound quality provided by this algorithm, referring to sound harsh or unnatural. They concluded that the great sound change caused by frequency transposition account for this negative reaction, requiring a long-term use so that a reorganization of the cortical tonotopic representation may occur.

Robinson et al. (2007) considered severe high frequency hearing loss when average of pure tone thresholds at 4.0, 6.0 and 8.0 kHz was less than 75 dB HL, and stressed that this loss is prevalent in 24% of individuals over 60. The authors presented a new method of frequency transposition applied only on sounds that have significant energy at high frequencies. They set the dead region frequency edge considering this feature a fundamental key in the individual formatting of the frequency transposition algorithm. They concluded that the benefit observed in the use of frequency transposition was reduced by the confusion generated between phonemes. Thus, they hypothesized that frequency transposition,

instead of improving discrimination of consonants, improves detection of fricatives. The authors evaluated the new algorithm to detect the phonemes /s/ and /z/ in final word position. They used 24 pairs of words that differed by the presence or absence of /s/ and /z/ in final syllables, recorded by a single female speaker. The assessment was performed by seven subjects with hearing loss and cochlear dead regions at high frequencies. Participants were instructed to select visually one of the words of the pair. As a result, authors found that frequency transposition significantly improved task performance, compared to the control condition.
In next sections we present some frequency compression and transposition algorithms we developed and evaluated for helping the hearing impaired with severe high frequency hearing loss.

2. Comparison of two frequency lowering algorithms for digital hearing aid

In this section we present a new frequency-lowering algorithm that uses frequency transposition instead of frequency compression. Furthermore, the frequency transposition is applied only over fricatives and affricates, leaving the other speech sounds untouched, as previous works have shown that frequency lowering only benefits high frequency phonemes. To perform comparison, we have also implemented a frequency compression algorithm based on Hick's method (Hicks et al., 1981).
Results of subjective preference (considering speech quality) indicate better performance of our frequency transposition method compared to Hick's frequency compression method. We also present subjective intelligibility tests over 20 subjects, showing that in this case performance (now considering speech intelligibility) depends on which are the specific phonemes being processed by these two algorithms.

2.1 Audiometric data acquisition and processing

The first step of both frequency-lowering algorithms consists in audiometric data acquisition of the hearing impaired subject. The audiometric exam is employed for measuring the degree of the hearing impairment of a given patient. In this exam, the listener is submitted to a perception test by continuously varying the sound pressure level (SPL) of a pure sinusoidal tone in a discrete frequency scale. The frequency values most frequently used are 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 6 kHz and 8 kHz. For each of these frequencies, the minimum SPL in dB for which the patient is capable of perceiving the sound is registered in a graph.
The audiogram is the result of the audiometric exam, which is presented by a graph with the values in dB SPL for each of the discrete frequencies. This graph is done separately for each subject's ear. Since the level of 0 dB SPL is considered the minimum sound pressure level for normal hearing, the positive values in dB registered on the vertical axis of the audiogram can be considered as the hearing losses of the patient's ear.
If the average loss at frequencies of 500, 1000 and 2000 Hz is equal or inferior to 20 dB, the subject is considered as having normal hearing. From 21 to 40 dB, hearing loss is classified as mild. Average loss greater than 40 dB but inferior to 70 dB is considered moderate. From 71 to 90 dB, we consider that patient has severe hearing loss and more than 95 dB of loss is classified as profound.
The threshold of discomfort, for normal or impaired listeners, is always below 120 dB SPL. Indeed, commonly the threshold of discomfort for the hearing impaired is lower than for

normal hearing subjects. Although less common, some audiograms bring both the threshold of discomfort and the threshold of hearing (Alsaka & McLean, 1996), as one can observe in Fig. 1. In this figure, points of audiogram corresponding to the right ear are signalled with a round mark and those corresponding to the left ear are signalled with an X mark. These marks are worldwide used in this way by audiologists. The dynamic range of hearing for each frequency is the threshold of discomfort minus the threshold of hearing.

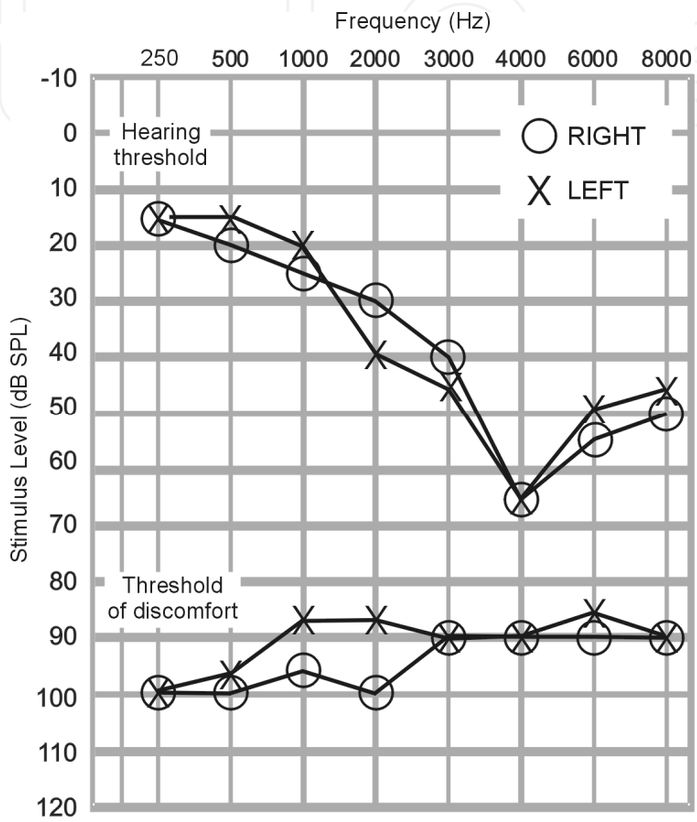

Fig. 1. Sloping sensorineural hearing loss case

Based on the acquired audiometric data, the algorithm analyses the range of frequencies where there is still some residual hearing. The criterion used is the following: first, it is verified if patient has a ski–slope kind of losses, i.e., if the losses are increasing with frequency. Only patients with this type of impairment can be aided by any frequency lowering method.

After that, the first frequency where there is a profound loss is determined. If this frequency is between 1.2 kHz and 3.4 kHz, a destination frequency to which the high–frequency spectrum will be transposed is calculated. Otherwise, no frequency transposition is needed (residual hearing above 3.4 kHZ) or profitable (residual hearing below 1.2 kHZ). This destination frequency is considered as the geometrical mean between 900 Hz and the highest frequency where there is still some residual hearing. The geometrical mean was empirically chosen because it provides a good trade-off between minimum spectrum distortion and maximum residual hearing profit. In order to obtain more accuracy in the losses thresholds, audiogram points are linearly interpolated.

As will be explained further, tests were done first in normal hearing people, simulating hearing losses by means of low-pass filtering. In this case, for destination frequency calculation as well as for definition of other algorithm parameters (see 2.2), the low-pass filter cut-off frequency is considered as the highest frequency where there is still some residual hearing.

2.2 Speech data acquisition and processing

Speech signals are sampled at 16 kHz and Hamming windowed with 25 ms windows. These windows are 50% overlapped, what means that the signal is analyzed at a frame rate of 1/12.5 kHz. A 1024-point FFT is used for representing the short–time speech spectrum in the frequency domain.

