Linköping University Postprint

Heart sound cancellation from lung sound recordings using recurrence time statistics and nonlinear prediction

Ahlstrom, C., Liljefeldt, O., Hult, P. and Ask, P.

N.B.: When citing this work, cite the original article.

Heart Sound Cancellation From Lung Sound Recordings Using Recurrence Time Statistics and Nonlinear Prediction

Christer Ahlstrom, *Student Member, IEEE*, Olle Liljefeldt, Peter Hult, and Per Ask, *Senior Member, IEEE*

Abstract—Heart sounds (HS) obscure the interpretation of lung sounds (LS). This letter presents a new method to detect and remove this undesired disturbance. The HS detection algorithm is based on a recurrence time statistic that is sensitive to changes in a reconstructed state space. Signal segments that are found to contain HS are removed, and the arising missing parts are replaced with predicted LS using a nonlinear prediction scheme. The prediction operates in the reconstructed state space and uses an iterated integrated nearest trajectory algorithm. The HS detection algorithm detects HS with an error rate of 4% false positives and 8% false negatives. The spectral difference between the reconstructed LS signal and an LS signal with removed HS was 0.34 ± 0.25, 0.50 ± 0.33, 0.46 ± 0.35, and 0.94 ± 0.64 dB/Hz in the frequency bands 20–40, 40–70, 70–150, and 150–300 Hz, respectively. The cross-correlation index was found to be 99.7%, indicating excellent similarity between actual LS and predicted LS. Listening tests performed by a skilled physician showed high-quality auditory results.

Index Terms—Bioacoustics, heart sound (HS), lung sound (LS), nonlinear prediction, recurrence time statistics.

I. INTRODUCTION

AUSCULTATION of lung sounds (LS) is often the first resource for detection and discrimination of respiratory diseases, such as chronic obstructive pulmonary disease (COPD), pneumonia, and bronchiectasis [1]. However, diagnosis based on lung sounds is a complex task, and it is desirable to make the signal as audible as possible. Recorded LS signals contain noise from several sources, such as heart sounds (HS), friction rubs, and the surrounding environment. The latter sounds can be reduced with adequate and firm microphone placement and by using sound proof rooms, but HS noise is unavoidable. HS and LS have overlapping frequency spectra, and even though highpass filtering is often employed to reduce HS, this results in loss of important signal information [2]. Previous approaches to HS cancellation from recorded LS include wavelet-based methods [2], adaptive filtering techniques [3], and fourth-order statistics [4], all resulting in reduced but still audible HS. Recent studies indicate that cutting out segments containing HS followed by interpolation of the missing data yields promising results [5], [6].

Fractals and chaos theory have been used extensively to describe physiological dynamics, and the respiratory system is no exception. The three-dimensional branching patterns of the airways form a fractal structure, where each branch repeats itself over multiple length scales [7]. Power laws are closely related to fractals, and it can be shown that dynamic processes propagating from fractal structures also exhibit fluctuations in time that follow power law distributions [8]. This indicates that LS are indeed fractal. We propose a new method for HS cancellation, elaborating on the linear approaches presented in [5] and [6] with nonlinear tools. Most of the shortcomings with the short time Fourier transform (STFT) method in [5] was taken care of with the linear prediction method in [6] (the time resolution of the HS detection algorithm was improved and linear prediction was introduced instead of two-dimensional interpolation). However, the choice of linear model in [6] is dependent on a flow signal, while our method only needs the LS signal. Further, our method allows nonlinear behavior in the LS signal.

II. SUBJECTS AND DATA COLLECTION

This study was approved by the ethical committee at Linköping University Hospital and was performed on six healthy male subjects aged 28.5 ± 3.7 years. LS were acquired with a contact accelerometer (Siemens EMT25C), which is a sensor that has been used extensively in the LS research community. The recording site was the second intercostal space along the left sternal border, and the sensor was fixed with an adhesive elastic tape. The acquisition protocol consisted of three phases: 30 s of tidal breathing (period 1), about 60 s of breathing with continuously increasing breath volumes up to vital capacity (period 2), and 10 s of breath hold (period 3). Respiration rate was not controlled. A standard three-lead ECG was recorded as a reference for HS detection. Both signals were digitized at 6 kHz with 12 bps (National Instruments, DAQCard-700), after passing an anti-aliasing filter with a cutoff frequency of 2 kHz. The LS signals were normalized to unity. Acquisition and processing of data were conducted in Labview (National Instruments) and Matlab (The MathWorks), respectively.

