Frequency Compression/Transposition of Fricative Consonants for the Hearing Impaired with High-Frequency Dead Regions

Francisco J. Fraga, Leticia P. Costa S. Prates, Maria Cecília M. Iorio

1 Engineering and Social Sciences Center, Federal University of ABC (UFABC), S. André, Brazil
2 Speech and Auditory Disorders Department, Federal University of S Paulo (UNIFESP), SP, Brazil
francisco.fraga@ufabc.edu.br, lepcsprates@hc.ufmg.br, iorio.otor@epm.br

Abstract

In this paper we present the first evaluation results of a carefully designed piecewise linear frequency compression curve, to improve the recognition of fricative consonants, for patients with high-frequency dead regions in the cochlea. Our original frequency compression/transposition algorithm takes into account the average short-time spectrum of the most frequent Brazilian Portuguese fricatives. A specific speech material, composed by twenty four different monosyllables spoken by eight speakers (four male and four female), was recorded for a consonant discrimination test. The algorithm was tested over ten normal hearing listeners with simulated dead regions above 1500 and 2000 Hz. One-way and two-way ANOVA results have shown a statistical significant improvement in the correct identification of fricatives in initial syllabic position.

Index Terms: digital hearing aid, frequency lowering, consonant discrimination.

1. Introduction

Moore et al. [1] called dead regions 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 unbenefficial 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 extensive review on the first proposed frequency lowering methods was done by Braida et al. [2].

All early suggested methods involve signal distortion, more or less noticeable, generally depending on the amount of the frequency shifting. Hicks et al. [3] developed a technique involving pitch–synchronous monotonic compression of the short–term spectral envelope, while at the same time avoiding some problems observed in previous methods. Reed et al. [4] have conducted consonant discrimination experiments on normal hearing listeners using Hick’s frequency lowering scheme. They observed a slight better performance for fricative and affricate sounds if compared to 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.

An overview of more recent studies in frequency compression/transposition is provided by Robinson et al. [5]. 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 cancelled 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 research, 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 paper, the dead regions were simulated by low-pass filtering of the speech material presented to normal hearing listeners.

2. Experiment design

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/, /th/, /n/, /f/, /v/) 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 1, 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/.

<table>
<thead>
<tr>
<th>Labiodentals</th>
<th>Alveolar</th>
<th>Post alveolar</th>
</tr>
</thead>
<tbody>
<tr>
<td>/nas/</td>
<td>/nas/</td>
<td>/nas/</td>
</tr>
<tr>
<td>/la/</td>
<td>/la/</td>
<td>/la/</td>
</tr>
<tr>
<td>/lis/</td>
<td>/lis/</td>
<td>/lis/</td>
</tr>
<tr>
<td>/fi/</td>
<td>/fi/</td>
<td>/fi/</td>
</tr>
</tbody>
</table>

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 two different high-frequency dead regions, both the processed and unprocessed (original) speech material was low-pass filtered at the cutoff frequencies of 1.5 and 2.0 kHz. Thus, the final speech database was formed by two sets of WAV files, stored in different folders containing 192 processed and 192 unprocessed utterances (control condition) for each simulated dead region (DR).

3. Piecewise linear frequency compression

The frequency compression/transposition algorithm was implemented in MATLAB. The signal analysis computations are done in the frequency domain and the processed speech is re-synthesized in the time domain using the well-known overlap-and-add technique [6]. After normalizing the dynamic range of the speech signals (each recorded utterance should have the same rms value), they are 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 DR (low-pass filtering).

3.1. Frame classification criteria

In our algorithm, the frequency compression curve should be applied only over 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 determine the noise-like or tone-like nature of a given speech frame. We develop a method based on the original work of Johnston [7] but with some modifications.

3.2. Frequency compression curve

In the first section of this paper we mentioned that the design of our piecewise linear frequency compression /transposition curve was made taking into account the average short-time spectrum of the BP fricatives.

Those average fricatives spectra were obtained from a different speech database, formed by nonsense CV syllables where the initial consonant was always one of the six BP fricatives (/s/, /z/, /ʃ/, /ʒ/, /ʃ/, /ʒ/) and the vowels were /a/, /i/ and /u/. These 18 different CV syllables were pronounced once by 6 speakers, 3 male and 3 female. Thus, each fricative phoneme has 18 distinct realizations in this database (9 men and 9 by women), which were used to compute the average short-time spectra showed in figures 1 (unvoiced fricatives) and 2 (voiced fricatives).

All speakers are also different from those who pronounced the words that form the speech material used in the phoneme identification and fricative detection test. In order to get significant average spectra, each short-time fricative spectrum was normalized and smoothed by a 16-tap moving average filter prior integrating the sum. As expected, we can see in Figure 1 that the unvoiced fricatives spectral shapes look like its voiced counterpart in Figure 2; the only difference is that the latter have a smaller high-frequency to low-frequency energy ratio. In both figures we can verify that the spectral cues for the 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.