If in the previous audiometric data analysis a ski–slope kind of loss was detected and the frequency transposition criterion was matched, a destination frequency has already been determined. Then, we have to find out (in a frame–by–frame basis) if the short–time speech spectrum presents significant information at high frequencies that justify the frequency transposition operation. The criterion used for transposing or not the short–time spectrum of each speech frame depends on a threshold. When the signal has high energy in high frequencies the algorithm transposes high frequency information to lower frequencies. The threshold is set for suppressing the processing of all vowels, nasals and the semivowels, while activating frequency transposition for fricatives and affricates.

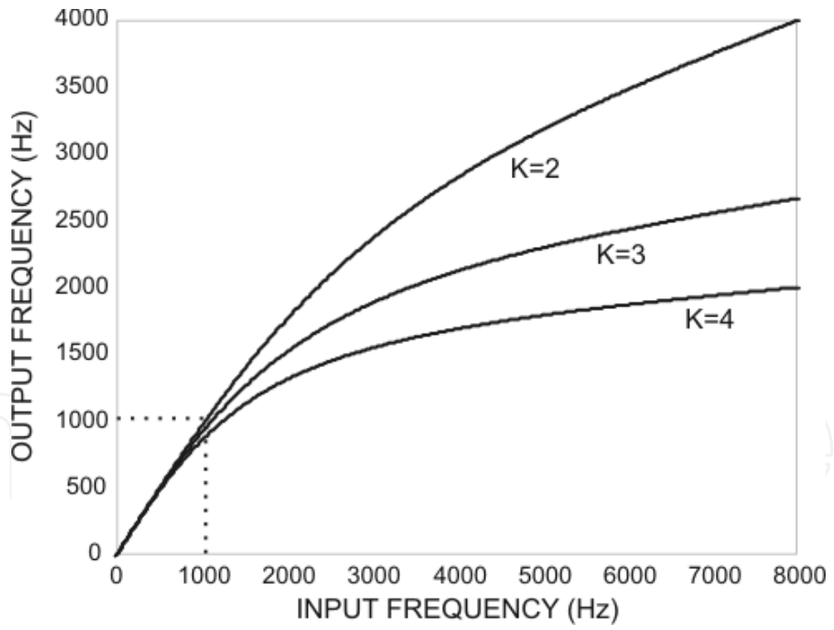

Fig. 2. Input-output frequency compression curves

To decide which part of the spectrum will be transposed, the energy of 500 Hz bandwidth sliding windows are calculated with 100 Hz spacing, from 1 kHz to 8 kHz (Nyquist frequency). This is done with the aim of find out an origin frequency. The origin frequency is the frequency 100 Hz below the beginning of the 500 Hz bandwidth window with maximum energy. The part of the spectrum that will be transposed corresponds to the range

of all frequencies above the origin frequency. This empirical criterion guarantees that the unavoidable distortion due to the frequency lowering operation will be profitable. Because in this way the most important part of high–energy spectrum will be transposed to lower frequencies, maintaining untouched low–frequency information.

For comparison, the Hick's frequency compression scheme was already implemented, but now only when the same frequency lowering criterion (high/low frequency energy ratio) used for transposition was matched, i. e., only for fricatives and affricates. The frequency compression was done by means of the same equation used by Reed et al. (1983). But in practice, it is more useful to implement its inverse equation, which is

$$\frac{f_{IN}}{f_S} = \frac{1}{\pi}\tan^{-1}\left[\left(\frac{1-a}{1+a}\right)\tan\left(K\pi\frac{f_{OUT}}{f_S}\right)\right] \quad , \quad a = \frac{K-1}{K+1} \tag{1}$$

where f_{IN} is the original frequency, f_{OUT} is the corresponding compressed frequency, K is the frequency compression factor, a is the warping parameter and f_S is the sampling rate. For minimum distortion at low frequencies, the warping parameter must be chosen by the ratio defined in second part of equation (1).

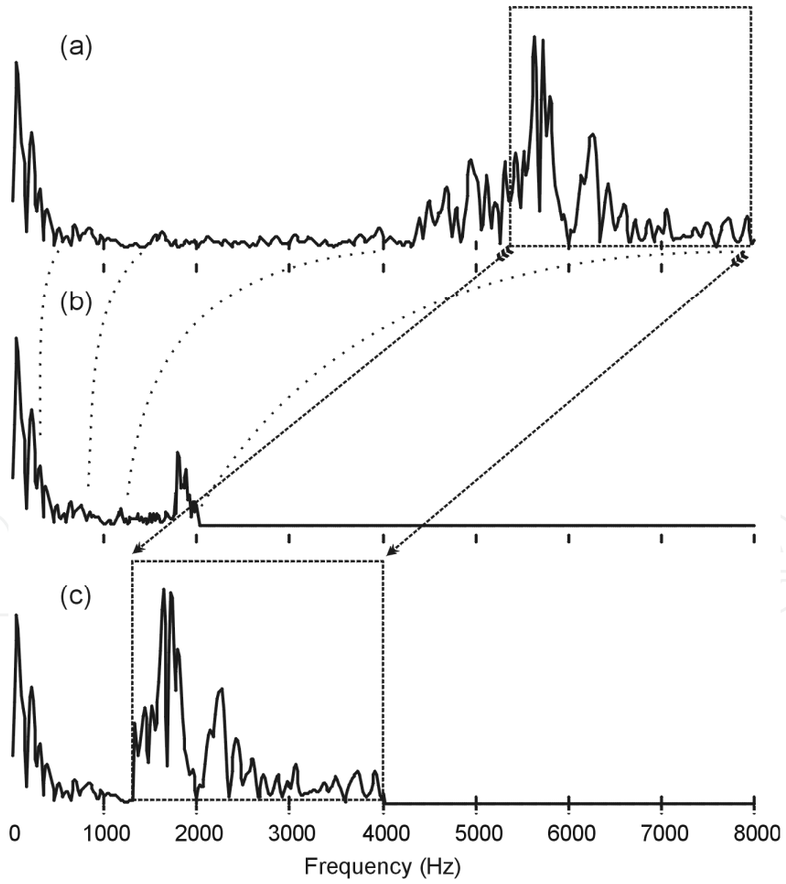

Fig. 3. Comparison of frequency lowering schemes: compression (b) versus transposition (c)

As it occurs with the destination frequency calculation (see 2.1), the compression factor K was determined according to the listener's loss degree. Fig. 2 shows the curves of equation (1) for K = 2, 3 and 4. In this figure we can see that low frequency spectral content (below 1000 Hz) is barely compressed.

For better understanding and comparison of frequency transposition and frequency compression spectral effects, in part (a) of Fig. 3 one can see the original short-time spectrum of a speech frame taken from a female pronunciation of phoneme /s/, in part (b) the same frame is shown compressed by a factor K = 4 and part (c) presents the frame after frequency transposition. It is easy to observe that frequency transposition preserves the spectral shape, what does not occur in the case of frequency compression, where one can clearly note a great amount of shape distortion at high frequencies, but still preserving low frequency spectral content.

2.3 Preliminary qualitative test

Both frequency lowering algorithms (frequency compression and frequency transposition) were not already tested with hearing impaired subjects. But we got some preliminary results with normal listeners, first considering only qualitative aspects of processed speech. In this case, a simple low pass filtering process simulates the losses above the frequency where there is no more residual hearing. In this preliminary qualitative test, cut-off frequency was fixed to 2 kHz.

