Hybrid Electroglottograph and Speech Signal based Algorithm for Pitch Marking

Hussein Hussein and Oliver Jokisch

Laboratory of Acoustics and Speech Communication
Dresden University of Technology, 01062 Dresden, Germany
{Hussein.Hussein,Oliver.Jokisch}@ias.et.tu-dresden.de

Abstract

Pitch marking is very significant in speech signal processing. In a text-to-speech (TTS) system based on the Time-Domain Pitch-Synchronous Overlap-Add (TD-PSOLA) method, robust estimation of pitch marks (PM) is especially important to the modification of the pitch and time scale of speech signals in order to match it to that of the target speaker. The aim of this paper is to improve the accuracy of automatic Pitch Mark Algorithms (PMA). Therefore, we propose a hybrid method for pitch marking that combines the advantages of the Electroglottograph (EGG) and the speech signals. We evaluate this hybrid algorithm for pitch marking against pitch mark algorithm used by Praat program [1]. The results of the evaluation indicate that the suggested method provides better performance than PMA based on EGG signal or speech signal.

Index Terms: Speech synthesis, Pitch marking

1. Introduction

Speech signal is one of the most important and natural means of communication between people. The signal generated by the vocal folds is the basis for speech synthesis based on TD-PSOLA method. The measurement of the glottis opening and closing in an indirect method from the outside of the neck would lead to obtain the source signal. The TD-PSOLA algorithm is well-known in speech synthesis systems for the modification of the pitch and time scale of speech segments. Accurate estimation of pitch marks for each pitch period in TTS system based on the TD-PSOLA method is necessary to generate optimal quality of synthetic speech.

The electroglottograph is an inexpensive and noninvasive sensor for studying the vibration of vocal folds. The EGG records electrical impedance in the glottis during speaking between two electrodes which are situated on the surface of the throat. Pitch mark is defined as the beginning of a pitch period. Some of the common pitch marking methods include maximum or minimum point in the speech signal, zero-crossing, which precedes or follows an extremum in the speech signal and Glottal Closure Instant (GCI) of the EGG signal. GCI i.e. the exact moments of excitation of the vocal tract during the phonation of vowels and semi-vowels, is generally chosen as the place of the marks. EGG and speech signal are recorded at the same time as the voice signal. The GCI of the EGG signal corresponds to the most negative-going peak of the speech signal. Because of the importance of pitch marking, numerous techniques and methods for pitch marking have been suggested over the past two decades. There are two major methods for determination of pitch marks. Firstly, direct from the speech signal [1][2][3][4][5]. Secondly, from the EGG signal [6].

There are different reasons which make the speech processing task difficult. For example, irregular excitation signal, a large number of time structures in speech signal and the fact that speech is not stationary. Therefor, no algorithm for the automatic pitch marking works error free [7]. The EGG signal makes the detection of the open and closed glottal segments in each period easier. Thus, pitch marks can be more precisely found in the EGG signal [8][9]. In contrast, the EGG signal possesses the following disadvantages: a very low level signal and the loss of some parts of the signal. The problems for it are: the movement of the larynx during speaking, the incomplete contact between electrodes and skin, the individual anatomy of the larynx and the neck, and the hairiness by male speaker.

![Figure 1: EGG missing region between two voiced speech units](image-url)

It is noticed by comparison between EGG and speech signal that the vocal cords do not open completely at the beginning and at the end of a vowel. It is possible that the vocal cords continue to vibrate without generation any important acoustic energy. There are a number of segments that are assigned as voiced by the EGG signal-based algorithm, but are assigned as voiceless by the speech signal-based algorithm.

Our purpose is to prove that the hybrid algorithm for pitch marking unites the advantages of the EGG and the speech signals, which can lead to performance improvements over PMA based on “speech-only” or “EGG-only”.

