The fundamental frequency of cough by autocorrelation analysis

A. Van Hirtum, D. Berckmans

Department of Agro-Engineering and Economics
K.U.Leuven, Kasteelpark Arenberg 30, 3001 Leuven, Belgium
daniel.berckmans@agr.kuleuven.ac.be

Abstract

The presented research evaluates the quantitative characterization of human cough sounds by estimating the fundamental frequency or pitch. The fundamental frequency was determined by autocorrelation analysis on both the rough time-signal and the linear predicted time-signal. Differences between ‘spontaneous’ and ‘voluntary’ cough sounds are put forward. The experimental cough database was registered in the free acoustical field on respectively 3 pathological and 9 healthy non-smoking subjects.

1. Introduction

As reflex-generated perturbation of the respiratory function ‘spontaneous’ cough is an important symptom in many respiratory diseases [1]. Auditive characterization of the cough sound resulted in several common labels as brassing, barking, whooping, etc. A more quantitative and objective cough-characterization is obtained by time-frequency analysis [2, 3]. In order to become a better understanding of the mechanism of cough sound creation it is attempted to interpret time- and spectral features in relation to the localization of cough sound production. In general the character of free field cough is said to reflect the state of the airway tissue and the behavior of the glottis [2]. The cough sound timbre as expressed in the auditive labelling is assumed to be determined by the glottis [2]. The cough sound timbre as expressed in the auditive labelling is assumed to be determined by the glottis [2]. The character of free field cough is said to reflect the state of the airway tissue and the behavior of the glottis [2]. The cough sound timbre as expressed in the auditive labelling is assumed to be determined by the glottis [2].

2. Data

Free field acoustic registration at 22050 Hz is performed with a standard multi-media microphone and sound-card. Since one of the objectives is to consider differences between ‘voluntary’ and ‘spontaneous’ cough sounds each individual cough sound is labelled in accordance. For the ‘spontaneous’ coughs this excludes any further diagnostics towards the character of the cough sound as e.g. described in [2, 3]. The ‘spontaneous’ cough class consists of 48 cough samples, the ‘voluntary’ out of 36 originating from respectively 3 pathological and 9 healthy non-smoking subjects, all aged between 20 and 30.

3. Methods

3.1. Linear predictive coding

In linear predictive coding (LPC) a given discrete time signal, $s(n)$, is approximated with a linear combination of the past $p$ samples, $\tilde{s}(n)$ [6]. This is expressed in Formula 1 where the coefficients $a_1, \ldots, a_p$ are assumed to be constant over the speech analysis frame.

$$s(n) \approx a_1 s(n-1) + \ldots + a_p s(n-p) = \tilde{s}(n)$$

Defining the prediction error $e(n)$ as $e(n) = s(n) - \tilde{s}(n)$ the predictor coefficients are found by minimizing the squared error with respect to each $a_k$. In terms of the autocorrelation function, Formula 2, the LPC equations can be expressed as in Formula 3. In accordance with [6] the order $p$ is set to 10 since frequency-information up to 10kHz is considered.

$$r_{n-k} = \sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m+i-k)$$

The presented research in this paper intends to retrieve $f_0$ for an experimental human cough database containing both ‘voluntary’ and ‘spontaneous’ cough sounds registered in the free acoustical field. At first resulting $f_0$’s will be assessed towards existence. At second retrieved $f_0$-values for ‘voluntary’ cough sounds will be compared to the $f_0$-values obtained for ‘spontaneous’ cough sounds. At third the usefulness of cough-characterization with $f_0$-envelope or pitch-contour will be considered. The fundamental frequency $f_0$ will be determined by autocorrelation analysis both on the rough time-signal as on the linear predicted pre-processed time-signal.
The pitch or fundamental frequency \( f_0 \) of a signal is determined by computing the autocorrelation of the pre-processed time-signal as is common for speech [7]. Pre-processing is aimed to spectrally flatten the signal so as to eliminate the effects of the vocal tract spectrum on the detailed shape of the resulting autocorrelation function. In addition the periodicity of the signal is enhanced. The three-step processing consists of successively windowing, filtering and clipping. Windowing tapers the signal to zero outside the analysis frame as defined by the window-size which is necessary for computing the short time autocorrelation function, already mentioned in Formula 2. The windowing is followed by low-pass filtering in order to eliminate the effects of the higher formant harmonics on the autocorrelation. A low-pass 9 order Butterworth filter with a cut-off frequency of 1000 Hz is applied. In addition to filtering non-linear signal-clipping is used to remove the effects of the first formant. Compressed center clipping re-scales the high-amplitude parts of the filtered signal without changing the signal-shape by reducing the amplitude with a fixed amount. The clipping threshold was set as a fixed percentage (60 percent) of the smaller of the maximum absolute signal level over the first and last one-thirds of the analysis frame, which is proven quite successful in speech.

**4. Results and discussion**

The fundamental frequency is estimated by autocorrelation analysis as described in Section 3.2 on both the rough time-signal and the LPC predicted time-signal. The first step in the pre-processing is performed by applying a hamming-window. The window-width or the analysis frame is chosen to agree with the signal-duration of the 'spontaneous' cough database as can be seen from the mean and median in Table 1. The windowing is followed by low-pass filtering in order to eliminate the effects of the higher formant harmonics on the autocorrelation. A low-pass 9 order Butterworth filter with a cut-off frequency of 1000 Hz is applied. In addition to filtering non-linear signal-clipping is used to remove the effects of the first formant. Compressed center clipping re-scales the high-amplitude parts of the filtered signal without changing the signal-shape by reducing the amplitude with a fixed amount. The clipping threshold was set as a fixed percentage (60 percent) of the smaller of the maximum absolute signal level over the first and last one-thirds of the analysis frame, which is proven quite successful in speech.

### Table 1: Mean and median duration (sec) of ‘spontaneous’ and ‘voluntary’ cough-sounds in the data-set.

<table>
<thead>
<tr>
<th>cough</th>
<th>number</th>
<th>mean</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td>spontaneous</td>
<td>48</td>
<td>1.0</td>
<td>0.9</td>
</tr>
<tr>
<td>voluntary</td>
<td>36</td>
<td>0.8</td>
<td>0.5</td>
</tr>
</tbody>
</table>

In addition to filtering non-linear signal-clipping is used to remove the effects of the first formant. Compressed center clipping re-scales the high-amplitude parts of the filtered signal without changing the signal-shape by reducing the amplitude with a fixed amount. The clipping threshold was set as a fixed percentage (60 percent) of the smaller of the maximum absolute signal level over the first and last one-thirds of the analysis frame, which is proven quite successful in speech.

Figure 1: Exemplar rough time-signal of ‘spontaneous’ (top) and ‘voluntary’ (bottom) cough-effort.

Both cough-classes is larger than the mean duration-value of e.g. respectively 600 and 410 ms for ‘spontaneous’ and ‘voluntary’ cough sound found in literature, but the observation of a prolonged duration for ‘spontaneous’ cough compared to ‘voluntary’ cough is supported [2, 8]. As a consequence the used analysis frame for the ‘spontaneous’ cough is generally larger than the frame used for ‘voluntary’ cough, which is important towards pitch detection. Since one window is applied to each of cough-sound the resulting pitch-estimate is an averaged pitch estimate for each cough-event. All references [2, 3, 4] and own experience with the searched database [9] mention both for ‘spontaneous’ as for ‘voluntary’ cough the presence of signal-energy up to 6 or 7 kHz. The second pre-processing step involving a low-pass filtering with a cut-off frequency of 1000 Hz seems therefore reasonable. Moreover the filter-threshold of 1000 Hz is well above the in literature mentioned pitch values ranging from 300 to 700 Hz for ‘voluntary’ cough [2]. Finally the non-linear compressed center clipping explained in Subsection 3.2 ends the signal-processing. So all output samples after clipping are proportional in amplitude to the amount by which they exceed the clipping threshold. After pre-processing the autocorrelation analysis is completed by computing the autocorrelation on both the 3-step pre-processed rough time-signal and the LPC predicted time-signal. The resulting correlation show peaks at the pitch period and the pitch-value is estimated by peak-detection. However if the normalized peak correlation value fell below a threshold (0.25), reliable selection of the pitch was not possible and no pitch was detected for the signal. As no pitch could be estimated the cough-signal is said to be unvoiced. If the cough-signal could be characterized with an averaged pitch- or \( f_0 \)-value the sound is said to be voiced. The number and percentage of voiced respectively ‘spontaneous’ and ‘voluntary’ cough-
sounds is given in Table 2 for pitch-estimation on both the rough time-signal and the LPC-predicted time-signal. From the table can be seen that autocorrelation analy-

<table>
<thead>
<tr>
<th>cough</th>
<th>rough</th>
<th>#</th>
<th>%</th>
<th>LPC</th>
<th>#</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>spontaneous</td>
<td>36</td>
<td>75</td>
<td>36</td>
<td>75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>voluntary</td>
<td>19</td>
<td>53</td>
<td>20</td>
<td>56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

sis on both the rough or the LPC predicted time-signal gives the same percentages for both cough-classes. So in spite of the statement that the cough sound timbre is determined by the fundamental frequency [2] respectively only ±75 % of the ‘spontaneous’ and even only ±55 % of the ‘voluntary’ cough sounds could be characterized by an average pitch-value. The lack of a pitch-estimation for ±45 % of the ‘voluntary’ cough sounds might be explained by considering one cough-effort as consisting out of a double cough sound as assumed in [2, 4] and illustrated in Figure 1 presenting the rough time-signal for both one ‘spontaneous’ (top) and one ‘voluntary’ (bottom) cough-effort defined as one expiratory flow peak in a cough sequence. The occasional absence of f₀-estimations might be reasoned because the vocal chords do not take part in the first cough sound creation as they are widened to its maximum extent in order to provide a minimum resistance [2, 4]. The percentage difference between respectively ‘spontaneous’ and ‘voluntary’ cough might be due to the presence of airway-narrows in case of ‘spontaneous’ cough having the same function as a reed instrument. The estimated fundamental frequencies for both ‘spontaneous’ as ‘voluntary’ cough sounds are 2D-depicted in Figure 2. f₀-values obtained by autocorrelation on the rough time-signal are plotted in ascending order on the x-axis, the y-axis shows the corresponding returns for autocorrelation analysis on the LPC predicted signal. f₀-estimations on both the rough time and LPC-predicted signal in general results in the same pitch-values which is illustrated in the figure by a straight line with ±45 % slope containing the majority of the 2D data-points. Without considering outliers estimated f₀-values for the total cough database are ranging from 90 Hz up to 600 Hz. So both the minimum as the maximum of the pitch-interval is lowered compared towards the pitch-interval from 300 up to 700 Hz mentioned in literature [2] for voluntary cough. Figure 3 shows the average pitch-values for ‘spontaneous’ cough sounds in the upper row and for ‘voluntary’ cough in the bottom row. The left column is obtained from the LPC-predicted signal, the right column from the rough time signal. Although the number of assessed cough-sounds is rather small majority of the pitch-estimates for ‘spontaneous’ cough sounds seems to be either below ±200 Hz or in the interval from ±350 up to ±450 Hz. Majority of pitch-values of ‘voluntary’ cough sounds seems to be situated from ±200 up to ±500 Hz or above ±450 Hz. Illustrating an average pitch-decrement from ‘spontaneous’ cough sounds compared to ‘voluntary’ cough sounds. The distinction of two pitch-zones which is certainly the case for ‘spontaneous’ cough might on one hand suggest that the highest zone corresponds to a higher harmonic instead of the pitch, indicating the importance of the higher harmonics in the cough sound or on other hand correspond to the mentioned distinction of the first and second cough sound in one cough-effort. Until now the average pitch for each cough sound is considered. However in accordance with [2, 3, 4] Figure 1 illustrates the possible difference in time-envelope between ‘spontaneous’ versus ‘voluntary’ cough sounds as the ‘spontaneous’ time-signal in general is more extended and possibly folded compared to the ‘voluntary’ time-signal. Moreover in general the ideal window-length for autocorrelation is considered to contain 2 or 3 pitch periods [7]. The average pitch-period however is much more as 2 or 3 times shorter as the window-length is determined by the signal-duration. So both to achieve a more accurate pitch-estimation starting from a more appropriate window-length as to consider pitch-changes during the cough-signal representing the variation in the time-envelope it is attempted to value the usefulness of pitch-contour characterization to gain additional information of the cough-sound. As exemplar the autocorrelation function (right column) computed for different successive parts of 1500 up to 2000 sample-points (left column) of the first ‘spontaneous’ cough of Figure 1 is showed in Figure 4. The plot illustrates the difference

Table 2: Number (\#) and percentage of respectively ‘spontaneous’ and ‘voluntary’ coughs recognized as voiced by autocorrelation analysis on respectively the rough time-signal and the LPC-reconstructed signal.

Figure 2: Overview f₀ estimated values by autocorrelation on the rough time (x-axis) and on the LPC predicted time-signal (y-axis).
in achieved \( f_0 \) between the beginning of the signal (first row), inside the first extended high amplitude part (second row) or at the signal end (last row). Table 3 quantifies

<table>
<thead>
<tr>
<th>number</th>
<th>averaged</th>
<th>before fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>110</td>
<td>216</td>
</tr>
<tr>
<td>4</td>
<td>418</td>
<td>412</td>
</tr>
<tr>
<td>1</td>
<td>88</td>
<td>103</td>
</tr>
</tbody>
</table>

the pitch-values for 8 two-folded ‘spontaneous’ cough-signals obtained respectively on the total time-signal and on the first extended high amplitude part before the fold. Both from Figure 4 and from table 3 further research towards pitch-contour characterization seems appropriate.

5. Conclusions

In this paper the fundamental frequency of human ‘spontaneous’ and ‘voluntary’ cough is assessed by autocorrelation analysis on both the rough time-signal and the linear predicted time-signal. Respectively 75 % of the ‘spontaneous’ versus 55 % of the ‘voluntary’ cough is characterized by periodicity expressed by an average fundamental frequency. Generally the achieved pitch-values for ‘spontaneous’ cough were situated either below 200 Hz or in the range from 350 up to 450 Hz. The pitch-values for ‘voluntary’ cough were mostly obtained in the intermediate interval from 200 Hz up to 350 Hz or above 450 Hz. Pitch is shown to vary during a single ‘spontaneous’ cough so further research towards pitch-contour characterization is suitable as well as towards cough-sound production.

6. References