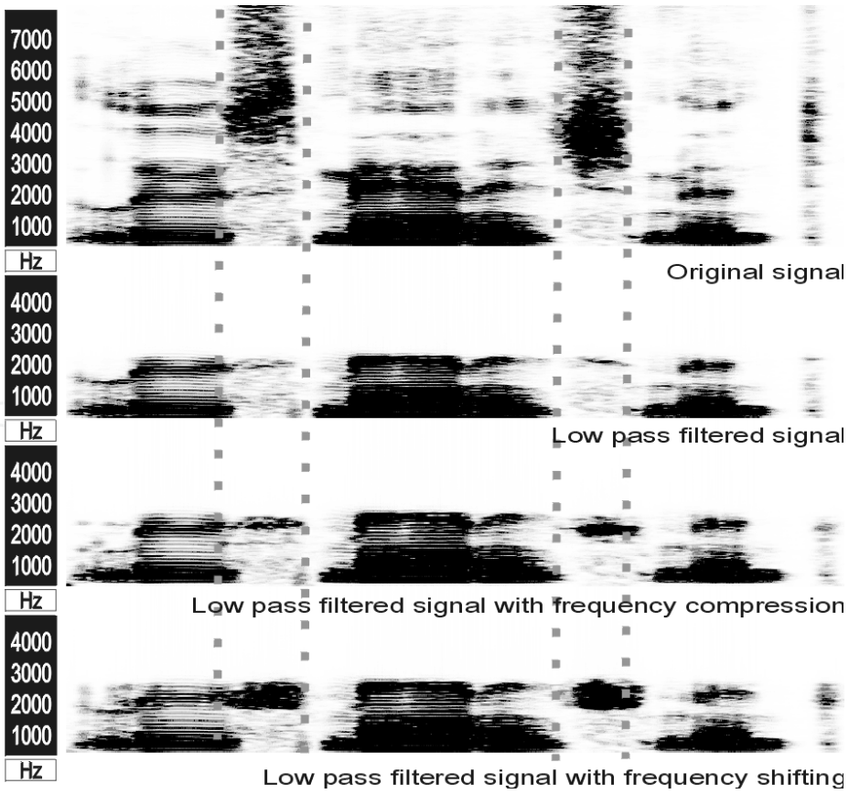

Fig. 4. Spectrograms of "loose management"

The experiment we have carried out consists of submitting the low-pass filtered speech signal to both frequency lowering algorithms. After, two normal hearing adults (one male and one female) listened to the processed speech signals, as well as to the control condition (low-pass filtering only). Listeners are not previously informed regarding signals' processing schemes and are asked for ranking the three speech sounds according to their subjective quality.

In this preliminary test, only two different speech signals were submitted to the algorithms. The original and processed spectrograms of one of these speech signals (a male pronunciation of the words 'loose management') are shown in Fig. 4, where we can appreciate again the visual difference between the two frequency lowering algorithms. It is important to remark that both listeners are native speakers of Brazilian Portuguese and are not used to listen to English words in their everyday tasks. This is important because in this way we intend to separate speech quality from speech intelligibility, although both aspects are correlated and cannot be perfectly assessed by subjective evaluation.

According to our prevision, only fricative speech sounds were frequency lowered in both algorithms. The unique exception is the phoneme / *l* /, which is not fricative but lateral approximant. But in this case, its pronunciation had high frequency energy, as we can observe in the spectrogram of the original speech signal (see Fig. 4).

Listeners' preferences were listed in Table 1. In this table, 'Signal 1' is the Portuguese word "pensando" (which means 'thinking'), pronounced by a Brazilian Portuguese native speaker (male), and 'Signal 2' is the English words 'loose management', the latter corresponding to the spectrograms of Fig.4.

Speech signal	Male listener	Female listener
Signal 1 – low pass filtering (control)	1st	3rd
Signal 1 – frequency compression	3rd	2nd
Signal 1 – frequency transposition	2nd	1st
Signal 2 - low pass filtering (control)	2nd	2nd
Signal 2 – frequency compression	3rd	3rd
Signal 2 – frequency transposition	1st	1st

Table 1. Listener's preferences

One can observe that listeners' preference are more consistent when listening to foreign words (english, in this case), where speech quality can be easily separated from speech inteligibility. These preliminary results indicate that the frequency shifting (or transposition) method was preferred by the listeners when compared to the frequency compression method. But it is important to remark that the subjective difference between the low pass filtered signal, the frequency–compressed signal and the frequency-shifted signal is very slight, as perceived by normal listeners.

2.4 Preliminary intelligibility test

We perform a syllable identification test over 42 normal hearing subjects, 31 male and 11 female; a simple low pass filtering process simulates the losses above the frequency where there is no more residual hearing. Speech material consists of 21 different CV phonetic syllables, which are composed with the seven Brazilian Portuguese fricative phonemes ([*f*], [*v*], [*ʒ*], [*ʃ*], [*s*], [*z*], [*x*]) followed by three vowels ([*a*], [*i*], [*u*]).

Six Brazilian Portuguese native speakers, three female and three male, pronounced once all these syllables. After processing, each utterance generates nine speech signals: low-pass filtered syllable (control), frequency compressed syllable and frequency transposed syllable. Previously to frequency lowering, all signals passed through three different low–pass filters with cutoff frequencies of 1.5, 2 and 2.5 kHz, thus forming a final speech database composed by 1134 WAVE files.

After has heard three times a phonetic syllable at random, the listener must choose one written syllable from a list of seven possibilities, because only syllables with the correct vowel is presented. Due to the random choice of the syllables that were presented to the listeners, there were some syllables that were less listened than others. But each of the 63 different processed fricatives was presented at least 5 times to each listener, and any of them was presented more than 15 times.

Processed Syllable	None	Compression	Shifting
[f] 1500	61,5	72,7	62,5
[f] 2000	40,0	44,4	69,2
[f] 2500	85,7	53,3	58,3
[v] 1500	78,6	80,0	81,8
[v] 2000	100,0	71,4	66,7
[v] 2500	77,8	61,5	90,9
[ʒ] 1500	25,0	28,6	33,3
[ʒ] 2000	50,0	81,8	86,7
[ʒ] 2500	69,2	62,5	77,8
[ʃ] 1500	0,0	20,0	45,5
[ʃ] 2000	44,4	62,5	55,6
[ʃ] 2500	77,8	100,0	84,6
[s] 1500	53,8	50,0	8,3
[s] 2000	73,3	36,4	33,3
[s] 2500	76,9	41,7	25,0
[z] 1500	57,1	46,7	75,0
[z] 2000	70,0	60,0	40,0
[z] 2500	44,4	60,0	33,3
[x] 1500	66,7	40,0	12,5
[x] 2000	46,2	21,4	38,5
[x] 2500	55,6	38,5	75,0

Table 2. Listener's correct decisions (%).

The results of this test are shown in Table 2, where column None means no processing further than low pass filtering, Compression means frequency compression and Shifting means frequency shifting (transposition). In the first column we have all the possible

fricatives for each (three) filter cutoff frequencies. In the table, numbers signaled in boldface correspond to the greatest percentage of correct decisions made for each processing type.

2.5 Discussion and conclusions

The slight perceived difference in the quality observed by both male and female listeners among the processed signals may be due to the fact that the disparity between the original signal (with frequencies up to 8 kHz) and the low pass filtered (2 kHz) signals is large. But for the impaired subject, that never (or for a long time) had any perception of sounds with frequencies above 2 kHz, may be the difference between the processed signals was not so slight.

