

Modulation Filtering for Heart and Lung Sound Separation from Breath Sound Recordings

Tiago H. Falk and Wai-Yip Chan

Department of Electrical and Computer Engineering, Queen's University, Canada

Email: {falkt, chan}@ee.queensu.ca

Abstract—Separation of heart and lung sounds from breath sound recordings is a challenging task due to the temporal and spectral overlap of the two signals. In this paper, the use of a spectro-temporal representation to improve signal separation is investigated. The representation is obtained by means of a frequency decomposition (termed modulation frequency) of temporal trajectories of short-term spectral components. Experiments described herein suggest that improved separability of heart (HS) and lung sounds (LS) is attained in the modulation frequency domain. Bandpass and bandstop modulation filters are designed to separate HS and LS signals from breath sound recordings, respectively. Visual and auditory inspection, quantitative analysis, as well as algorithm execution time are used to assess algorithm performance. Log-spectral distances below 1 dB corroborate our listening test which found no audible artifacts in separated heart and lung sound signals.

Index Terms: Spectro-temporal processing, modulation filtering, modulation spectrum, heart sounds, lung sounds.

I. INTRODUCTION

Auscultation of lung sounds is a useful procedure for detection of pulmonary diseases. Lung sounds (LS) in breath sound recordings, however, are corrupted by intrusive quasi-periodic heart sounds (HS), which alter the temporal and spectral characteristics of the recording. Heart sound cancellation is a difficult task as both HS and LS signals have overlapping frequency spectra, in particular at frequencies below 150 Hz. The simplest method to reduce HS effects is to apply a highpass filter with cutoff frequency varying from 50-150 Hz. However, due to overlapping LS/HS spectra, highpass filtering also degrades the LS signal, which may lead to an erroneous diagnosis of pulmonary dysfunction.

More complex methods to *reduce* (or cancel) HS from breath sound recordings have been described in the literature. Proposed methods have been based on adaptive filtering techniques [1], wavelet denoising [2], combination of HS localization-and-removal and LS prediction [3], time-frequency filtering [4], or independent component analysis [5], to name a few. Several such approaches have been shown to reliably reduce HS effects [6]. Subjective analysis, however, has suggested that due to the temporal and spectral overlap between heart and lung sounds, HS reduction may result in noisy or possibly “non-recognizable” lung sounds [6].

In this paper, we venture away from conventional temporal and/or spectral signal representations and explore using an alternate spectro-temporal representation to *separate* HS and LS signals from breath sound recordings. In the proposed

domain, henceforth termed modulation frequency domain, enhanced separability of heart and lung sounds is attained. Linear phase bandpass and bandstop modulation filters are designed to separate HS and LS signals, respectively. We emphasize that the proposed method, unlike many methods proposed in the literature, does not depend on a reference heartbeat signal (commonly obtained from an electrocardiogram), on heart sound localization, or on lung sound prediction. Experiments with data from two healthy subjects suggest that accurate signal separation is attained and that audible artifacts are *not* introduced by the proposed separation method. Moreover, it is shown that algorithm execution time is approximately one order of magnitude lower than a HS cancellation algorithm based on time-frequency filtering.

The remainder of this paper is organized as follows. Section II introduces the spectro-temporal signal representation. Section III describes the proposed modulation filtering methodology and experimental results are presented in Section IV. Lastly, conclusions are presented in Section V.

II. SPECTRO-TEMPORAL REPRESENTATION OF BREATH SOUND RECORDINGS

Spectro-temporal processing consists of the frequency decomposition of *temporal trajectories* of short-term signal spectral components, as illustrated in Fig. 1 (a). The breath sound recording is windowed and a base transform (e.g., Fourier transform) is applied to generate the time-frequency (termed acoustic frequency to distinguish from modulation frequency) representation of the signal. A second transform, termed modulation transform, is then applied to the temporal trajectory of the magnitude component of each acoustic frequency signal. The resulting modulation spectrum contains information regarding the rate of change of signal spectral components. Note that if invertible transforms are used and phase components are kept unaltered, the original signal can be perfectly reconstructed [7].