![Average normalized magnitude spectra of the unvoiced fricative consonants](image1)

![Average normalized magnitude spectra of the voiced fricative consonants](image2)

Moreover, based in the data shown in Figures 1 and 2 we can make the following inferences:

a) Any attempting to lower the speech spectrum of post-alveolar fricatives (/s/, /z/) above 3000 Hz should be done carefully, because it can lead to a confusion with the alveolar ones (/ʃ/, /ʒ/);

b) Lowering the spectra of /ʃ/ and /ʒ/ seems to be unprofitable since they are almost completely flat above around 1000 Hz;

c) Just the opposite, according to the figures, bringing the spectral cues of alveolar fricatives /ʃ/ and /ʒ/ present in the frequency range from 2500 to 6500 Hz approximately, to the frequency range between 1000 and 2000 Hz may contribute to differentiate them from other fricatives, considering the presence of a high-frequency DR above 2000 Hz;

d) All fricatives average spectra show a very similar shape from 0 to 2000 Hz. Besides, in every short-time spectrum from figures 1 and 2 we observe a “valley” between 1000 and 2000 Hz. So, it seems that we can apply a great amount of frequency compression to this region, because there are no cues for fricative discrimination in this range of frequencies.

Furthermore, a study of Turner and Hurtig [9] in normal listeners suggests that Compression Ratios (CR) below 0.6 (i.e., compression to less than 60% of the original bandwidth)
can lead to impairment of normal speech recognition. Joining this result from literature with these inferences, we designed the piecewise linear frequency compression curve showed in Figure 3.

**Figure 3: Frequency compression/transposition curve**

### 3.2.1. Rationale for the compression curve

The reasons for the design of each linear part of this original frequency compression curve are resumed below.

I. From 0.0 to 0.5 kHz: To preserve the pitch perception of voiced fricatives, this frequency region remains untouched (CR = 1.0);

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 are no relevant information for fricative discrimination in it;

III. From 3.0 to 8.0 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 /ʃ/ and /ʒ/ we do not expect significant differences in perception after compression due to their spectral flatness in this frequency range. In the case of /s/ and /z/ we hope this compression curve will make a small difference.

### 4. 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. The test was performed in the interior of an audiometric chamber, using the free-field set up in accordance with the ISO 8253-3 standard. The vocabulary words were played through a computer connected to an audiometer, at average intensity level of 65 dB (A scale).

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 choose 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).

For each different pronounced word there was a list of 12 possibilities from the 24-word vocabulary, since the vowel recognition does not matter for the test purposes. After the conclusion of the sessions, the software automatically calculated and stored the results, but no feedback was given to the listeners before, during or after the sessions.

Before each test session there was a mandatory short training section, with the aim of familiarizing the subjects with the test and the computer interface. No feedback was given to listeners in the training sessions too.

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.

We have also carried out a series of one-way ANOVA (COMP versus FILT) over the data, which results in the recognition rate of fricatives in initial syllabic position averaged across the 10 listeners can be observed in figures 4 and 5, for the simulated dead regions above 1500 and 2000 Hz, respectively. In the figures, three stars indicate a significance level of p < 0.001, two stars show a significance level of p < 0.01, and one star indicates a significance level of p < 0.05.

The results of both two-way and one-way ANOVA for fricatives with different articulation places were computed separately, but the voiced and unvoiced ones were grouped. We did it because the confusion between fricatives with the same articulation places was insignificant for all ten subjects (listeners). It was a predictable result since our algorithm does not affect low frequency patterns, which are responsible for the differentiation between voiced and unvoiced fricatives.

### 5. Discussion and conclusions

In the series of one-way ANOVA it was verified that the speaker gender influenced the algorithm performance in different directions, depending on the DR. For simulated DR above 1500 Hz, there was a better performance and significance in the overall recognition for male speakers (~63%, p < 0.01) relative to the female speakers (~54%, p < 0.05). Just the opposite, in the case of simulated DR above 2000 Hz, the performance for female speakers was better (~73%, p < 0.001) than for male speakers (~67%, p > 0.05).
We can see in this latter case that for the fricatives group “/s/, /z/” the signals in control condition were significantly better recognized than the frequency compressed ones. Both results are due to the fact that male speakers have longer vocal tracts, so their vocal tract resonant frequencies are lower than the female ones. It caused the effect that sometimes the “male” compressed fricatives from group “/s/, /z/” are confused with fricatives from group “/S/, /Z/”. Thinking in the final application, maybe the solution for this problem will be to adapt the piecewise linear frequency compression curve according to the DR of each patient.

Moreover, we conclude that these overall good results were due to the high improvement in performance for the correct identification of initial syllabic fricatives from group “/f/, /v/”. Simultaneously, in most situations our algorithm does not change significantly the performance for the other 2 fricatives groups, which explains the improvement in the performance of fricatives discrimination as a whole.

Finally, considering that the average spectra of the BP unvoiced fricatives used in the test are very similar to the average spectra of the same fricatives in Spanish [10], this frequency compression scheme probably will be effective if applied to consonants from other languages too.

6. References