Relatively to the results of the intelligibility test, results are difficult to analyze if we consider the set of syllables as a whole. But it is interesting to analyze each fricative sound in particular. For example, we can conclude from the results that for phone [s] the better is to do no further processing, but if we consider phone [ʃ] we conclude just the opposite: no processing leads to zero percent correct identification when the highest audible frequency is 1.5 kHz, but it improves to 45,5% when submitted to frequency transposition (shifting). In the case of fricative sound [ʒ], the better results are also obtained with our frequency shifting algorithm, independently of the signal bandwidth. For all other situations, the optimal solution depends on the specific phone and cutoff frequency considered.

Considering the set of phonemes as a whole, based on the preliminary syllable identification test, we observed that there was no statistical significance between both frequency lowering algorithms when compared to low-pass filtering ($p > 0.05$), as well as compared to eachother ($p > 0.07$). But considering invidual phonemes, we can conclude that if we incorporate a simple automatic phoneme classifier in the system, it is possible to choose the better frequency lowering algorithm to be applied for each specific phone, given the maximum frequency where there is some residual hearing. This is not difficult to do, considering the advances observed in the performance of automatic phoneme recognition algorithms over the last years (Scanlon et al., 2007). Finally, it is important to remark that both algorithms have demonstrated to be fast enough to enable their usage in digital hearing aid devices.

3. Frequency compression and its effects on human speech recognition

In this section we present the development and evaluation of the same frequency compression algorithm described in section 2, but 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 research was to conduct a descriptive analysis of results in normal individuals, considering the compression ratio applied and the familiarity with the words of the test.

3.1 Methods

This study was conducted at the Department of Integrated Care, Research and Teaching on Hearing (NIAPEA), Federal University of São Paulo - PaulistaSchool of Medicine, after

approval by the Research Ethics Committee of the Federal University of São Paulo / Hospital São Paulo, under the protocol 0150. 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/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: group F, composed by Speech-Language Pathologists/Audiologists and group P, composed by the remaining participants.

Participants had hearing thresholds better than 20 dB in the frequencies from 250 to 8000 Hz, measured before the beginning of the evaluation. Speech material used in this study consisted of monosyllabic words applied through TDH 39 headphones at 60 dB NA intensity, in silence, monotic task, both ears. The subjects were instructed to repeat, exactly, the monosyllables presented. The word recognition rate (WRR) was established by counting the number of words repeated correctly and dividing by the number of words heard.

For the word recognition test (WRT) we used a list of 25 monosyllabic words phonetically balanced (Pen & Mangabeira, 1973) and available on CD (Pereira & Schochat, 1997). A new organization of the words list 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, we used the Aurical™ audiometer from Madsen Electronics™, coupled to a personal computer. The speech procedures were applied in a sound proof booth using a portable compact disc player, model 4147 from Toshiba™, coupled to the Aurical™ audiometer and TDH 39 headphones, besides the CD containing speech samples.

The words in the lists had the speech spectrum modified by frequency compression. We performed all speech processing using Matlab™, at the Engineering and Modeling Center of the ABC Federal University (UFABC. After processing, speech material were assembled in a computer and recorded on CD.

Frequency compression was performed by non-linear method - i.e. performing smaller compression in low frequencies and further compression on high frequencies (6). Speech signals are sampled at 16 kHz and Hamming windowed with 25 ms windows. These windows are 50% overlapped, what means that the signal is analyzed at a frame rate of 1/12.5 kHz. A 1024-point FFT is used for representing the short–time speech spectrum in the frequency domain.

Three following compression ratios were used (or compression factor - K) in the words lists: 1:1 (K = 1), 2:1 (K = 2) and 3:1 (K = 3), thus composing three lists of frequency compressed words. 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 16 kHz. Compression ratios of 2:1 and 3:1 (or compression factor K = 2 and K = 3) mean application of frequency compression in different proportions.

The higher the compression ratio is, the greater the degree of frequency lowering - which creates major changes on the speech spectrum. The frequency compression curves used in this study can be observed in Figure 1. These curves were implemented directly in the frequency domain, in a frame-by-frame basis, using the equation shown in the lower right

corner of the figure, where variable a controls the degree of nonlinearity of the curves (a = 0 turns the curve into a straight line). Speech processing back to the time domain were performed by the well-known overlap-and-add method (Nawab & Quatieri, 1998).

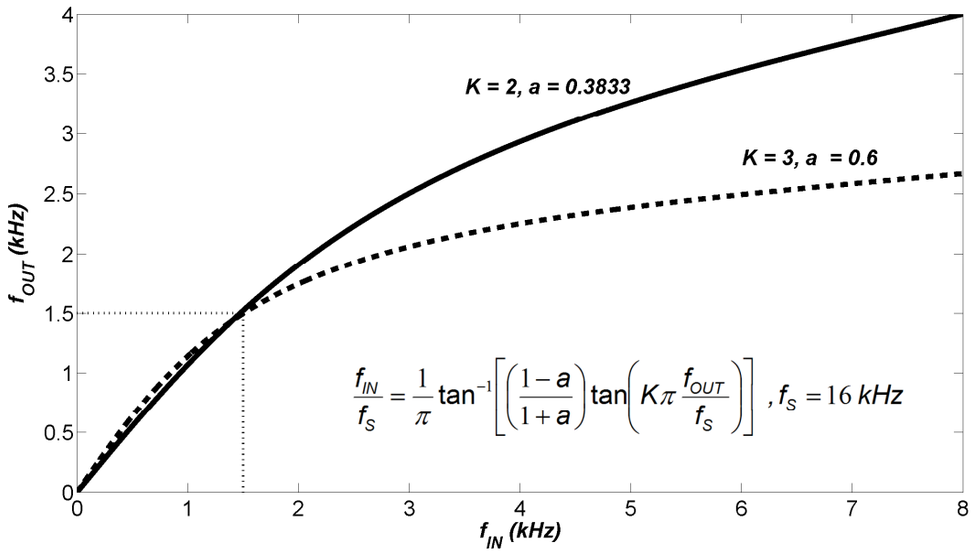

Fig. 5. Frequency compression curves used in the pilot study. Horizontal axis depicts the frequency range of input (original) signal and vertical axis the frequency range of output (processed) signal for the compression factors K = 2 (full line) and K = 3 (dashed line).

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 output frequencies (of processed signal) correspond exactly to half the values of input frequencies (of 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 and with no compression in the range from 0 to 1 kHz. In this study, the approximate range of linearity (and no compression) was extended up to 1.5 kHz, aiming to reduce as most as possible the perceptual distortion of main pitch harmonics and formants of the original speech signal.

Figure 6 displays the spectrogram of the Portuguese monosyllable "jaz" (/ʒas /, male speaker) 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, one 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. This was done in order to not provide clues that could facilitate the recognition of words, once all 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.