In Section 2 of this paper, we present the detailed strategy of the hybrid algorithm for pitch marking. Section 3 provides an overview of database, existing algorithms for pitch marking and evaluation criteria. Section 4 presents an evaluation of the results. Finally, Section 5 contains the conclusion.
2. Hybrid pitch mark algorithm

The hybrid algorithm for pitch marking, which was implemented in [10], is transferred and modified in this paper. The algorithm is based on the division of EGG signal into two regions: stable and unstable EGG regions. Pitch marks are extracted from the EGG signal in the stable EGG regions. On the other hand, for the lost parts and EGG regions with low amplitude, pitch marks are extracted from the speech signal. The extracted pitch marks from the speech signal, which correspond to unstable EGG regions, are transferred and adapted to the extracted pitch marks from the EGG signal. The framework of the algorithm is shown in Figure 2.

\[ \Delta t \approx (0.15/343) + (0.20/343) \pm 0.00015 \quad (1) \]

where the speed of sound is about 343 m/s. Therefore

\[ \Delta t \approx 1 \pm 0.2 \text{(ms)} \quad (2) \]

4. Determination of EGG Missing Region: If the length of the region between two sequential EGG-based pitch marks is at least 10ms, then this region is regarded as EGG missing region. There are basically three cases of EGG missing regions:

- The EGG missing region is located between two vowels (see the Figure 1).
- The EGG missing region is found before a vowel.
- The EGG missing region is determined after a vowel.

5. Insertion of Speech-based PM in EGG Missing Region: The Speech-based PM, which correspond to EGG missing region, are transferred to the EGG-based PM in this region. The Speech-based PM do not always agree with the GCI of the EGG signal, therefore the Speech-based PM, which are superimposed with the EGG-based PM in the EGG missing region, are adjusted with a shift. The EGG-based pitch marks are considered as reference for pitch marking. Thus the shift is equal to the average value of the shift between the existing synchronized pitch marks in the two signals. The shift must not be more than 0.5 ms [12]. The pitch marks are called in this case overlapped PM.

6. Checking the overlapped PM: Sometimes the speech signal based PMA mark pitch marks in the voiceless segments or low amplitude levels, e.g. at the breathing. Therefore the transferred speech-based PM are examined by an energy threshold method. The speech-based PM in the missing EGG regions, whose energy (3) falls below the determined threshold value, are ignored. Since voiceless segments are low energy segments, the speech signal is smoothed with the "sliding average value" (4), in order to eliminate voiceless and very noisy segments. The threshold is computed as 5% of the average value of the energy of the smoothed speech signal. The Figure 3 shows the short time energy for the speech signal and the smoothed speech signal (N=160 samples for short time energy and N=20 samples for the smoothing).

\[ x_{\text{ener}}(k) = \sum_{i=k-N}^{k} x^2(i) \quad (3) \]

\[ x_{\text{aver}}(k) = \frac{1}{N} \sum_{i=0}^{N-1} x(k+i) \quad (4) \]

3. Experiments

3.1. Database

In the European Union project TC-STAR (Technology and Corpora for Speech to Speech Translation) [12], language resources for UK English, Spanish and Mandarin were produced. In this paper the experiments are implemented only for the UK English. A platform with 3 synchronized channels were used. The channel 1: large membrane microphone with a distance of 20 cm from the speaker, channel 2: laryngograph and channel 3:
The performance of the pitch mark algorithms with the subset of TC-STAR database is shown in tables 1 to 4. The results with regard to the success rate showed that the hybrid algorithm get the best results (SR=95.67\% for speech based reference and SR=98.77\% for EGG based reference). Taking accuracy into consideration, the CHFA PMA get the best results for speech based reference (ACC=92.61\%) and the EGG PMA get the best results for EGG based reference (ACC=98.28\%).

The proposed algorithm with EGG based reference obtained better results than those with speech based reference, because the hybrid algorithm based basically on pitch marking from the EGG signal in the stable EGG regions (see tables 1 and 2). The results in tables 1 and 3 showed that the pitch mark algorithms achieves higher success rates and accuracies with female speech than with male speech.

Figure 4 shows the success rates of the pitch mark algorithms, which were evaluated in this paper, and the success rates of algorithms, which were read in the literature. The hybrid algorithm provides performance improvements over speech signal based PMA and EGG signal based PMA.

### Table 1: Evaluation results of PMA for speech based reference (female speaker).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Ref</th>
<th>Test</th>
<th>SR(%)</th>
<th>ACC(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Praat</td>
<td>10461</td>
<td>10893</td>
<td>87.96</td>
<td>72.23</td>
</tr>
<tr>
<td>EGG</td>
<td>10461</td>
<td>9856</td>
<td>90.46</td>
<td>87.43</td>
</tr>
<tr>
<td>CHFA</td>
<td>10461</td>
<td>10748</td>
<td>94.14</td>
<td>92.61</td>
</tr>
<tr>
<td>Hybrid (EGG+CHFA)</td>
<td>10461</td>
<td>10832</td>
<td>95.67</td>
<td>92.35</td>
</tr>
</tbody>
</table>
Table 2: Evaluation results of PMA for EGG based reference (female speaker).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Ref</th>
<th>Test</th>
<th>SR[%]</th>
<th>ACC[%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Praat</td>
<td>1720</td>
<td>1854</td>
<td>97.84</td>
<td>58.05</td>
</tr>
<tr>
<td>EGG</td>
<td>1720</td>
<td>1741</td>
<td>98.32</td>
<td>98.28</td>
</tr>
<tr>
<td>CHFA</td>
<td>1720</td>
<td>1811</td>
<td>82.60</td>
<td>69.95</td>
</tr>
<tr>
<td>Hybrid (EGG+CHFA)</td>
<td>1720</td>
<td>1853</td>
<td>98.77</td>
<td>97.85</td>
</tr>
</tbody>
</table>

Table 3: Evaluation results of PMA for speech based reference (male speaker).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Ref</th>
<th>Test</th>
<th>SR[%]</th>
<th>ACC[%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Praat</td>
<td>1127</td>
<td>1157</td>
<td>60.62</td>
<td>27.38</td>
</tr>
<tr>
<td>EGG</td>
<td>1127</td>
<td>1165</td>
<td>76.05</td>
<td>52.87</td>
</tr>
<tr>
<td>CHFA</td>
<td>1127</td>
<td>1124</td>
<td>75.76</td>
<td>57.66</td>
</tr>
<tr>
<td>Hybrid (EGG+CHFA)</td>
<td>1127</td>
<td>1255</td>
<td>77.56</td>
<td>53.28</td>
</tr>
</tbody>
</table>

Table 4: Evaluation results of PMA for the whole speech based reference (female + male speaker).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Ref</th>
<th>Test</th>
<th>SR[%]</th>
<th>ACC[%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Praat</td>
<td>11588</td>
<td>12050</td>
<td>82.82</td>
<td>69.31</td>
</tr>
<tr>
<td>EGG</td>
<td>11588</td>
<td>11021</td>
<td>88.76</td>
<td>83.36</td>
</tr>
<tr>
<td>CHFA</td>
<td>11588</td>
<td>11872</td>
<td>91.98</td>
<td>88.50</td>
</tr>
<tr>
<td>Hybrid (EGG+CHFA)</td>
<td>11588</td>
<td>12087</td>
<td>93.54</td>
<td>87.76</td>
</tr>
</tbody>
</table>

Figure 4: Comparison of the success rate between PMAs

5. Conclusions

The modified hybrid algorithm for pitch marking was presented. The algorithm combines advantages of both single, EGG based and speech signal based approaches. To evaluate the performance, a subset of the TC-STAR database with one female and male speaker including a manually segmented PM reference track was used. The described algorithm was compared to the established pitch mark algorithm used by Praat program. Evaluation criteria after [14] has been used for measurement the errors. Our experimental results indicate that the hybrid algorithm performs better than both single algorithms by providing about 1.5% improvement compared to the advanced speech signal based PMA. By increasing the performance of the underlying EGG signal and speech signal based algorithms, we expect further improvements regarding the hybrid method. In this study, the PM analysis of the female voice performed better than the one for the male voice.

The proposed algorithm was used for pitch detection (PDA) during an automatic annotation task (phrases) [15] within TC-STAR project.

6. References