Comparison between DyWT- and EGG-based Estimation of Glottal Closure Instant for Speech Signal

Jongwon Seok, Youngho Son, and Keunsung Bae

School of Electronic and Electrical Engineering, Kyungpook Natl. Univ., Taegu 702-701, KOREA

Abstract: The glottal closure instants obtained from the wavelet analysis of speech signal are investigated through the comparison with those obtained from the EGG signal. Epochs estimated using the cubic spline wavelet show zigzag patterns around the actual instants of glottal closure, i.e., EGG-based epochs. About 96% of DyWT-based epochs are included within the range of ±0.5 ms with respect to the EGG-based epochs.

ESTIMATION OF GLOTTAL CLOSURE INSTANT

Air pressure built up below the larynx by the efforts of muscles and lungs provides the energy for the speech signal. The vocal folds in the larynx are positioned so that air can flow through the glottis. The puffs of air produced by opening and closing the glottis with vocal folds generate a periodic excitation for the vocal tract and this operation results in voiced sounds. During the closing and opening operation of glottis, the main excitation is known to occur at the instant of glottal closure that is called the GCI (Glottal Closure Instant) or epoch. Determination of the epochs plays an important role in some applications of speech processing such as voice source analysis, pitch detection and speech synthesis.

The electroglottograph (EGG) signal represents the vibratory motion of the vocal folds. It is known that the declining portion of the EGG waveform corresponds to the glottal closing phase and the GCI coincides well with the negative peak point in the differentiated EGG signal \[1\]. Therefore, EGG-based estimation of epochs, i.e., the negative peak points in the differentiated EGG for corresponding pitch periods can be considered as the actual instants of glottal closure.

Mallat \[2\] has shown that, with a proper choice of wavelet function, the local maxima of wavelet transformed signal indicate the sharp variation in the signal. Since the glottal closure causes sharp discontinuities in the speech signal, dyadic wavelet transform (DyWT) can be useful for detecting abrupt change in the voiced sounds, i.e., epochs. So epoch detection using the wavelet analysis has recently attracted an increasing amount of attention \[3,4\]. Most of them, however, focused on the development of epoch detection algorithm and few researchers paid attention to the accuracy of estimated epochs compared with the actual instant of glottal closure. In this paper, we investigate the glottal closure instants obtained from the wavelet analysis of speech signal and compare them with those obtained from the EGG signal.

EXPERIMENTS AND DISCUSSION

Synchronized speech and EGG signals sampled at 10 kHz with 16 bits/sample, respectively, are used for analysis. Two male and two female speakers uttered five English sentences for data collection. The cubic spline wavelet given in [2] is used for wavelet analysis, and epochs are determined by visual inspection of the wavelet transformed speech signals at scale range from 3 to 5.

Using the EGG-based epochs as a reference, detection accuracy of DyWT-based epochs has been studied. DyWT-based epochs show zigzag patterns around the actual instants of glottal closure. It is thought to be caused by the source-tract coupling during the utterance. The distribution of DyWT-based epochs with respect to EGG-based epochs is plotted in Figure 1 for male and female speakers, respectively. Table 1 shows the average detection ratio of DyWT-based epochs for the given tolerance at each wavelet scale. It is shown that wavelet scale 3 and 4 give more accurate estimation of epochs than scale 5 and about 96% of epochs locates within the range of ±5 samples with respect to EGG-based epochs.

The pitch period of voiced sound can be obtained by computing the time interval between epochs. Figure 2 shows the difference of pitch period between DyWT-based method and EGG-based one. It can be seen that the variation of pitch period is far less than that of epochs.
TABLE 1. Epoch detection ratio with given tolerance at each wavelet scale

<table>
<thead>
<tr>
<th>Scale</th>
<th>±3 samples</th>
<th>±4 samples</th>
<th>±5 samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale 3</td>
<td>80.38 %</td>
<td>90.00 %</td>
<td>96.20 %</td>
</tr>
<tr>
<td>Scale 4</td>
<td>80.37 %</td>
<td>90.58 %</td>
<td>95.76 %</td>
</tr>
<tr>
<td>Scale 5</td>
<td>72.78 %</td>
<td>81.07 %</td>
<td>87.75 %</td>
</tr>
</tbody>
</table>

ACKNOWLEDGEMENTS

This work has been partially supported by the Korea Science and Engineering Foundation under Grant KOSEF-971-0917-103-2.

REFERENCE