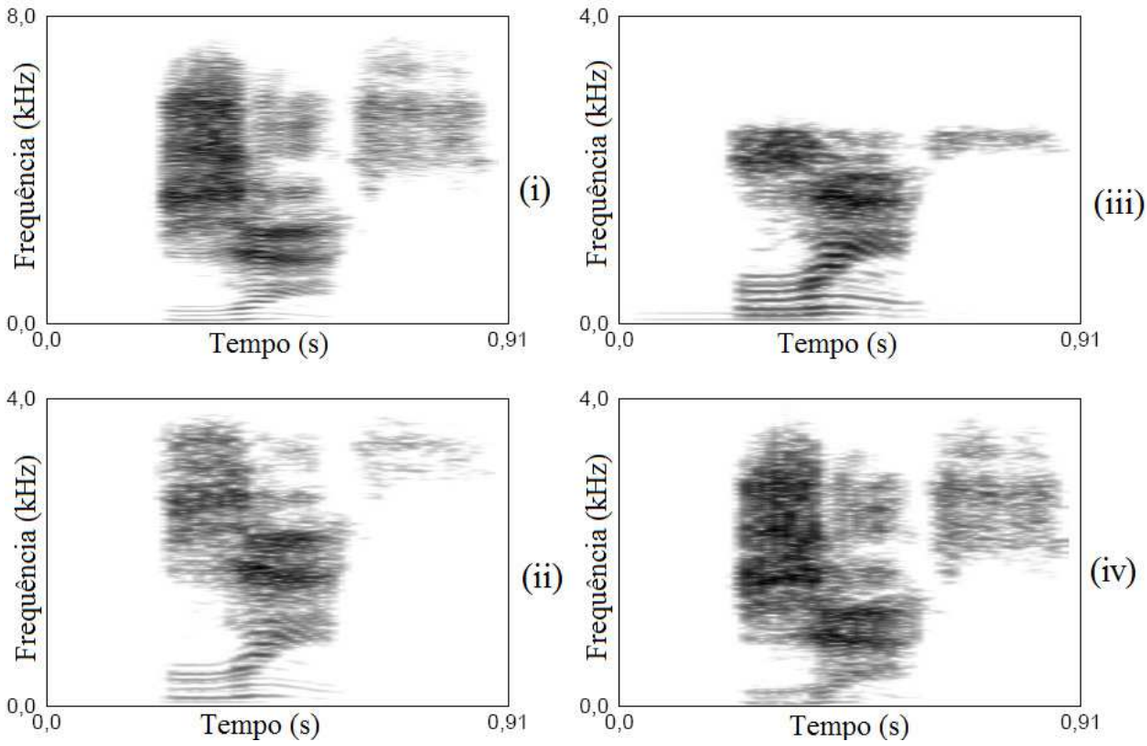

Fig. 6. Spectrograms of /ʒas/ (male speaker): (i) original (K = 1 and a = 0); (ii) non-linear compression with K = 2 and a = 0.3833; (iii) nonlinear compression with K = 3 and a = 0.6; and (iv) linear compression, with K = 2 and a = 0 (situation not assessed in this study).

3.2 Results

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

As there were no statistically significant differences between the WRR obtained for the right and the left ear in both groups, as demonstrated by the p-values at bottom line of Table 3, we chose to perform the remaining analysis considering the values of both ears. Thus, the samples are doubled, making the results statiscally more reliable.

Thus, in Figure 7 we show the WRR average values obtained in groups P and F (joining both ears), considering the compression ratio (compression factor K).

3.3 Discussion

Evaluation of human speech recognition using frequency compression 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 methods as an attempt to improve speech recognition, especially for the hearing impaired with losses at high frequencies.

	Group P						Group F					
	K3		K2		K1		K 3		K 2		K 1	
	OD	OE	OD	OE	OD	OE	OD	OE	OD	OE	OD	OE
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
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
Q2	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-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

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

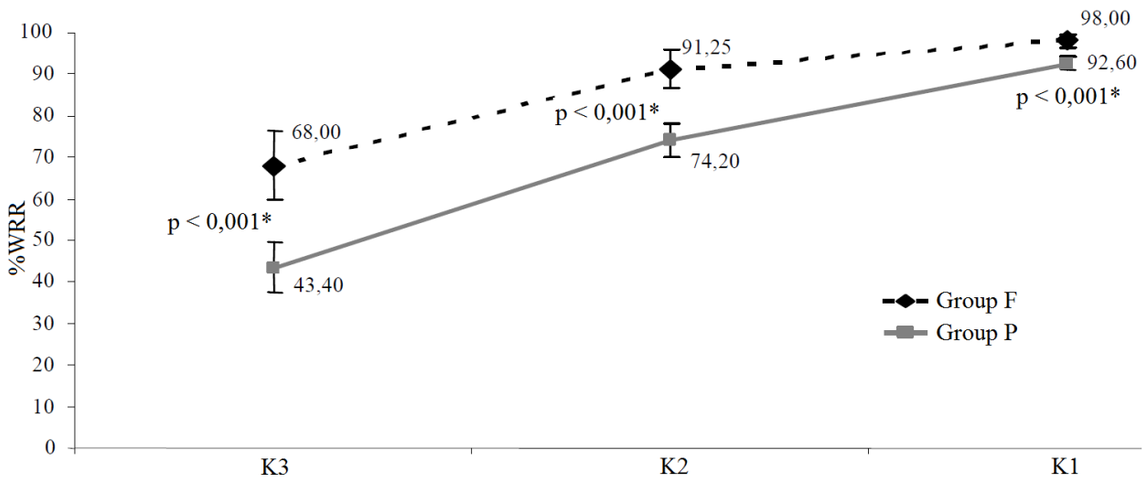

Fig. 7. Comparative graphic of average WRR values obtained in groups P and F, with compression factors K = 3, K = 2 and K = 1, now considering both ears (joined). Statistical significance (p-values) was established with Mann-Whitney Test.

With the discovery of dead regions in the cochlea (Moore et al., 2000), and successive studies demonstrating its negative impact on the ability of word recognition (Gordo & Iorio, 2007), 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 cochlear dead regions.

The purpose of this study was to develop a frequency compression algorithm and to assess, in normal individuals, the word recognition rate (WRR) using this algorithm. With the aim of conducting a pilot study, we used frequency compression in three distinct ratios: 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.

As expected, we found poorer performance on tests of word recognition the higher the compression ratio was in both groups evaluated. Figure 7 shows that group F presented better performance in all compression ratios evaluated ($p < 0.001$). Based on 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 prior training using a hearing aid with this algorithm embedded can be a way to improve word recognition by the hearing impaired.

Still in Figure 3, it can be noticed by the crescent lines a gradual improvement in WRR as the compression ratio decreases. This trend could be observed for both groups. A study conducted with normal hearing participants using a linear 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 (Turner & Hurtig, 1999). 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, with less signal distortion.

Moreover, in the present work we used non-linear frequency compression, while this study used only the linear compression. Other authors (Turner & Hurtig, 1999; Simpson et al., 2005; Baskent & Shannon, 2006) concluded that frequency lowering algorithms should be implemented cautiously, in order to avoid strong signal distortion. Almost all authors believe that prior training with the algorithm facilitates the recognition of words because the patient learns how to listen to new speech clues.

In contrast, the perceived effects of distortions in speech spectrum caused by frequency lowering are greater in normal hearing individuals as compared to hearing impaired subjects, once normal hearing are not accustomed to listen to degraded speech signals.

The idea of conducting a pilot study with normal hearing subjects allowed evaluating the variables that could influence the test applied in hearing impaired ones. It is intended, in future, to continue this study applying frequency compression for the hearing impaired with dead regions in the cochlea. As this was just a pilot study, we can and should be questioning the methodology applied. We believe that we used too high frequency compression ratios and, therefore, it would be important to study lower compression ratios to promote less distortion in the speech signal, as other authors suggest (Turner & Hurtig, 1999).

Moreover, we believe it will be necessary to use speech material more appropriate to the proposal of this study, with a larger sample, using recordings from both male and female speakers (Baskent & Shannon, 2006). Also, it would be important to design a WRT with more repetitions of the same phonemes, enabling us to analyze the recognition of phonemic groups separately (Simpson et al., 2005). This would allow the study of frequency compression effects for each sound in particular and the precise benefits and harms of this algorithm for human word recognition.