The assumption investigated here is that the spectral content of heart sounds change at rates different from that of the spectral content of lung sounds. Our experiments have suggested that important modulation spectral content of heart sounds typically fall between approximately 2-20 Hz. Lung sounds, on the other hand, are shown to be broadband with more important modulation frequency content situated at low frequencies (< 2 Hz); such behavior is expected due to the “stationarity” of lung sounds. The modulation spectra

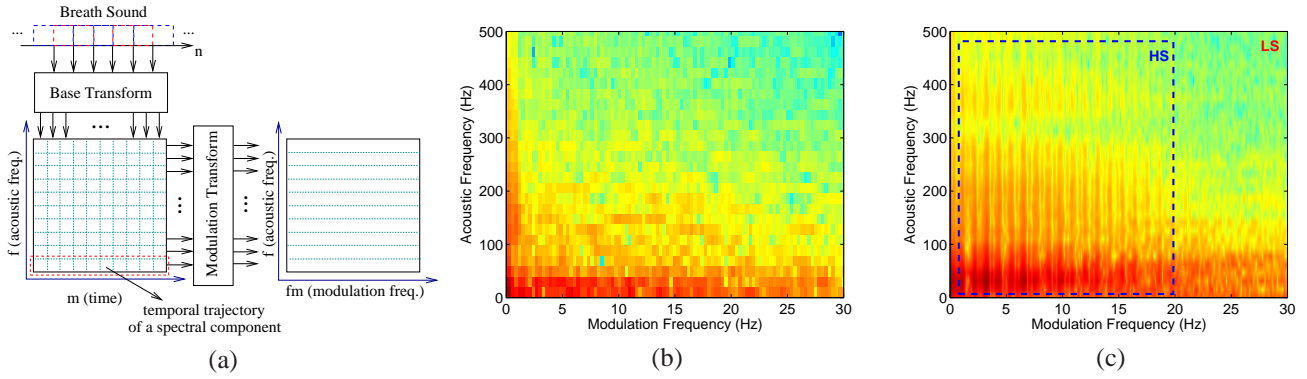


Fig. 1. Modulation spectral analysis – (a) signal processing steps, and the spectro-temporal representation of a breath sound recording with (b) showing mostly LS, and (c) showing overlapping HS and LS. Modulation spectral plots are zoomed in to depict acoustic frequencies below 500 Hz.

shown in Fig. 1 (b) and (c) assist in illustrating this behavior. Subplot (b) depicts the modulation spectrum of a segment of the breath sound recording containing mostly LS (recorded at the right 4th intercostal space during high airflow rate). Subplot (c) depicts the modulation spectrum of a segment of breath sound recorded at the center of the chest during medium airflow, thus containing both HS and LS. As noted, both signals contain important (and overlapping) frequency content at acoustic frequencies ranging from 15-50 Hz. The modulation frequency axis, however, introduces an additional dimension at which improved separability can be attained. As a consequence, modulation filtering is proposed to blindly separate heart and lung sounds from breath sound recordings.

III. MODULATION FILTERING: METHODOLOGY

Modulation filtering is described as filtering of the temporal trajectories of short-term spectral components. For the sake of notation, let $s(f, m)$, $f = 1, \dots, N$ and $m = 1, \dots, T$, denote the short-term spectral component at the f^{th} frequency bin and m^{th} time step of the short-term analysis. N and T denote total number of frequency bands and time steps, respectively. For a fixed frequency band $f = F$, $s(F, m)$, $m = 1, \dots, T$, represents the F^{th} band temporal trajectory. Two finite impulse response modulation filters are employed and depicted in Fig. 2. The first is a bandpass filter with cutoff modulation frequencies at 1 Hz and 20 Hz (dotted line); the second is the complementary bandstop filter (solid line). Modulation frequencies above 20 Hz are kept as they are shown to improve the naturalness of separated LS signals.

