Quantitative Assessment of the Use of Continuous Wavelet Transform in the Analysis of the Fundamental Frequency Disturb of the Synthetic Voice

Adriane Parraga\textsuperscript{a,\,*}, Milton A. Zaro\textsuperscript{b}, Adalberto Schuck Jr.\textsuperscript{c}

\textsuperscript{a} M.Sc. Student CPGEE-DELET-UFRGS Porto Alegre, RS, Brazil
\textsuperscript{b} PROMEC-UFRGS, SBEB, Porto Alegre, RS, Brasil
\textsuperscript{c} DELET-UFRGS, IEEE-EMBS, SBEB, Porto Alegre, RS, Brazil

Abstract

The aim of this work is to investigate quantitatively the capability of the Continuous Wavelet Transform (CWT) as a tool to estimate (calculate) Jitter and Shimmer, assessing the error between these indices calculated in each Wavelet decomposition and the ones for the original signal, for several dilatation levels. It were generated 2 synthetic vowel /a/ with the fundamental frequency of 120Hz for male and 220Hz for female, by an autoregressive 22 coefficients all-pole model, and introduced Jitter and Shimmer on the signal using five different percent variations. The signals were decomposed by CWT in 8 levels of dilatation (1, 2, 4, 8, 16, 32, 64 and 128), using the Mexican Hat, Meyer and Morlet real bases. Jitter and Shimmer were calculated for the original signals and all the 8 levels of decompositions and then the errors between the indices in the decompositions and the original signals were calculated. It can be concluded that CWT can be used as a tool for pre-processing the signal to measure Shimmer preferentially, and Jitter, instead of using the original signal to do that. The Mexican Hat base provided the lowest errors for Shimmer analysis, where the best dilatation level was 8 (error below 0.1%). In addition, the errors associated with Shimmer index, in general, are lower than the ones associated with Jitter index.

Keywords: Speech processing, Continuous Wavelet transform, Acoustic analysis of voice, Fundamental frequency analysis.

1. Introduction

Several acoustical indices like Jitter and Shimmer analysis, Harmonic-to-Noise ratios, Breathiness and others were proposed to assess in an objective form and to quantify the perceptual aspects of the dysphonic voice [1].

The Wavelet Transform, Discrete (DWT) and Continuous (CWT) has been used in Biomedical Signal Processing in different fields applications [2], mainly in EEG and ECG signal analysis and feature extraction, and in voice analysis [3],[4],[5].

In [6], the authors had shown that the use of DWT with Haar base in vocal fold assessment presented some inherent problems in the low resolution levels coefficients (low frequency part of the signal), caused by aliasing and phase sensitivity [7],[8]. This feature can cause problems in a kind of Jitter and Shimmer (voice’s fundamental frequency perturbation analysis) analysis with these coefficients.

One of the suggestions, attempting to solve the problems of DWT, was the use of CWT for the signal decomposition. In [9], the authors showed that the use of CWT, with 4 different decomposition bases, 3 real and 1 complex, solved the above mentioned problems.

It was also concluded that, qualitatively speaking, using the same dilatation level as in DWT, the use of the Mexican Hat base gave the best results, because it’s graphics apparently are smooth and have a better resolution in the high dilatation levels, independently of the patient’s fundamental frequency and disease.

The aim of this work is to investigate quantitatively the capability of the CWT as a tool to estimate (calculate) Jitter and Shimmer, assessing the error between these indices calculated in each Wavelet decomposition and the ones for the original signal, for several scales.

The coefficients of CWT are calculated by [10]:

\[ \text{CWT}(a,b) = \frac{1}{\sqrt{a}} \int x(t)\psi(\frac{t-b}{a}).dt \]  

(1)

where \( \psi(t) \) is the Mother Wavelet, \( a \) is the dilation parameter and \( b \) the translation in time. The Wavelet basis function \( \psi_{a,b}(t) \) is a family of \( \psi(t) \) prototype function.
Jitter and Shimmer are calculated in the original signal \(x(t)\) and in the coefficients of the CWT(a,b) for several discrete values of \(a\). The signal is decomposed using CWT of the Wavelet Toolbox Matlab\textsuperscript{TM} v.5.3.