3.4 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.

4. Frequency compression/transposition of fricative consonants for the hearing impaired with high-frequency dead regions

Moore et al. (2000) called Dead Regions (DR) those parts of the cochlear basilar membrane with complete absence of inner hair cells. They alerted that simple sound amplification (by a hearing aid) over a dead region may be unbeneficial and may even impair speech intelligibility. Face to this difficulty, frequency compression or transposition have been suggested by many authors in the attempting to bring the high-frequency speech information to lower frequencies.

An overview of more recent studies in frequency compression/transposition is provided by Robinson et al. (2007). They developed a new frequency transposition method too, applied only to fricative and affricate sounds. But their results showed that there was no statistical significant improvement for fricatives discrimination. They concluded that the increasing in the confusion between some fricative phonemes have canceled the effect of the better recognition of others. Based on these negative results, the primary target of this research was the development of a frequency compression algorithm to be applied only to fricative consonants and that does not increase the confusion between them. We have also not observed in previous works a direct concern in making frequency compression according to the average spectral shape of fricatives or any other speech sound.

In this section, we present the design of our original piecewise linear frequency compression/transposition curve was made taking into account the average short-time spectrum of the most frequent Brazilian Portuguese (BP) fricatives. In the first phase of our research, which is described in this section, the dead regions were simulated by low-pass filtering of the speech material presented to normal hearing listeners.

4.1 Piecewise linear frequency compression

The frequency compression/transposition algorithm was implemented with Matlab™. The signal analysis computations are made in the frequency domain and the processed speech is re-synthesized in the time domain using the well-known overlap-and-add technique (Nawab & Quatieri, 1998). After normalizing the dynamic range of the speech signals (each recorded utterance should has the same rms value), they were divided in frames of 50 ms (800 samples) with an overlap of 75% between adjacent frames.

Then a 2048-point FFT is applied to each speech frame, which was previously multiplied by a Hamming window in the time domain. For the control condition, we just eliminate the frequency domain samples corresponding to the simulated dead region (low-pass filtering).

In our algorithm, the frequency compression curve should be applied only over non-vocalic speech sounds, i.e., just for noise-like consonants (fricative and affricates). To perform such sound classification, we calculate the Spectral Flatness Measure (SFM) of each signal frame, which is used to determinate the noise-like or tone-like nature of a given speech frame. We develop a method based on the original work of Johnston (1998) but with some modifications.

In a recent published research on frequency transposition (Robinson et al., 2007) it was used a much simpler criterion to do the same task, which is based in the energy ratio between high-frequency and low-frequency power. In a previous work, which was presented in section 2, we also used this same simple criterion and we have actually tried to use it again, but we did not achieve a hundred percent efficiency in the frame classification task. Experimentally we have observed that the straightforward high-frequency to low frequency energy ratio criterion has failed sometimes to classify correctly a noise-like speech sound, mainly in the case of voiced fricatives. Otherwise, using our method (based on SFM), all speech frames of all fricative consonants from our database were properly classified as noise-like ones. In addition, any frame belonging to a vowel or a silence segment in the speech material was misclassified.

Applying our SFM criterion, it was possible to verify whether a short-time spectrum of the audio signal has a noise-like or tone-like nature. This means, in practice, that the frequency compression curve acted only and always on fricative phonemes –voiceless or voiced.

SFM calculation consisted on the following steps:

1. Power spectrum of each frame was obtained by multiplying each FFT sample by its complex conjugate;
2. The geometric mean (G_q) of the power spectrum of the current speech frame q was calculated including only the frequency range from 800 to 2800 Hz;
3. The arithmetic mean (A_q) of the power spectrum of the current frame q was calculated including only the frequency range from 800 to 2800 Hz;
4. SFM of the current frame q was calculated in decibels, according to (1)

$$SFM_{dB} = 10\log_{10}\left(\frac{G_q}{A_q}\right) \tag{2}$$

5. The factor a_q of the actual frame was defined and calculated by (2)

$$\alpha_q = \min\left(\frac{SFM_{dB}}{-40}, 1\right) \tag{3}$$

6. The a_m factor was calculated as the output of a moving average filter applied to the factor a of the last P frames:

$$\alpha_m = \frac{1}{P}\sum_{p=0}^{P-1}\alpha_{q-p} \tag{4}$$

7. Similarly, the same moving average filter was applied to the last P values of arithmetic mean A_q:

$$A_m = \frac{1}{P}\sum_{p=0}^{P-1}A_{q-p} \tag{5}$$

8. If the a_m factor was greater or equal to the tonality threshold a_T, the current frame was considered of tone-like nature and the mean tone-like average A_T was updated with the value of the arithmetic mean A_q of the current power spectrum;

9. If the a_m factor was lower than the tonality threshold a_T, the current frame was considered of noise-like nature if, additionally, the arithmetic mean A_q was at least four times lower than the mean tonal value A_T.

Experimentally, it was verified that the values P = 4 and a_T = 0.03 were the ones that produced the best results. For these values, the algorithm achieved 100% efficiency on the speech classification task (tone-like or noise-like nature) of frames of fricative consonants when applied on 192 recorded monosyllables. Thus, the frequency compression algorithm was only applied on the current speech frame if it was classified as a noise-like nature signal by the above described method.

Considering the spectral characteristics of studied sounds, the SFM calculation was applied only in the frequency range between 0.8 and 2.8 kHz to allow the detection of sounds of a non-tonal (noise-like) nature without changing the identity of the remaining sounds. It is well known that the presence or absence of voicing is expressed mainly at low frequencies (Russo & Behlau, 1993; Robinson et al., 2007). So, in order to correct classifying the voiced fricatives as noise-like nature signals, the frequencies in the range between 0 and 0.8 kHz were not considered for the SFM calculation as the compression of all fricatives, both voiced and voiceless, was desired. The sounds above 2.8 kHz were also excluded from the SFM calculation because the main vowel formants correspond to frequencies between 500 and 3000 Hz (Behlau, 1984) and consequently the harmonic structure of vowel sounds is stronger in this frequency range.

Based on our own experiments as well as in literature about the average spectral distributions of Britain English, European Portuguese and Spanish fricatives (Jesus 2001; Manrique & Massone, 1981), we confirmed that spectral cues for discrimination between fricatives with different articulation places are all above 2000 Hz. This fact is the major reason behind the difficulty in the fricatives differentiation observed in patients presenting high-frequency dead regions in the cochlea. Joining this result from literature with our own, we designed the piecewise linear frequency compression curve shown in Figure 8.

Following, the compression curve is described justifying each designed part (I, II and III) with their respective compression ratios (CR) according to the frequency ranges of the original speech signal:

i. From 0.0 to 0.5 kHz: To preserve the pitch perception of voiced fricatives, this frequency regions remains untouched;
ii. From 0.5 to 3.0 kHz: In this part of the spectrum, only the fricatives /ʃ/ and /ʒ/ offer cues for phoneme identification. But these cues (basically an increase of spectral power) continue until 6500 Hz approximately. Thus, we applied a strong compression (CR = 0.2) to this region since there is no relevant information for fricative discrimination;
iii. From 0.5 to 3.0 kHz: For this frequency range the CR becomes 0.67 in order to preserve the original speech information, because most of the cues for fricative discrimination belong to this region. Just for clearness reasons, we divide this frequency range in two parts: from 3.0 to 4.5 kHz, which will be mapped to 1.0-2.0 kHz after compression (see Figure 3), and from 4.5 to the Nyquist frequency, mapped to 2.0-4.33 kHz after compression. The main purpose of this research is help the hearing impaired with dead regions above 1.5 or 2.0 kHz, so the first part of this region is the major one. We hypothesize that the transposition of these frequencies to the frequency range from 1.0 to 2.0 kHz will be effective to improve the perception and discrimination of fricative consonants, mainly for /ʃ/ and /ʒ/. For /f/

and /v/ we do not expect significant differences in perception after compression due to their spectral flatness in this frequency range.

