Pitch Detection and Voiced/Unvoiced Decision Algorithm based on Wavelet Transforms

Léonard Janer, Juan José Bonet
Dept. TSC Universitat Politècnica de Catalunya
08034 Barcelona, Spain
Email: leonard@gps.tsc.upc.es

Eduardo Lleida-Solano
GTC Dept. IEEC Centro Politécnico Superior de Ingenieros
50015 Zaragoza, Spain
Email: lleida@mcps.unizar.es

ABSTRACT

An improvement of an existing Pitch Detection Algorithm is presented in this paper. The solution reduces the computational load of its precedent algorithm and introduces a voiced/unvoiced decision step to reduce the number of errors. The efficiency of this improved system is tested with a semi-automatically segmented speech data base according to the information delivered by an attached laryngograph signal. The results show its periodicity detection.

1. Introduction

In the last years, Wavelet Transforms [WT] have been intensively applied in different pitch detection algorithms [1][2][3][4][5]. Most of those algorithms are based on Dyadic Wavelet Transforms (DyWT), what it means a constant dilation factor equal to 2. The algorithm we present in this paper is a modification of an existing one based on a Modulated Gaussian Wavelet Transform based Speech Analyser (MGWTSA) [3]. This is a system with 17 bands distributed in frequency along a Bark scale (logarithmic frequency distribution) that determines pitch period values sample per sample on continuous speech.

The algorithm we present in the following two sections of this paper introduces two important points with respect to the previous solution:

1. The first improvement is the reduction in the computational load of the system, without an important loss in the performance rate of the global system, with an associated increment in the number of errors in the Pitch Detection System. In reducing the number of scales taken into account in the algorithm, we have introduce more Gross Pitch Errors, or errors above 1 msec, in the estimated period. Opposite to this effect the number of voiced segments of speech detected as unvoiced has been reduced (drastically for male speakers).

2. The second advantage is a voiced/unvoiced segment decision step at the end of the system. The consequence of this system is a reduction in the Gross Error Pitch Rate.

Finally, in the fourth section the results of this new version of the algorithm are presented with the same speech data base used in the original work [3][6]. In the last section we detail the conclusions of this work.

2. Pitch Detection Algorithm

2.1. First Algorithm Implementation

As a first stage for a better understanding of the system presented in this paper we must explain the basic ideas introduced in its original version. That is a system [3] working with 17 bands, that decomposes the speech signal into different scale versions of the signal followed by a maxima detection step and a peak confirmation and correlation block to get the pitch period estimation sample by sample (see Figure 1). The system works both in the time domain finding the maxima for all its scales, and in the frequency domain joining them to build a first Pitch Period Estimation, that will be post-processed with three alternative algorithms, based on a distortion measure computed at the maxima detection stage, or a mode/median decision block.

2.2. New Pitch Detection Algorithm

The new version of the algorithm presented in this paper, takes its fundamentals from the same basic: a time-frequency evaluation of the speech signal, taking advantages of the MGWTSA. In the time domain the system finds the maxima of the considered scales of the MGWTSA. In this new release we work only with 5 scales, taking only the lowest scales of the system, the ones where the periodicity information is the highest. The reason is an important computational load reduction, with a few reduction in the performance rates. The second main difference of the system is the binary threshold introduced at the maxima detection stage. The threshold is evaluated by the following rules:

1. Each time a maximum is found we calculate the averaged value for the surrounding maxima (the 5 maxima around it, with a maximum time interval allowed). This value will be referred as Decision Value.
**Input Speech Signal (20 KHz.)**

**Modulated Gaussian Wavelet Transform based Speech Analyser (MGWTSA)**
17 scales with BARK distribution

**Maxima Detector (with a different prefixed threshold per scale)**

**Peaks Confirmation (with a Distortion Measure evaluation)**

**Peaks Combination onto the first Pitch Period Sequence**

**Window centred at each peak (with its width adapted to the Pitch Period value)**

**Three different post-processing solutions:**
1. Distortion Measure
2. Mode Computation
3. Median Computation

**Median filter Post-Processing**

![Figure 1: Block Diagram for the complete system in the original algorithm](image)

2. The threshold is imposed according to the **Decision Value**. If the value is higher than a prefixed number, then the threshold takes its minimum value, if not the threshold goes up to its maximum value.

3. Then the maximum is compared with this variant threshold.

Once the maxima have been detected, the algorithm works in the same way as it worked in its first implementation but only with 5 bands of the MGWTSA, as shown in Figure 8. The objective of this new version of the system is the improvement in the threshold decision step. In the previous algorithm, the threshold took a preestablished value, and was too strong in some low energy voiced segments. In the version presented in this paper the threshold takes into consideration the energy of the speech signal around a window centred at each maximum localization.

**Figure 2: Speech Segment “/sun where dis/” from the sentence “the northwind and the sun were disputing which was the stronger when a traveller came along, wrapped in a warm cloak”’ pronounced by a male speaker.**

**Figure 3: Reference Pitch Period Labels and Evaluated Pitch Period Labels**

3. **Voiced/Unvoiced Decision Algorithm**

As a new feature for this Pitch Period Estimation Algorithm, we have introduced a Voiced/Unvoiced Decision Algorithm, that is based on the same system, and modifies some of the errors produced by the PDA. Its fundamentals are the following ones:

1. Once the maxima have been confirmed by the **Peaks Confirmation Stage**, and joined together by the **Peaks Combination Stage**, the **Voiced/Unvoiced Stage** is introduced in the system.