The measure was done for 3 real bases (Meyer, Morlet and Mexican Hat), comparing which one gives the lowest errors.

2. Materials and Methods

In order to have controlled fundamental frequency perturbations set of voices, 2 synthetics emitted sustained vowel /a/ were generated, using the same approach as in [5]. It were used 8 healthy adult male voices and 4 healthy adult female voices separately to produce in Matlab\textsuperscript{TM} v.5.3 the coefficients of 2 autoregressive (AR) all-pole models. The order used for both models was 22 (coefficients), and the command used to calculate them was Matlab Signal Processing Toolbox’s LPC (Linear Predict Code). The estimated model \(\tilde{s}[n]\) for the true speech signal in the time domain \(s[n]\) is given by [11]:

\[
\tilde{s}[n] = \frac{1}{a_0} \left\{ b_0 d[n] - \sum_{i=1}^{22} a_i \tilde{s}[n-i] \right\}
\]  

(1)

The mean LPC coefficients obtained for women are presented in Table 1 (where \(a_0\) and \(b_0\) are 1)

<table>
<thead>
<tr>
<th>(a_1)</th>
<th>(a_2)</th>
<th>(a_3)</th>
<th>(a_4)</th>
<th>(a_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7635</td>
<td>0.5419</td>
<td>0.4022</td>
<td>0.0952</td>
<td>0.0511</td>
</tr>
<tr>
<td>0.2996</td>
<td>0.1062</td>
<td>0.0935</td>
<td>0.1217</td>
<td>0.0667</td>
</tr>
<tr>
<td>0.1999</td>
<td>0.1709</td>
<td>0.1537</td>
<td>0.1358</td>
<td>0.0193</td>
</tr>
<tr>
<td>0.1144</td>
<td>0.1795</td>
<td>0.0061</td>
<td>0.0486</td>
<td>0.1178</td>
</tr>
<tr>
<td>0.0362</td>
<td>0.1019</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 – Average LPC coefficients for women

The mean LPC coefficients obtained for men are presented in Table 2.

<table>
<thead>
<tr>
<th>(a_1)</th>
<th>(a_2)</th>
<th>(a_3)</th>
<th>(a_4)</th>
<th>(a_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.9943</td>
<td>1.1012</td>
<td>0.0598</td>
<td>0.1054</td>
<td>0.3533</td>
</tr>
<tr>
<td>0.0296</td>
<td>0.1062</td>
<td>0.0935</td>
<td>0.1217</td>
<td>0.0667</td>
</tr>
<tr>
<td>0.2199</td>
<td>0.1709</td>
<td>0.1537</td>
<td>0.1358</td>
<td>0.0193</td>
</tr>
<tr>
<td>0.1144</td>
<td>0.1795</td>
<td>0.0061</td>
<td>0.0486</td>
<td>0.1178</td>
</tr>
<tr>
<td>0.0362</td>
<td>0.1019</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 – Average LPC coefficients for men

The input signal \(d[n]\) for the model is a delta train whose periodicity is given by the fundamental period \(T_p=1/f_0\) . For the adult female, the fundamental frequency average used was \(f_0=220\) Hz and for the adult male was \(f_0=120\) Hz [12]. Shimmer was introduced in the input signal by adding and subtracting alternately a percent value of the amplitude of the delta train. For the Jitter, in the input signal were produced pulses separated by a period in which were introduced alternately positive and negative variations.

Because there are many different ways proposed to calculate Jitter and Shimmer indices and in order to achieve a standard index for comparison, it has been chosen to calculate the indices using the average of the measurement (fundamental frequency and amplitude) and its standard deviation.

The values used for the simulations were 0.5, 1, 2, 5 and 10% for Jitter and 5, 10, 15, 20 and 25% for Shimmer. These percentages represents typical values calculated from a set of 60 normal and with some kind of disease in their vocal tract (voice's fundamental frequency perturbation) subjects and then chosen some intermediate values.

The data voices were sampled at \(f_s=22050\) Hz, the same as the synthetic vowels. Each generated vowel has the size of 8192 points. The synthetic voices for female and male are in Fig. 1 and Fig. 2, respectively.

![Fig. 1 – Synthetic vowel /a/ for male](image1)

![Fig. 2 - Synthetic vowel /a/ for female](image2)

The average and the standard deviation of Jitter and Shimmer were calculated using the software SAD2\textsuperscript{TM}, developed at LMM-PROMEC-UFRGS (at [http://www.ufrgs.br/lmm](http://www.ufrgs.br/lmm) for download).

The Shimmer and the Jitter indices are the ratio of the standard deviation (std) by the average (mean) of the respective measurement in percentage, as in equation 2 and 3:

\[
\text{Jitter}(x) = \frac{\text{std}(f_0)}{f_0} \times 100
\]  

(2)

\[
\text{Shimmer}(x) = \frac{\text{std}(A_m)}{A_m} \times 100
\]  

(3)
where $x$ is the signal, $\bar{f}_0$ is the mean of the fundamental frequency and $\text{std}(f_0)$ is its standard deviation. The $\bar{A}_m$ is the mean of the fundamental frequency period peak amplitude of signal $x$ and $\text{std}(A_m)$ its standard deviation.

Then, a CWT analysis is performed by Matlab™ v.5.3, with its Wavelet Toolbox v. 1.2, over all signals. It was used 3 different bases for decomposition (Morlet, Meyer and Mexican Hat) with 8 different dilatation scales: $c_1=1$, $c_2=2$, $c_3=4$, $c_4=8$, $c_5=16$, $c_6=32$, $c_7=64$ and $c_8=128$, the same used in [6] and [9].

The CWT decompositions and the real bases of decomposition chosen are present in this toolbox. The number of points and the effective “compact” support used for each base, are the standard values for the Toolbox: 256 points and effective support of $[-4,4]$ for Morlet, $[-8,8]$ for Meyer and, $[-5,5]$ for Mexican Hat, as shown in Fig. 3, 4 and 5 respectively.

For each level of the decomposed signal, the Jitter and Shimmer indices were calculated using equation 2 and 3. With all the results, the errors indices were measured between the signal and the signal decomposed in each dilatation level, as in equation 4 and 5.

$$JE(\%) = \left( \frac{\text{Jitter}(x) - \text{Jitter}(c_n)}{\text{Jitter}(x)} \right) \times 100$$ (4)

$$SE(\%) = \left( \frac{\text{Shimmer}(x) - \text{Shimmer}(c_n)}{\text{Shimmer}(x)} \right) \times 100$$ (5)

where $JE(\%)$ is the Jitter error and $SE(\%)$ is the Shimmer error in percent, $\text{Jitter}(x)$ and $\text{Shimmer}(x)$ are the indices of the original $x$, the standard signal considered, and $\text{Jitter}(c_n)$ and $\text{Shimmer}(c_n)$ are the coefficients in each decomposition level of CWT, denoted by $c_n$, where $n$ are the dilatations levels used.

### 3. Results

With the exception of the coefficients $c_7$ and $c_8$, all the other results obtained are summarised in table 3. It shows the mean of $JE(\%)$ and $SE(\%)$ for women and men for all the bases and the decompositions used.

<table>
<thead>
<tr>
<th>Base</th>
<th>Scales</th>
<th>$JE(%)$</th>
<th>Scales</th>
<th>$SE(%)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morlet</td>
<td>C1</td>
<td>C2</td>
<td>C3</td>
<td>C4</td>
</tr>
<tr>
<td>0.5</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>X</td>
<td>20</td>
<td>X</td>
</tr>
<tr>
<td>10</td>
<td>X</td>
<td>25</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mexican Hat 0.5</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Mexican Hat 1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>10</td>
</tr>
<tr>
<td>Mexican Hat 2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>15</td>
</tr>
<tr>
<td>Mexican Hat 5</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>20</td>
</tr>
<tr>
<td>Meyer 0.5</td>
<td>O</td>
<td>X</td>
<td>5</td>
<td>O</td>
</tr>
<tr>
<td>Meyer 1</td>
<td>O</td>
<td>X</td>
<td>10</td>
<td>X</td>
</tr>
<tr>
<td>Meyer 2</td>
<td>O</td>
<td>O</td>
<td>15</td>
<td>X</td>
</tr>
<tr>
<td>Meyer 5</td>
<td>O</td>
<td>X</td>
<td>20</td>
<td>X</td>
</tr>
<tr>
<td>Meyer 10</td>
<td>O</td>
<td>X</td>
<td>25</td>
<td>X</td>
</tr>
</tbody>
</table>

Legend:
- **X**: Error > 10%
- **O**: Error < 0.1%
- **0.1% < Error < 1%**
- **1% < Error < 5%**
- **5% < Error < 10%**
- **Error = 0%**
The coefficients $c_7$ and $c_8$ aren’t included because they have given high percent errors for Jitter and Shimmer (errors around 200%), in all bases and scales here used.

Some details that were hidden in the summarising process should be commented here:

a) For Morlet base, in the Shimmer analysis, the errors were below 3% for women and below 1.5% for men. In the Jitter analysis, for a Jitter value of 0.5%, the errors were very high for all levels and gender. In addition, the average error decreases with the increase of the Jitter value, both for men and women;

b) For the Mexican Hat base, the errors for Jitter of 0.5% were higher than others Jitter values, but still are lower in comparison to the other 2 bases used. The errors also decreases with the increase of Jitter values, both for men and women. The best results for Jitter were obtained for the levels $c_1$, $c_2$, $c_3$ and $c_4$. For Shimmer, in the coefficients $c_5$, $c_6$ and $c_7$ the errors were below 0.4% for women, and for men the coefficients $c_1$,$c_2$ were below 1%;

c) For Meyer base, the errors for Jitter of 0.5% were higher than the other variations of Jitter. Moreover, the value of errors decrease with the increase of the Jitter value, both for men and women.

4. Conclusions

Analysing all the results, it can be observed that:

a) For Jitter analysis, the error decreases as the Jitter value introduced increases, both for men and for women and for all 3 bases used. However, it was not true for Shimmer, which gave smooth results independently of the Shimmer variation values;

b) For a Jitter analysis, in general, it can be noted that the results with errors below 1% are obtained only in Jitter values above 5 %;

c) The coefficients $c_1$ and $c_8$ for all the bases used have given high percent errors for Jitter and Shimmer (errors around 200%). This happens because in scales 64 and 128, the Wavelet decomposition process filters the fundamental frequency of the original signal, making it hard to be detected in these coefficients;

d) For Shimmer analysis, comparing the bases here used, the level $c_4$ of the Mexican Hat was the one that gave the best results (errors below 0.1%). For the first 2 levels, the performance of the bases Morlet and Mexican Hat are similar; from $c_3$ to $c_5$, the Mexican Hat has a better performance, and for the $c_6$, Morlet is better;

e) In comparison with the other bases, the Meyer base is the one with the worst performance for Jitter, in any scale, because there are more results with errors above 5% than the others;

f) In general, quantitatively speaking, the CWT of the signal is more appropriate for measuring Shimmer than Jitter, because in Jitter analysis there are more coefficients with error above 5%. In addition, in a Shimmer analysis, the Mexican Hat is the base that has given the best results for it (levels $c_3$, $c_4$ and $c_5$). This was also reported in [9], in a qualitative way.

It can be concluded that CWT can be used as a tool for pre-processing the signal to measure Shimmer preferentially, and Jitter, instead of using the original signal to do that.

References


[7] Antonini M., Barlaud M., Mathieu P. and Daubechies I:


