Abstract: In this paper, an improved Multiscale Product-based Method is evaluated for the open quotient (OQ) estimation from the noisy speech signal. The method consists of making the multi-scale wavelet transforms coefficients product at three scales. Our proposed approach is based upon correlation functions computed on the negative and the positive parts of the speech Multiscale Product (MP). It operates without determining the glottal opening and closing instants. Over each frame, the pitch period is given by the first maximum of the MP negative part autocorrelation and the open phase is given by the first maximum of the inter-correlation between the negative and positive parts. OQ is the ratio of the open phase and the pitch period. Tested on the Keele University database, our new approach proves to be robust to noise degradation.

Keywords: open quotient, speech, multi-scale product, correlation, noise.

I. INTRODUCTION

According to the acoustic theory of the speech production, the acoustic source signal produced by the vibrating vocal folds is filtered by the vocal tract to produce the speech output signal [1]. For voiced speech, the glottal vibration is periodic, with the folds opening and closing repeatedly in a regular manner. Thus, one period of the voice source signal includes open phase and closed phase. During the closed phase, the vocal folds are in full contact and there is no air flow passing through the glottis. The open phase is itself divided into an opening phase during which the vocal cords begin to separate gradually and a closing phase during which the separated folds start to be in close. Therefore, during the open phase the air passes through the glottis and the vocal cords are totally or partially detached.

The instant of vocal folds full contact is called the glottal closure instant (GCI) and one of vocal folds complete separation is the glottal opening instant (GOI). GCI and GOI are events of great interest for the glottis excitation. The open quotient is another interesting parameter characterising the source signal. It is defined as the ratio between the open phase and the cycle period.

Inverse filtering (IF) is a common and useful technique for voice source analysis. The principle of the IF is to cancel the vocal tract effect from a recorded speech signal to acquire a glottal flow [2].

Direct measurement of the glottis parameters from the radiated speech signal is still a challenging problem in speech analysis and synthesis domains. Though, numerous parameterisation approaches have been suggested. Time-based methods consist of detecting significant events such as glottal opening and closing instants to compute the glottis parameters [3], [4]. Frequency-based methods use the properties of the flow magnitude spectrum such as the level difference of the harmonics [5], [6]. In [7], Hanson uses the difference between the magnitudes of the first two spectral harmonics (H1–H2) as an indication of the open quotient.

The electroglottographic recordings are used by many researchers to extract the glottal source features. Recently, Henrich et al. have proposed a correlation-based method called DECOM [8]. Her algorithm uses the correlation of the DEGG signal to estimate the fundamental frequency (F0) and the open quotient (OQ).

In this study, we focus on applying the Henrich correlation algorithm on the speech MP to estimate the open quotient from a noisy speech. The idea is born from the fact that the speech MP is strongly close to the EGG signal.

This paper is organised as follows. Section 2 reviews the principle of the multi-scale product analysis. Section 3 describes the Correlation Multiscale Product-based method for measuring the open quotient from the speech signal. In section 4, we evaluate the performance of our approach on clean and noisy speech data. Section 5 concludes this work.

II. MULTISCALE PRODUCTS FOR SPEECH ANALYSIS

Wavelet transform is a multiscale analysis widely used in image and signal processing. Due to the efficient time-frequency localisation and the multiresolution characteristics, the wavelet transforms are quite suitable for processing signals of transient and non-stationary nature. In [9], Mallat has shown that
multiscale edge detection is equivalent to find the local maximum of its wavelet representation. Glottal closure and opening instants are such events characterising the speech signal. The peak displaying the discontinuity in the wavelet transform is often damaged by noise when the scale is so fine or smoothed when the scale is large.

To improve edge detection using wavelet analysis, the multiscale product method is proposed. The latter consists of making the product of the wavelet transform coefficients of the acoustic signal over three scales. It enhances the peak amplitude of the modulus maxima line and eliminates spurious peaks due to the vocal tract effect.

The product of the wavelet transform of a function $f(n)$ at scales is:

$$p(n) = \prod_j W_{sj} f(n)$$  \hspace{2cm} (1)

Were $W_{sj} f(n)$ represents the wavelet transform of the function $f(n)$ at scale $sj$.

The product $p(n)$ shows peaks at signal edges, and has relatively small values elsewhere. An odd number of terms in $p(n)$ preserves the edge sign.

The MPM was first related to the edge detection problem in image processing [10]. Besides, the MPM is proposed by Bouzid and Ellouze to extract crucial information concerning the vocal source from both the speech and the electroglottographic signal (EGG) such as glottal opening and closure instants, the fundamental frequency, the open quotient and the voicing decision [11], [12].

III. MULTISCALE PRODUCT CORRELATION-BASED METHOD FOR OPEN QUOTIENT MEASUREMENT

As illustrated in Fig. 2, our proposed approach for the open quotient estimation from the speech signal operates following three stages. The first stage consists of computing the MP of a voiced speech signal and then dividing it into frames of a fixed length. The second stage consists of separating the speech MP into two parts: a negative part $MP_c$ which contains information concerning glottal closure peaks, and a positive part $MP_o$ which contains information about glottal opening peaks. The $MP_c$ signal is derived from the original signal by replacing any positive value by zero. In the same way, the $MP_o$ signal is derived from the original signal by replacing any negative value by zero.

The third stage concerns the calculation of the intercorrelation function between the positive and negative parts ($MP_o$ and $MP_c$) to estimate the open phase, and the autocorrelation function of the $MP_c$ to estimate the pitch period over each frame. The open phase and the pitch period are respectively given by the non null index matching with the first maximum of the intercorrelation and autocorrelation functions. The open quotient is then deduced by calculating the ratio between the open phase and the pitch period.

Fig. 1: Overview of the proposed method.
To compute the MP, we multiply the wavelet transforms of the speech signal at scales 2, 5/2 and 3 using the quadratic spline function.

To divide the MP signal into frames of a length \( N \), we multiply it by a sliding rectangular window \( w[N] \). The MP over a window of index \( i \) is given by the following equation:

\[
MP_{wi}[k] = MP[k - iN]w[k] \tag{2}
\]

Where \( k \) is within \([1, N]\) and \( i \) is the frame index.

The intercorrelation function between MP\(o\) and MP\(c\) over a frame \( i \) is calculated as follows:

\[
R_o(k) = \sum_{l=1}^{N} MP_{wi}^o(l)MP_{wi}^c(k + l) \tag{3}
\]

As the same way, the autocorrelation function of MP\(c\) over a frame \( i \) is calculated as follows:

\[
R_c(k) = \sum_{l=1}^{N} MP_{wi}^c(l)MP_{wi}^c(k + l) \tag{4}
\]

Fig. 2 : Speech MP and the autocorrelation function of its positive part.

Fig. 3 : Speech MP and the inter-correlation between its negative part and positive one.

The non null index matching with the first maximum of the MP\(c\) autocorrelation function shown in the second part of the Fig. 2 corresponds to the pitch period which is defined as the distance separating two consecutive GCIs.

Fig. 3 shows the speech MP followed by the intercorrelation calculated between its positive and negative parts. The non null index matching with the first maximum of the inter-correlation function corresponds to the time between an opening peak and the consecutive closing peak which is termed as the open phase.

IV. EVALUATION RESULTS

In this section, we evaluate the performance of our proposed method for OQ estimation using the Keele University database. This database includes the acoustic speech signals and laryngograph signals (single speaker recording). Five adult female speakers (fi) and five adult male speakers (mi) with \( i \in \{1, ..., 5\} \) are recorded in low ambient noise conditions using a sound-proof room. Each utterance consists of the same phonetically-balanced English text. In each case, the acoustic and laryngograph signals are time-synchronised and share the same sampling rate value of 20 kHz [13].

To evaluate the performance of our OQ estimator, we compute the standard deviation (\( \sigma \)) of the error measured as a difference between the OQ estimated from the speech and the EGG signals.

To study the noise effect on the accuracy of our open quotient estimator, we add noise to the original speech signal at various SNR levels. The noise is taken from the noisex-92 database [14]. Babble and vehicle noises are considered in this work.

Table 1 shows the performance of our approach for OQ estimation from the clean and noisy speech.

On clean speech approach estimates OQ with a standard deviation ranging from 0.03 for f2 to 0.08 for m5. It’s a considerable accuracy for estimating open quotient from the speech signal. In fact, works developed in this field usually use the EGG recordings.

In the presence of noise at SNR levels ranging from 5dB to -5dB, we can notice that the noise has insignificant effect on the accuracy of the proposed approach. The majority of speakers save the same standard deviation value when adding noise. For others, the deviation increases finely when the SNR level reaches -5 dB.

V. CONCLUSION

In this paper, we have proposed an improved Multiscale Product-based method for estimating the open quotient from clean and noisy speech signal. The proposed method exploits the correlation of the speech multiscale product which reminds the derivative of the EGG signal shape representing the global source activity.

The OQ estimation is obtained by calculating the ratio of the open phase over the pitch period. The open phase is referred as the index non null of the first
maximum localised on the inter-correlation function between the positive and the negative parts of the speech MP. As the same way, the pitch period is the non null index matching with the first maximum of the speech MP correlation function.

Standard deviation between OQ estimated from the speech signal and OQ measured from the EGG signal is measured to evaluate our method. The evaluation is done on the Keele University database in a noisy environment. Noise is extracted from the noisex-92 database. The proposed approach is proved to be accurate and robust.

REFERENCES

Table 1: Performance of Multiscale Product Correlation method for open quotient estimation using the Keele database on clean and noisy speech

<table>
<thead>
<tr>
<th>Noise Type</th>
<th>SNR</th>
<th>Female speakers</th>
<th>Male speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F1</td>
<td>F2</td>
</tr>
<tr>
<td>Clean speech</td>
<td>5 db</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>Babble</td>
<td>0 db</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>-5 db</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Vehicle</td>
<td>5 db</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>-5 db</td>
<td>0.07</td>
<td>0.06</td>
</tr>
</tbody>
</table>