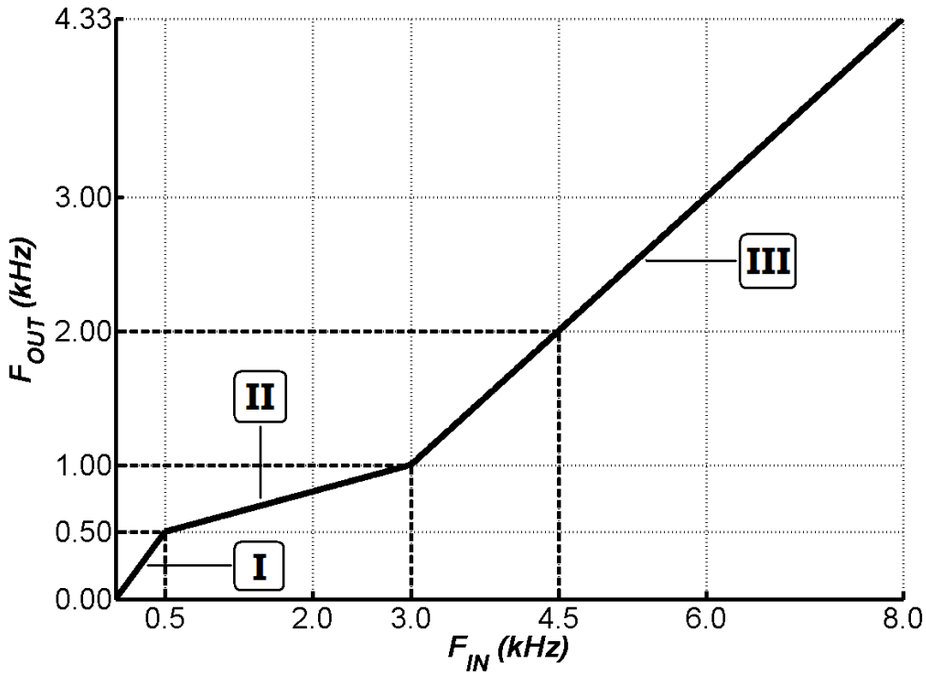

Fig. 8. Frequency compression curve designed and used in this work, with two knee points delimiting regions I, II and III with different compression ratios: (1:1), (5:1) and (1.5:1), respectively

Considering dead regions above 2.0 kHz, for example, the frequency range from 3.0 to 4.5 kHz (first subpart) is transposed to the range from 1.0 to 2.0 kHz. This is actually one of the strengths of this new algorithm: to obtain an effect of frequency transposition by means of a frequency compression curve with two knee points.
In order to facilitate the visualization of the effect of this two knees frequency compression curve on voiced fricatives, these three frequency ranges (I, II and III) are delimited by vertical grey lines in Figure 9.

4.2 Speech test material

We have designed an experiment for simultaneously evaluate consonant discrimination (fricatives in initial syllabic position) and fricative detection (in final syllabic position). In this paper we will focus only in the results of the consonant discrimination test.
The vocabulary for the speech recognition test was formed by the combination of the six most used BP fricative phonemes (/s/, /z/, /f/, /v/, /ʃ/, /ʒ/) with the vowels /a/ and /i/ and a final /s/ in half the situations, forming the set of 24 CV and/or CVC Brazilian Portuguese syllables shown in Table 1. From these, 19 monosyllables form known words in Portuguese, but the remaining 5, marked with [N] in the table, are nonsense syllables. In Table 4, the articulation places and voicing manner of course refer only to the consonant in initial syllabic position, because the final one (when it exists) is always the post alveolar unvoiced fricative /s/.

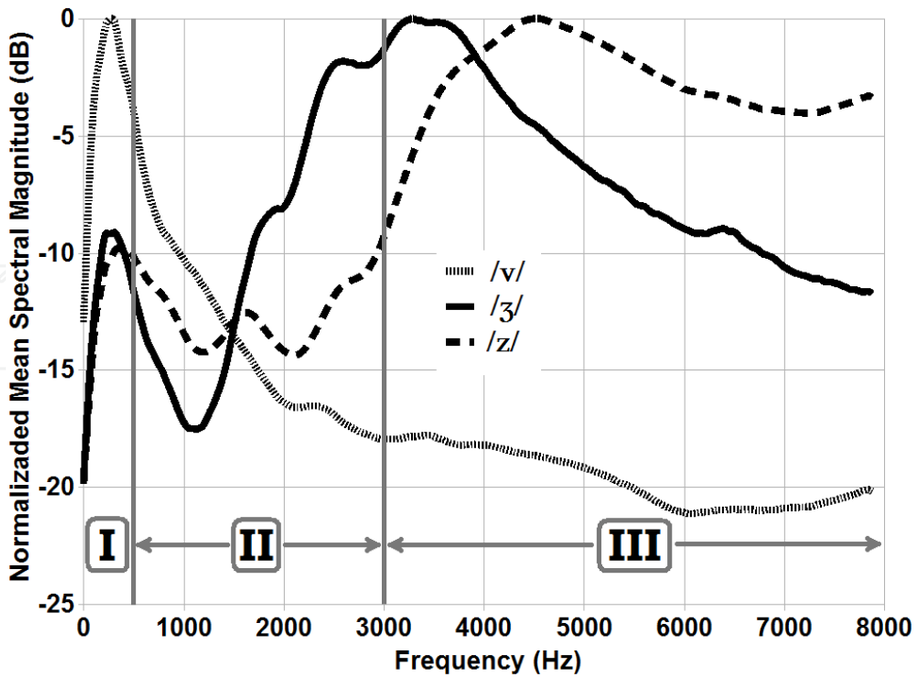

Fig. 9. Average spectral distributions of voiced fricative phonemes /v/ (dotted line), /ʒ/ (continuous line), and /z/ (dashed line), pronounced by male and female speakers. The two knee-point frequencies of the frequency compression curve delimiting the three frequency ranges (I, II and III, see Figure 8) are shown by vertical grey lines.

These words were recorded from 8 speakers, 4 female and 4 male, which pronounced once each different monosyllable. Thus, the original database was composed by 192 utterances, digitized at a 16 kHz sampling rate and stored in separated WAV files.
All utterances were then processed by the same frequency compression/transposition algorithm, which will be presented in the next section. In order to simulate 3 different high-frequency dead regions, both the processed and unprocessed (original) speech material was low-pass filtered at the cutoff frequencies of 1.5, 2.0 and 3.0 kHz. Thus, the final speech database was formed by 3 sets of WAV files, stored in different folders containing 192 processed and 192 unprocessed utterances (control condition) for each simulated dead region.

Labiodentals		Alveolar		Post alveolar	
unvcd.	voiced	unvcd.	voiced	unvcd.	voiced
/fas/	/vas/	/ʃas/	/ʒas/	/sas/ N	/zas/
/fa/	/va/	/ʃa/	/ʒa/	/sa/	/za/
/fis/	/vis/	/ʃis/	/ʒis/	/sis/ N	/zis/ N
/fi/	/vi/	/ʃi/	/ʒi/ N	/si/	/zi/ N

Table 4. Brazilian Portuguese monosyllables used in the fricatives identification and detection experiment.