2. For all the maxima, we evaluate the number of maxima around its position within a window as width as its Pitch Period estimated value.

3. With the number of maxima we take a binary decision. All the points where we have more than 3 maxima are labelled as voiced, and their Pitch Period will be the one computed by the system, and those with a number below 3 will be labelled as Unvoiced segments of speech.

The system takes into consideration the fact that where the speech will be voiced, approximately all the bands will find a

**Figure 4: MGWTSA output as a scalogram**
maximum, above its own threshold, and then the number of maxima at any voiced segment will be near 5 (the number of considered scales from the MGWTSA). Where the signal will be unvoiced, only one or two of all the scales as a maximum will find a peak, and then the voiced/unvoiced decision step will eliminate those erroneous pitch period values. Finally, the transitions, between a voiced and an unvoiced segment or vice versa, will have a number of maxima, usually, higher than 2, and the system will label them as voiced/unvoiced in function of the next peak position (take in mind the 5 maxima window for threshold calculation).

4. Evaluation Results

We will show the results for the original algorithm, the new PDA version and finally for the voiced/unvoiced decision system, in Tables 1, 2 and 3.

The speech database is composed by 5 male and 5 female speakers, with 40 seconds per speaker of English sentences sampled at 20 KHz. The pitch period reference signal is computed with a semi-automatical system based on the associated laryngograph signal [7].

In all the algorithms we present the following statistics: 

**GPER** (Gross Pitch Error Rate, or errors greater than 1 msec.), **PDER** (Pitch Doubling Error Rate, or female periodicity detected as male periodicity), **PHER** (Pitch Halving Error Rate, or the male periodicity detected as a female one), **VuVER** (Voiced/Unvoiced Error Rate), **FPER** (Fine Pitch Error Rate, or errors in pitch detection below 1 msec.) and **PR** (Performance Rate, or percentage of Fine Errors and No-errors in pitch period estimation).

<table>
<thead>
<tr>
<th>Evaluation Results for Distortion Measure Solution</th>
<th>GPER (%)</th>
<th>PDER (%)</th>
<th>PHER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>5.30</td>
<td>0.00</td>
<td>1.22</td>
</tr>
<tr>
<td>Female</td>
<td>2.30</td>
<td>0.25</td>
<td>0.03</td>
</tr>
<tr>
<td>VuVER (%)</td>
<td>6.63</td>
<td>(3.8%)</td>
<td><strong>83.65</strong></td>
</tr>
<tr>
<td>Male</td>
<td>4.2%</td>
<td>(2.9%)</td>
<td><strong>91.95</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluation Results for Mode Evaluation Solution</th>
<th>GPER (%)</th>
<th>PDER (%)</th>
<th>PHER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>5.16</td>
<td>0.00</td>
<td>1.38</td>
</tr>
<tr>
<td>Female</td>
<td>3.15</td>
<td>0.25</td>
<td>0.03</td>
</tr>
<tr>
<td>VuVER (%)</td>
<td>6.63</td>
<td>(3.8%)</td>
<td><strong>84.74</strong></td>
</tr>
<tr>
<td>Male</td>
<td>4.2%</td>
<td>(2.9%)</td>
<td><strong>91.95</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluation Results for Median Computation Solution</th>
<th>GPER (%)</th>
<th>PDER (%)</th>
<th>PHER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>5.12</td>
<td>0.00</td>
<td>1.30</td>
</tr>
<tr>
<td>Female</td>
<td>3.90</td>
<td>0.38</td>
<td>0.25</td>
</tr>
<tr>
<td>VuVER (%)</td>
<td>6.85</td>
<td>(2.7%)</td>
<td><strong>77.13</strong></td>
</tr>
<tr>
<td>Female</td>
<td>3.99</td>
<td>(2.7%)</td>
<td><strong>90.81</strong></td>
</tr>
</tbody>
</table>

Table 1: Pitch Period Detection Results with the original algorithm

The results for the same speech database as in the original work show a better performance for male speakers but a worse periodicity detection for female speakers.

With the new algorithm the difference between male and female speakers has been reduced, as a first step to get a speaker independent pitch detection algorithm.

Future works for this PDA are the improvement in the voiced/unvoiced decision step to reduce the number of Gross Pitch Errors.

6. Acknowledgments

This work has been supported by the Spanish Ministry of Education and Sciences (MEC) grant TIC95-0884-C04. We would like to thank Ignasi Esquerda, Eric Mouset and Juan.
Speech Segment “/traveller came/” from the sentence “the north wind and the sun were disputing which was the stronger when a traveller came along, wrapped in a warm cloak”, pronounced by a male speaker.

Figure 5: Reference Pitch Period Labels and Evaluated Pitch Period Labels

Luis Navarro for their contribution on this work.

7. REFERENCES