III. METHODOLOGY

A recurrence time statistic sensitive to changes in the reconstructed state space of a signal was used for detecting HS segments. The detections were removed, and the resulting gaps

were filled with predicted LS using forward and backward prediction, exploiting the evolution of nearest neighbors in the reconstructed state space.

A. State-Space Reconstruction

The dynamics of a time discrete system is determined by its possible states in a multivariate vector space (called state space or phase space). The transitions between the states are described by vectors, and these vectors form a trajectory describing the time evolution of the system. An observed signal S is a projection from this multivariate state space onto a one-dimensional time series. S can be considered as a set of n scalar measurements

$$S = \{s_1, s_2, \ldots, s_n\} \tag{1}$$

from which a sequence of N d-dimensional vectors a_k can be constructed using Takens' delay embedding theorem

$$a_k = \{s_k, s_{k+\tau}, s_{k+2\tau}, \ldots, s_{k+(d-1)\tau}\}, \qquad k = 1, 2, \ldots, N \tag{2}$$

where τ is a delay parameter, and d is the embedding dimension [9]. The purpose of the embedding is to unfold the projection back into a reconstructed state space that is dynamically and topologically equivalent to the state space that generated the process S [10]. Since the dynamics of the reconstructed state space contains the same topological information as the original state space, characterization and prediction based on the reconstructed state space is as valid as if it were made in the true state space.

Takens' theorem assumes that S is infinitely long and noise free. These conditions are seldom met, and the selection of τ and d affects how accurately the embedding reconstructs the systems state space. When dealing with a finite time series, the choice of embedding dimension is not too crucial if d is sufficiently large [11]. A more important consideration is the window length needed to reconstruct each vector a_k, $w = \tau \cdot d$. Typically, univariate time series exhibit some sort of periodicity, and w should be chosen to span several of these oscillations. In this letter, τ and d were chosen based on the standard techniques of average mutual information and false nearest neighbors [9]. An average value of d was used for the whole study, while τ was calculated adaptively by automatic detection of the first minimum of the average mutual information.

B. HS Detection

HS have a transient waveform that is superpositioned upon LS. As HS and LS originate from different sources, they have different attractors (see Fig. 1). These changes in signal dynamics can be detected with a recurrence time statistic Ψ, commonly known as the Poincaré recurrence time or the recurrence time of the first kind [12]. An arbitrary state a_{ref} is chosen on the trajectory whereupon all recurrences within a hypersphere of radius r are selected, i.e., $B_r(a_{\text{ref}}) = \{a : \|a - a_{\text{ref}}\| \leq r\}$. Ψ is then defined as the total amount of states in the set B_r. Ψ is related to the information dimension via a power law, which motivates its ability to detect weak signal transitions based on the signals amplitude, period, dimension, and complexity [13].
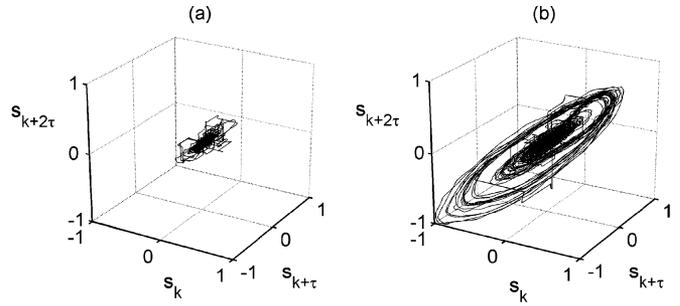

Fig. 1. State-space trajectories ($d = 3, \tau = 12$) of LS with HS (a) removed and (b) present. The transition between the two attractors is reflected in the recurrence time statistic, hence indicating when an HS is present.