4.3 Results

Ten normal hearing volunteers, 5 men and 5 women, all Brazilian native speakers between 23 and 30 years old, have been selected to participate. All listeners did the test first for the simulated DR above 2.0 kHz and after for the DR above 1.5 kHz, in order to offer an ever-increasing level of difficulty in the consonant discrimination task. For each listener, the test was applied in a different day for each simulated DR and the average running time spent in the sessions was 67 minutes. The subjects have to choice the written form of the word they have just listened to, in a computer screen. Before deciding, it was necessary to listen at least 3 times to each word, automatically chosen by specific software in a random sequence among 384 utterances (192 processed and 192 unprocessed).

Two-way repeated measures ANOVA were performed over the speech recognition results of the 10 listeners, in terms of correctness (%) in the identification of fricatives in initial syllabic position. The data analysis was done separately according to the speaker gender, using as fixed factors the processing type (COMP versus FILT) and the simulated DR size (above 1500 versus above 2000 Hz). Using Tukey simultaneous tests it was verified that the performance of our original frequency compression algorithm (COMP) was significantly superior (p-value = 0.00005 for female and p-value = 0.007 for male speakers) compared to the simple low-pass filtering (FILT) results, for both simulated dead regions.

4.4 Conclusion

The two-way ANOVA results have shown that our piecewise linear frequency compression curve, applied only to fricative sounds, has presented a statistical significant better performance than low-pass filtering (control condition) in all situations, for both male and female speakers.

5. References

Hicks, B.L., Braida, L.D., Durlach, N.I. (1981). Pitch invariant frequency lowering with non–uniform spectral compression. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 81), 6, 121-124.

Reed, C.M., Hicks, B.L., Braida, L.D., et al. (1983). Discrimination of speech processed by low-pass filtering and pitch-invariant frequency-lowering. J Acoust Soc Am, 74(2), 409-491, ISSN: 0001-4966

Reed, C.M., Schultz, K.I., Braida, L.D., et al. (1985). Discrimination and identification of frequency-lowered speech in listeners with high-frequency hearing impairment. J Acoust Soc Am, 78(6), 2139-2141, ISSN: 0001-4966

Turner, C.W., Hurtig, R.R.(1999). Proportional frequency compression of speech for listeners with sensorineural hearing loss. J Acoust Soc Am, 106(2), 877-886, ISSN: 0001-4966

Simpson, A., Hersbach, A.A., McDermott, H.J. (2005). Improvements in speech perception with an experimental nonlinear frequency compression hearing device. Int J Audiol, 44, 281-292, ISSN 1499-202

Simpson, A., Hersbach, A.A., McDermott, H.J. (2006). Frequency-compression outcomes in listeners with steeply sloping audiograms. Int J Audiol, 45, 619-629, ISSN 1499-202

McDermott, H.J., Dean, M.R.(2000) Speech perception with steeply sloping hearing loss: effects of frequency transposition. Br J Audiol, 34(6), 353-361, ISSN: 0300-5364

Moore, B.C.J., Huss, M., Vickers, D.A., et al. (2000). A test for the diagnosis of dead regions in the cochlea. Br J Audiol, 34, 205-224, ISSN: 0300-5364

Kuk, F. (2007) Critical factors in ensuring efficacy of frequency transposition. Part 1: individualizing the start frequency. Hearing Review [serial online]. 2007, Mar. Available from http://www.hearingreview.com

Kuk, F., Keenan, D., Peeters, H., Lau, C. and Crose, B. (2007) Critical Factors in Ensuring Efficacy of Frequency Transposition Part 2: Facilitating Initial Adjustment. Hearing Review [serial online]. 2007, Apr. Available from http://www.hearingreview.com

Robinson, J.D., Baer, T., Moore, B.C.J.(2007). Using transposition to improve consonant discrimination and detection for listeners with severe high-frequency hearing loss. Int J Audiol, 46(6), pp. 293-308, ISSN 1499-202

Alsaka, Y. A., McLean, B. Spectral Shaping for the Hearing Impaired (1996). Proceedings of the IEEE Southeastcon '96, pp 103-106, ISBN: 0-7803-3088-9, Tampa, FL , USA, 11-14 Apr, 1996

Scanlon, P., Ellis, D.P.W., Reilly, R.B. Using broad phonetic group-experts for improved speech recognition, Ieee Transactions on Audio, Speech and Language Processing, Vol. 15, No. 3 (March 2007), pp 803-812, ISSN: 1558-7916

Gordo A, Iorio MCM. (2007) 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, ISSN: 0034-7299

Baskent D, Shannon RV. (2006). Frequency transposition around dead regions simulated with a noise band vocoder. J Acoust Soc Am. 2006;119(2):1156-63, ISSN: 0001-49

Johnston, J.D. (1998). Transform Coding of Audio Signals Using Perceptual Noise Criteria. IEEE Journal on Selected Areas in Communications, vol. 6,. N.o 2 (February 1988), ISSN: 0733-8716

Russo, I.C.P., Behlau, M. (1993). Percepção da fala: análise acústica do português brasileiro (1st ed.). São Paulo: Lovise, ISBN: 85-7241-404-5

Behlau, M. (1984). Uma análise das vogais do português brasileiro falado em São Paulo: perceptual, espectrográfica de formantes e computadorizada de freqüência fundamental. [PhD thesis]. São Paulo: Universidade Federal de São Paulo

Jesus, L.M.T.(2001). Acoustic phonetics of European Portuguese fricative consonants [PhD thesis] Southampton: University of Southampton

Manrique, A.M., Massone, M.I.(1981). Acoustic analysis and perception of spanish fricative consonants. J Acoust Soc Am, 69(4), pp. 1145-1153, ISSN: 0001-4966

Pen M, Mangabeira-Albernaz PL (1973). Desenvolvimento de testes para logoaudiometria - discriminação vocal. Anales Del Congresso Pan-americano de Otorrinolaringologia y Broncoesofagia, Lima - Peru; 1973, pp. 223-226.

Pereira LD, Schochat E. (1997), Manual de avaliação do processamento auditivo central. São Paulo: Lovise; 1997, ISBN: 8585274441

Nawab, S.H. and Quatieri, T.F. (1998). Short-time Fourier transform. Advanced Topics in Signal Processing, Chapter 6, ed. by J.S. Lim and A.V. Oppenheim, Prentice-Hall, ISBN: 0471446785

Speech Technologies

Edited by Prof. Ivo Ipsic

ISBN 978-953-307-996-7
Hard cover, 432 pages
Publisher InTech
Published online 23, June, 2011
Published in print edition June, 2011

This book addresses different aspects of the research field and a wide range of topics in speech signal processing, speech recognition and language processing. The chapters are divided in three different sections: Speech Signal Modeling, Speech Recognition and Applications. The chapters in the first section cover some essential topics in speech signal processing used for building speech recognition as well as for speech synthesis systems: speech feature enhancement, speech feature vector dimensionality reduction, segmentation of speech frames into phonetic segments. The chapters of the second part cover speech recognition methods and techniques used to read speech from various speech databases and broadcast news recognition for English and non-English languages. The third section of the book presents various speech technology applications used for body conducted speech recognition, hearing impairment, multimodal interfaces and facial expression recognition.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen