DISTINCTION BETWEEN VOLUNTARY COUGH SOUND AND SPEECH IN VOLUNTEERS BY SPECTRAL AND COMPLEXITY ANALYSIS

1Department of Pathological Physiology and 2Institute of Physiology, Jessenius Faculty of Medicine, Comenius University, Martin, Slovakia

Objective monitoring of cough sound for extended period is an important step toward a better understanding of this symptom. Because ambulatory cough monitoring systems are not commercially available, we prepared own monitoring system, which is able to distinguish between voluntary cough sound and speech in healthy volunteers. 20-min sound records were obtained using portable digital voice recorder. Characteristics of the sound events have been calculated in time and frequency domains and by a nonlinear analysis. Based on selected parameters, classification tree was constructed for the classification of cough and non-cough sound events. We validated the usefulness of our algorithm developed against manual counts of cough obtained by a trained observer. The median sensitivity value was 100% (the interquartile range was 98-100) and the median specificity was 95% (the interquartile range was 90-97). In conclusion, we developed an algorithm to distinguish between voluntary cough sound and speech with a high degree of accuracy.

Key words: classification tree, cough sound, sample entropy, spectral analysis, speech

INTRODUCTION

Cough is the commonest symptom forcing the patients to seek a medical advice (1, 2). To date, the methods used to assess the cough have been primarily subjective and only broadly reflect the impact of cough and/or cough therapies on the quality of life (3). Because the cough is episodic in its nature, data collection over many hours is required for its objective monitoring requiring subsequent real-time aural analysis, which is equally time-consuming. Ambulatory cough monitoring systems have been proposed recently based either
on sound recordings alone (4) or on simultaneous sound and electromyography recordings (5-6), but their use has remained restricted to the research setting mainly due to the need of a trained operator to manually identify cough events from the recordings, which is an arduous task.

In order to make cough monitoring applicable to clinical practice, it is necessary to develop accurate automatic cough monitoring system for recording, detection and counting of coughs. With the availability of digital recording devices and the advances in digital storage media, battery powered mp3 player/recorders can be used to make high quality ambulatory sound recordings. Data can be transferred to personal computer and the recordings can be used to develop algorithms for cough sounds identification (7). Several systems for automatic cough recognition and monitoring based on sound recording have been described more recently (7-10). They are based on automatic cough detection algorithms that operate relatively reliably in the ambulatory setting. Different methods of analysis were applied to resolve the problem of recognition of cough sounds present in a given recording, while rejecting other sounds with similar characteristics.

Because these mathematical algorithms are not commercially available, we set out to prepare own mathematical algorithm for the analysis of cough frequency, based on spectral analysis and nonlinear method of sample entropy quantifying irregularity of a sound signal. A first step toward automated recognition of cough within a continuous sound recording is to distinguish blindly between cough sound and speech. We have some experience with the cough sound analysis (11, 12). The aim of the present study was to develop an algorithm based on the classification tree for the automatic recognition of cough sound and speech, as the first step in the development of a cough monitoring system.

MATERIAL AND METHODS

The study was approved by a local Ethics Committee and was performed in accordance with the Helsinki Declaration of 1975 for Human Research. We made a comparison between speech and voluntary cough sounds recorded from healthy subjects. Firstly, the signal was screened for periods of sound exceeding noise level within the recordings. These sound events were extracted, and the periods of silence were omitted from further analysis. The extracted sound events were stored into separate files and underwent digital signal processing to calculate the sound characteristics. The sound events were then classified into cough and non-cough events using classification tree (13).

Subjects

The study group included 20 healthy subjects (15 female - median age 34.3 yr, range 18 - 56 yr; 5 male - median age 47.2 yr, range 26 - 66 yr). All subjects were without any respiratory disease according to personal history and basic examination. Two subjects were excluded from our study because they were not able to perform voluntary cough appropriately.
**Recording system**

The recording system consisted of a portable digital voice recorder (Sony ICD-MX20, Sony Corporation, China) with the sampling frequency of 8 kHz and a miniature omnidirectional condenser microphone (ATR35s, Audio-Technica U.S., Philippines) with a flat frequency response between 50-18 000 Hz. The microphone was attached to the subject's chest and was covered by plastic foam membrane to suppress sounds coming from the outer environment. The audio signal from the microphone was initially recorded to the memory card of the digital recorder as a MSV file (memory stick voice file). After recording, we transferred this file into the PC and converted it to 11 kHz 16-bit mono digital wave file (WAV format) using Digital Voice Editor 2.31 software (Sony Corporation).

**Protocol**

All subjects performed continuous reading of a text from a book with voluntary coughs (46 cough events) performed at the instants indicated in the text. The recording lasted for about 20 min. Before reading the text, the subjects coughed voluntarily three times to obtain their individual cough sound pattern.

**Determination of sound events**

The first step in the sound analysis was the isolation of sound events from the raw recording. We used the moving window without an overlap over the whole audio signal to calculate the standard deviation (SD) of the signal for each position of a window. The length of the moving window was 200 points corresponding to duration of 18.2 ms. SD for each window position was compared with an empirically determined threshold value exceeding the background noise level. Portions of the signals containing no sound events showed only small SD related to the inherent noise present in the signal. Portions of the signal that were below the threshold value were excluded from further analysis. The sound event exceeding a given threshold was then subjected to a more detailed detection of its beginning and end (7). For this purpose, we found the time instants when the corresponding SD of the signal exceeded 40% of the previously used threshold at the beginning and end of the sound event. Since the cough sound can be composed of two relatively isolated sounds (double cough sounds), we searched for the sound immediately following a given sound event. If this sound was present both sounds were connected into one sound event. Detected sound events were stored in separate WAV files.

**Sound events analysis**

A second step was to calculate the characteristics of sound events. We quantified the duration of each sound file (parameter length).

Spectral analysis. Because of apparent non-stationarity of the sound signal, each sound event was analyzed using a short-time Fourier transform (STFT). We used the Hanning window (512 samples) to avoid spectral leakage and window was shifted with a step of 5 samples. The length of the used window corresponds to 45 ms of time. For each window position, power spectral density (PSD) (smoothed by moving averaging) was computed. Total power (TP) corresponding to the area under the PSD curve was computed as a measure of the sound intensity for a given shift of the window. The maximum TP value throughout the sound event (global maximum) was denoted as $TP_{max}$ - it is a measure of a maximum sound intensity. $TP_{mean}$ was a measure of the mean sound intensity and was computed as an arithmetic mean of the TP values for the whole sound event.

Next, we found all the local maxima and minima in the time course of TP in a sound event. We computed the measure ratio as a ratio of the sum of TPs of all local maxima divided by the sum of TPs of all local minima in given sound events.
From the spectrum corresponding to the global maximum of the sound, we also computed the skewness ($\text{skewness}_{\text{global}}$) and kurtosis ($\text{kurtosis}_{\text{global}}$) of the PSD value distribution (in the frequency band 0-1000 Hz) to distinguish between harmonic sounds (high peaks in the spectrum) and noise-like sounds (flat spectra). The time of the occurrence of the global maximum was marked as the parameter $\text{time}_{\text{global}}$.

In the next step, we found the first local maximum of TP and its value was divided by the time when it occurred (from the start of the sound event) - the parameter $\text{slope}$. This measure quantifies the sound intensity increase at the start of a sound event. The time of the occurrence of the first local maximum was marked as the parameter $\text{time}_{\text{local}}$. Analogously to the global maximum spectrum, skewness ($\text{skewness}_{\text{local}}$) and kurtosis ($\text{kurtosis}_{\text{local}}$) and the time of its occurrence ($\text{time}_{\text{local}}$) were also computed for the spectrum corresponding to the first local maximum.

**Nonlinear analysis.** For 512 samples corresponding to the local and global maxima, we computed sample entropy (SampEn) values. SampEn is a measure of irregularity and unpredictability of the signal. Therefore, it is higher for noisy signals compared to periodic oscillations. SampEn ($m$, $r$, $N$) is a negative natural logarithm of the conditional probability that two sequences similar for $m$ points remain similar at the next point. Algorithm for SampEn computation was published elsewhere (13).

We presumed that cough sound would have a higher degree of irregularity compared with the speech. The SampEn was calculated for two values of the input parameter $r$ (tolerance; $r = 0.1$ and $r = 0.2$ times SD of the window). The length of compared sequence ($m = 2$) and the length of analyzed window ($N = 512$ samples) was fixed. The SampEn for a local maximum was denoted as $\text{SampEn}_{\text{local}}(0.1)$ and $\text{SampEn}_{\text{local}}(0.2)$. SampEn values corresponding to the global maximum were denoted as $\text{SampEn}_{\text{global}}(0.1)$ and $\text{SampEn}_{\text{global}}(0.2)$.

**Classification tree**

The sound events were then classified into "cough" and "non-cough" sounds using a classification tree. The input parameters for tree construction were all the assessed variables ($\text{length}$, $\text{TPmax}$, $\text{TPmean}$, $\text{ratio}$, $\text{slope}$, $\text{skewness}_{\text{local}}$, $\text{kurtosis}_{\text{local}}$, $\text{skewness}_{\text{global}}$, $\text{kurtosis}_{\text{global}}$, $\text{time}_{\text{local}}$, $\text{time}_{\text{global}}$, $\text{SampEn}_{\text{local}}(0.1)$, $\text{SampEn}_{\text{local}}(0.2)$, $\text{SampEn}_{\text{global}}(0.1)$, $\text{SampEn}_{\text{global}}(0.2)$) and the output parameter of the tree was the presence of the cough or non-cough events.

**Statistics**

Differences in the sound parameters between cough and non-cough sounds were evaluated using a nonparametric Mann-Whitney U-test. For the classification of sound events, the classification tree was constructed. Values are presented as medians and interquartile ranges due to non-gaussian distribution of the variables. Statistical analysis was performed using statistical package Systat 10, SPSS Inc.

**RESULTS**

All variables were significantly different when comparing cough sound and speech ($P<0.001$ for all). The variables significantly higher in cough sounds (compared to speech) include: $\text{length}$, $\text{TPmax}$, $\text{TPmean}$, $\text{slope}$, $\text{SampEn}_{\text{local}}(0.1)$, $\text{SampEn}_{\text{local}}(0.2)$, $\text{SampEn}_{\text{global}}(0.1)$, and $\text{SampEn}_{\text{global}}(0.2)$, while the following variables were significantly lower: $\text{time}_{\text{local}}$, $\text{skewness}_{\text{local}}$, $\text{kurtosis}_{\text{local}}$, $\text{ratio}$, $\text{time}_{\text{global}}$, $\text{skewness}_{\text{global}}$, and $\text{kurtosis}_{\text{global}}$.

We made 18 recordings from which we obtained 6590 sound files, consisting of cough and non-cough sounds. The sound files included 892 cough sounds and
5698 non-cough sounds. The algorithm of the tree construction selected 6 sound characteristics (length, TPmean, slope, ratio, SampEn$_{local} (0.1)$, and SampEn$_{local} (0.2)$) as the variables that were most useful for the classification (Fig. 1). The selected variables provide relatively independent information contained in the sound files, which is useful for the classification of sound events. The results of classification tree performance for individual subjects are summarized in Table 1. We compared the effectiveness of our designed algorithm with the manually classified, as cough and non-cough events, sound files. The system performance was evaluated calculating the sensitivity, specificity, and a number of true positives, true negatives, false positives, and false negatives. The median sensitivity was 100% (interquartile range was 98-100) and the median specificity was 95% (interquartile range was 90-97).

The performance of the algorithm is influenced by the energy level of the cough signal. Lower values of sensitivity and specificity were obtained in subjects who had lower intensity of cough and speech. In the individual recordings, a total number of coughs differed, because in some subjects voluntary cough evoked spontaneous cough and several volunteers forgot to cough at indicated instants.

Table 1. Results of sound classification in 18 subjects and comparison between performance of the developed algorithm vs. manual counting of coughs are shown. The recordings were classified by trained listeners and cough and non-cough sounds were marked. This classification was regarded as a gold standard for the evaluation of the classification tree effectiveness (5). Algorithm performance was evaluated by calculating the sensitivity, specificity, and a number of true positives, true negatives, false positives, and false negatives.

<table>
<thead>
<tr>
<th>Subject number</th>
<th>Manually marked</th>
<th>Automatic classification</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>True positive</th>
<th>True negative</th>
<th>False positive</th>
<th>False negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cough</td>
<td>Non-cough</td>
<td>Cough</td>
<td>Non-cough</td>
<td>%</td>
<td>%</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>1</td>
<td>48</td>
<td>460</td>
<td>73</td>
<td>435</td>
<td>100</td>
<td>95</td>
<td>48</td>
<td>435</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>70</td>
<td>55</td>
<td>65</td>
<td>100</td>
<td>93</td>
<td>50</td>
<td>65</td>
</tr>
<tr>
<td>3</td>
<td>51</td>
<td>151</td>
<td>57</td>
<td>145</td>
<td>100</td>
<td>96</td>
<td>51</td>
<td>145</td>
</tr>
<tr>
<td>4</td>
<td>49</td>
<td>253</td>
<td>75</td>
<td>128</td>
<td>98</td>
<td>90</td>
<td>48</td>
<td>227</td>
</tr>
<tr>
<td>5</td>
<td>55</td>
<td>491</td>
<td>86</td>
<td>460</td>
<td>100</td>
<td>94</td>
<td>55</td>
<td>460</td>
</tr>
<tr>
<td>6</td>
<td>52</td>
<td>23</td>
<td>55</td>
<td>20</td>
<td>100</td>
<td>87</td>
<td>52</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>52</td>
<td>294</td>
<td>54</td>
<td>292</td>
<td>98</td>
<td>99</td>
<td>51</td>
<td>291</td>
</tr>
<tr>
<td>8</td>
<td>52</td>
<td>148</td>
<td>55</td>
<td>145</td>
<td>98</td>
<td>97</td>
<td>51</td>
<td>144</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
<td>307</td>
<td>83</td>
<td>274</td>
<td>100</td>
<td>89</td>
<td>50</td>
<td>274</td>
</tr>
<tr>
<td>10</td>
<td>51</td>
<td>539</td>
<td>64</td>
<td>526</td>
<td>100</td>
<td>98</td>
<td>51</td>
<td>526</td>
</tr>
<tr>
<td>11</td>
<td>51</td>
<td>151</td>
<td>81</td>
<td>121</td>
<td>100</td>
<td>80</td>
<td>51</td>
<td>121</td>
</tr>
<tr>
<td>12</td>
<td>47</td>
<td>75</td>
<td>51</td>
<td>71</td>
<td>74</td>
<td>79</td>
<td>35</td>
<td>59</td>
</tr>
<tr>
<td>13</td>
<td>50</td>
<td>364</td>
<td>47</td>
<td>367</td>
<td>87</td>
<td>99</td>
<td>44</td>
<td>361</td>
</tr>
<tr>
<td>14</td>
<td>51</td>
<td>176</td>
<td>55</td>
<td>172</td>
<td>98</td>
<td>97</td>
<td>50</td>
<td>171</td>
</tr>
<tr>
<td>15</td>
<td>43</td>
<td>654</td>
<td>72</td>
<td>625</td>
<td>95</td>
<td>95</td>
<td>41</td>
<td>623</td>
</tr>
<tr>
<td>16</td>
<td>52</td>
<td>846</td>
<td>76</td>
<td>822</td>
<td>100</td>
<td>97</td>
<td>52</td>
<td>822</td>
</tr>
<tr>
<td>17</td>
<td>48</td>
<td>443</td>
<td>71</td>
<td>420</td>
<td>100</td>
<td>95</td>
<td>48</td>
<td>420</td>
</tr>
<tr>
<td>18</td>
<td>44</td>
<td>253</td>
<td>44</td>
<td>253</td>
<td>80</td>
<td>96</td>
<td>35</td>
<td>244</td>
</tr>
</tbody>
</table>
Fig. 1. Classification tree used for the distinction of cough sounds from speech. The algorithm chose the six most useful variables for correct classification of sound events. The algorithm performs a stepwise splitting. The top panel contains the entire sample, i.e., all sounds. For illustration, cough sounds, classified manually, form the column on the left-hand side of each panel, while non-cough sounds (speech) are displayed on the right-hand side. Each downward split panel contains a subset of the sample in the panel directly above it. Furthermore, each panel contains the sum of the samples in the corresponding panels below it. Each panel can be thought of as a cluster of objects, or cases, which are split down in the tree. The tree is binary, because each panel is split only into two subsamples. The variables automatically selected for the construction of a classification tree included:

- **length**: the entire length of the analyzed sound event;
- **ratio**: the ratio of the sum of total powers of all local maxima divided by the sum of total powers of all local minima;
- **slope**: the value of the first local maximum divided by the time of its occurrence;
- **TPmean**: the arithmetic mean of the total power for the whole sound event;
- **SampEn_{local} (0.1)**: sample entropy for 512 samples corresponding to the first local maximum for $r=0.1$;
- **SampEn_{local} (0.2)**: sample entropy for 512 samples corresponding to the first local maximum for $r=0.2$
DISCUSSION

In the present pilot study we presented an algorithm that is able to distinguish voluntary cough sound from speech. We consider that speech is the most common sound present in the 24-hour recordings obtained from subjects who perform their daily activities. Therefore, we assume that the distinction of the speech from cough sounds is the first step toward developing an ambulatory cough monitoring system.

The median sensitivity of our algorithm was 100% and specificity was 95%. The algorithm is based on time domain, spectral, and nonlinear analysis of the sound files. The sounds were classified as cough and non-cough using a classification tree. Generally, a high degree of accuracy with sensitivity over 95% was reached. Relatively small sensitivity in 3 patients (Table 1) can be attributed to a lower effort during voluntary cough in these subjects.

Recently described systems allowing the distinction between cough and non-cough sounds calculate time-varying spectral features of analyzed sounds and use hidden Markov models (10) or probabilistic neural network (9). The values of sensitivity and specificity in those studies are lower compared with our present results, but the comparison is hindered by differences in the measurement protocols. In the present study, we compared cough sounds vs. speech, but other authors compared cough sounds vs. sounds occurring during normal daily activities. Our protocol was standardized - the subjects performed continuous reading of the text from a book and they coughed voluntarily at the indicated instants.

One of the limitations of this study was a relatively small number of patients. Our protocol was aimed at distinguishing cough sounds from speech, but other sounds occur during daily activities, too. In the future, we want to investigate the performance of our algorithm for distinction between cough sounds and other sounds, including sneezing and sounds from outer environment. In addition, we compared voluntary cough sounds vs. speech. The characteristics of spontaneous cough sounds could be different compared with voluntary cough. Therefore, it is necessary to validate the algorithm in spontaneously coughing subjects.

We conclude that the algorithm developed for the distinction of speech and cough sounds reached a high degree of accuracy, indicating its potential usefulness in clinical practice. In future studies, we plan to improve the developed algorithm to enable the analysis of 24-hour monitoring of cough frequency in subjects performing normal daily activities.

Acknowledgements: Our thanks are due to T. Zatko and M. Vrabec for their outstanding technical assistance. This study was supported by European Social Fund Project SOP LZ 2005/NP1-027.

Conflicts of interest: The authors had no conflicts of interest to declare in relation to this article.
REFERENCES


Received: May 16, 2008
Accepted: August 21, 2008

Author’s address: Jozef Martinek. Department of Pathological Physiology, Comenius University, Jessenius Faculty of Medicine, Sklabinska 26 St., 037 53 Martin, Slovakia; phone: +421 43 4238213, fax: +421 43 4134807; e-mail: martinek@jfmed.uniba.sk