A sliding window is used to partition the recorded LS signal into overlapping segments (and hence obtaining time resolution), and Ψ is calculated for each segment. The size was set to 1200 samples (200 ms), and the overlap was excessively set to 1190 samples to get a fine time resolution. An increment of one sample could have been used to increase time resolution further, but ten samples was a good compromise to save CPU time and to reduce the size of the matrices. The r-value is a very important parameter in the detection algorithm. If r is chosen too low, the hypersphere would be low on data, and if r is chosen too high, the hypersphere will contain misleading information from erroneous parts of the reconstructed state space. An adaptive threshold determined by the slow varying envelope of the LS signal (extracted as the lowpass filtered output of a Hilbert filter) was used to control r, after being translated and scaled to match the r-values. Different translation parameters were investigated in the range $r = [0.08, 0.2]$, while the scaling factor was set to 0.2. When the time-varying r-values had been determined, $\Psi(r)$ was normalized to unit, and a threshold was set to 0.6 to pick out the HS (see Fig. 2). Extracted HS were compared to the ECG to determine detectability. R-peaks were used as markers for the first HS and the T-waves for the second HS, respectively.

C. Lung Sound Prediction

After removing all HS, the data in the resulting missing segments were replaced using nonlinear prediction. Similar trajectories in state space share the same waveform characteristics in time domain, and a way of predicting missing data is thus to mimic the evolution of neighboring trajectories. The prediction scheme used was based on nearest trajectories with iterated integrated prediction [11] (see Fig. 3). Six LS segments surrounding the HS segment were used to reconstruct the state space, and five nearest neighbors were used in the prediction. The components of S closest in time to the prediction were accentuated by applying a biquadratic weight to the neighbors [11]. Bringing the predicted data back into the time domain is simple since s_k is the first coordinate of the embedding vector a_k.

Since the prediction error grows exponentially with prediction length [9], both forward and backward prediction were used (hence dividing the missing segment in two parts of half the size). The number of predicted points was allowed to exceed past half of the segment, and the two predictions were merged
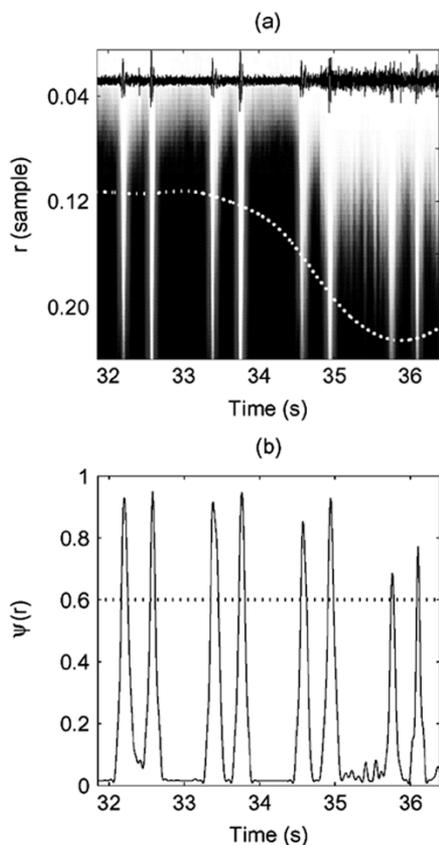
Fig. 2. In (a), Ψ is plotted over time for various r-values where the grayscale indicates the strength of Ψ. Superimposed in the figure is the LS signal (black waveform) and the chosen r-value (white waveform). $\Psi(r)$, determined by the r-values chosen in (a), is plotted in (b) along with a threshold to extract HS.
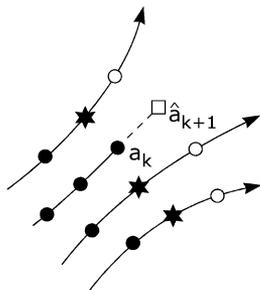

Fig. 3. Three trajectory segments and a (forward) predicted trajectory in a two-dimensional phase space. The average change between the nearest neighboring trajectory points (black stars) and their successors (white circles) is used to predict the next point (white square).

in the time domain close to the midpoint at an intersection with similar slope.

IV. RESULTS AND DISCUSSION

The complete data set consisted of 604 heart cycles (1208 first and second HS). Table I summarizes the results from the HS detection algorithm. A correct detection had to cover a whole HS (determined by visual inspection), and even though some HS were correctly marked, they were still considered erroneous if they did not cover the entire sound (when a HS is not completely covered, the prediction will be based on HS data and the results will not be satisfactory). The error rate for the whole

TABLE I
RESULTS FROM HS DETECTION

Subject	Number of HS	Number of false positive [b]	Number of false negatives [c]
1[a]	208	3	9
2	174	8	23
3	168	5	17
4	220	3	34
5	140	15	11
6	298	17	5
Total	1208	51 (4%)	99 (8%)

[a]Subject 1 had a third heart sound that was excluded in the study.
[b]Mostly friction rubs.
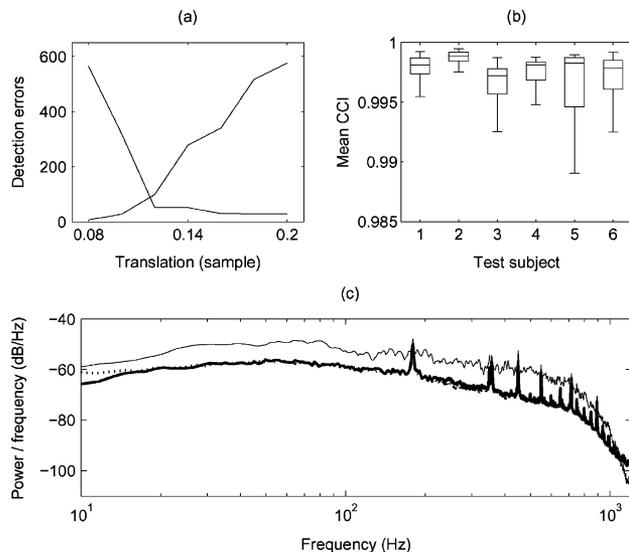[c]Most of which are found during high flows near vital capacity.

Fig. 4. (a) Total amount of false positives and false negatives for different choices of the translation parameter when determining the threshold in the detection algorithm. (b) Values of lower quartile, mean, and upper quartile (boxes) of the estimated CCI of each test subject. The whiskers show the range of the data. (c) Illustrates the PSD of the original signal (thin line), the original signal with HS removed (thick line), and the signal where HS have been replaced by nonlinear prediction (dashed). The spectra showed are averages over all subjects and all periods. The peak at 180 Hz and its accompanying harmonics are due to a computer fan in the measurement equipment.

material was 4% false positives and 8% false negatives. When matching the adaptive threshold with appropriate r-values, the choice of the scaling and translation parameters is rather sensitive [see Fig. 4(a)]. The difference between different data sets was, however, small, so the same parameters (translation $= 0.11$ and scaling $= 0.2$) were used throughout the whole material.

Differentiating between different detections is a problem yet to be solved. In this letter, the measurements contained several friction rubs (the main reason for false positive detections in Table I), and one test subject had a distinct third HS. Most of these friction rubs and third HS were detected and removed along with the HS. LS with similar characteristics will probably be marked by the method as well, and particularly, crackles present a potential problem. By including extra criteria, such as spectral characteristics or the degree of impulsiveness, it is possible that these false detections could be avoided.

The embedding dimension for phase-space reconstruction was found to be $d = 7$. The delay parameter, on the other hand, was allowed to vary since the average mutual information showed that τ fluctuated heavily between 8 and 17.

TABLE II
DIFFERENCES IN DECIBELS PER HERTZ (mean ± std) BETWEEN LS (WITH REMOVED HS) AND PREDICTED LS. THE DIFFERENT PERIODS REPRESENT THE THREE MEASUREMENT PHASES

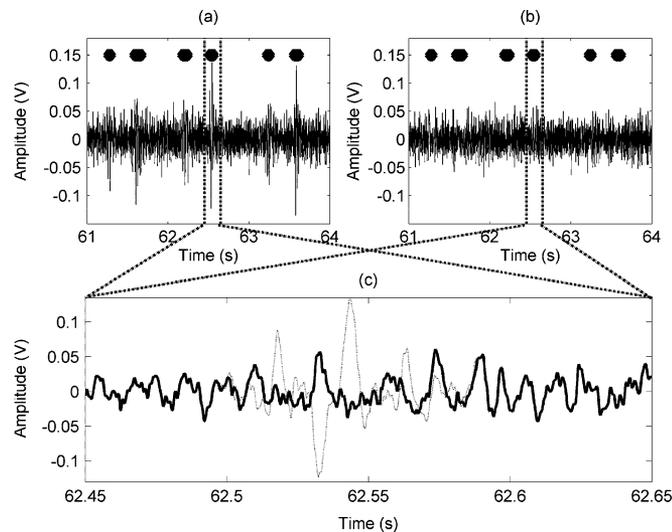| Phase	20-40 Hz	40-70 Hz	70-150 Hz	150-300 Hz
Period 1	0.62±0.37	0.61±0.33	0.44±0.31	1.14±0.84
Period 2	2.55±0.52	3.00±0.67	2.59±0.44	2.36±0.67
Period 3	1.58±0.82	5.18±1.52	2.89±1.51	2.27±1.22
All	0.34±0.25	0.50±0.33	0.46±0.35	0.94±0.64

Fig. 5. Example of the (a) recorded LS with HS present and (b) reconstructed LS with HS removed. The bars indicate HS detections. A zoomed in version of predicted LS (solid) and LS including HS (dashed) are shown in (c).

The HS cancellation was evaluated by comparing power spectral densities (PSDs) between the predicted signal and the original signal where the HS had been cut out (see Fig. 4). Table II shows the difference between the original data (without HS) and the predicted LS data, divided into four subbands: 20–40, 40–70, 70–150, and 150–300 Hz (since HS have most of their energy below 300 Hz). As can be seen in Fig. 4(c), the spectrum of the reconstructed signal is very close to the LS free of HS. However, comparing PSDs might be questionable as phase information is lost. The difference in the low-frequency part of the spectrum is an error caused by false HS detections (poorly cut out HS). If available segments contain HS in the end points, the prediction will try to recreate a HS, which results in predictions with similar frequency content as HS.

The waveform similarity between predicted segments and actual LS data is very high with a cross-correlation index of $CCI = 0.997 \pm 0.004$ [see Fig. 4(b)]. Since the main objective of the method was to give auditory high-quality results, a simple complementary listening test was performed by a skilled primary health-care physician. The impression was that most HS had been successfully replaced but that some predictions had a slightly higher pitch than pure LS. An example of HS cancellation is illustrated in Fig. 5. The presented method gives very good auditory results, though it must be confirmed by applying the method to signals from patients with different lung diseases.

REFERENCES

[1] R. Loudon and R. L. Murphy, Jr., "Lung sounds," *Amer. Rev. Respir. Dis.*, vol. 130, pp. 663–673, 1984.

[2] S. Charleston, M. R. Azimi-Sadjadi, and R. Gonzalez-Camarena, "Interference cancellation in respiratory sounds via a multiresolution joint time-delay and signal-estimation scheme," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 10, pp. 1006–1019, Oct. 1997.

[3] S. Charleston and M. R. Azimi-Sadjadi, "Reduced order Kalman filtering for the enhancement of respiratory sounds," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 4, pp. 421–424, Apr. 1996.

[4] L. J. Hadjileontiadis and S. M. Panas, "Adaptive reduction of heart sounds from lung sounds using fourth-order statistics," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 7, p. 642, Jul. 1997.

[5] M. T. Pourazad, Z. K. Moussavi, and G. Thomas, "Heart sound cancellation from lung sound recordings using adaptive threshold and 2D interpolation in time-frequency domain," in *Proc. 25th Annu. Int. Conf. IEEE Engineering Medicine Biology Society*, Cancun, Mexico, 2003, pp. 2586–2589.

[6] Z. K. Moussavi, D. Flores, and G. Thomas, "Heart sound cancellation based on multiscale products and linear prediction," in *Proc. 26th Annu. Int. Conf. IEEE Engineering Medicine Biology Society*, San Francisco, CA, 2004, pp. 3840–3843.

[7] H. Kitaoka, R. Takaki, and B. Suki, "A three-dimensional model of the human airway tree," *J. Appl. Physiol.*, vol. 87, pp. 2207–2217, 1999.

[8] B. J. West, "Physiology in fractal dimensions: Error tolerance," *Ann. Biomed. Eng.*, vol. 18, pp. 135–149, 1990.

[9] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. Cambridge, MA: Cambridge Univ. Press, 2004.

[10] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *J. Statist. Phys.*, vol. 65, pp. 579–616, 1991.

[11] J. McNames, "A nearest trajectory strategy for time series prediction," in *Proc. Int. Workshop Advanced Black-Box Techniques Nonlinear Modeling*, Leuven, Belgium, 1998, pp. 112–128.

[12] J. B. Gao, "Detecting nonstationarity and state transitions in a time series," *Phys. Rev. E*, vol. 6306, pp. 066 201–066 208, 2001.

[13] J. B. Gao, Y. Cao, L. Gu, J. G. Harris, and J. C. Principe, "Detection of weak transitions in signal dynamics using recurrence time statistics," *Phys. Lett. A*, vol. 317, p. 64, 2003.