A. Signal Processing

In our experiments, the Gabor transform is used for spectral analysis. The Gabor transform is a unitary transform (energy is preserved) and consists of an inner product with basis functions that are windowed complex exponentials. In our experiments, doubly over-sampled Gabor transforms are used and implemented based on discrete Fourier transforms (DFT), as depicted in Fig. 3. First, the breath sound recording is windowed by a power complementary window (more details in Section III-B). An N -point DFT is then taken and the magnitude ($|s(f, m)|$) and phase ($\angle s(f, m)$) components

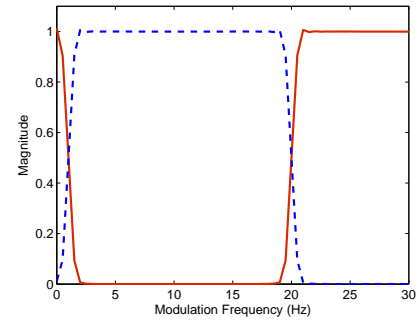


Fig. 2. Magnitude response of bandpass and bandstop modulation filters.

of each frequency bin are input to a so-called “modulation processing” module.

Modulation filtering and phase delay compensation are performed in the modulation processing module. The “per frequency bin” magnitude trajectory $|s(f, m)|$, $m = 1, \dots, T$ is filtered using the bandpass and the bandstop modulation filters to generate signals $|\hat{s}(f, m)|$ and $|\bar{s}(f, m)|$, respectively. The remaining modulation processing step consists of delaying the phase by an integer number of samples; the delay is dependent on the order of the linear phase filter used (more details in Section III-B). The outputs of the modulation processing modules are the bandpass and bandstop filtered signals and the delayed phase components $\angle \bar{s}(f, m)$. Two N -point IDFTs are then taken. The first IDFT (namely IDFT-1) takes as input the N $|\hat{s}(f, m)|$ and $\angle \bar{s}(f, m)$ signals to generate $\hat{s}(m)$. Similarly, IDFT-2 takes as input signals $|\bar{s}(f, m)|$ and $\angle \bar{s}(f, m)$ to generate $\bar{s}(m)$. The outputs of the IDFT-1 and IDFT-2 modules are windowed by the power complementary window and overlap-and-add is used to reconstruct HS and LS signals, respectively.

B. Implementation Details

In our experiments, a square-root Hann window of length 20 milliseconds with 50% overlap (frame shifts of 10 milliseconds) is used for the Gabor transform. In order to attain accurate resolution at 1 Hz modulation frequency, higher order filters are needed. Here, 150-tap linear phase filters are used; such filter lengths are equivalent to analyzing 1.5 s

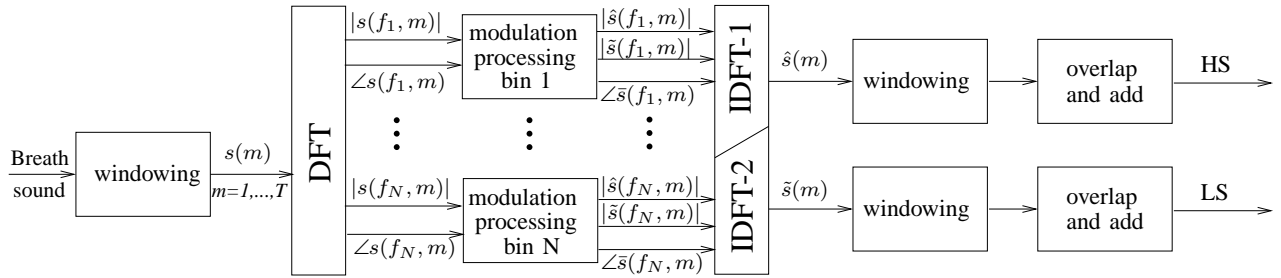


Fig. 3. Block diagram of the proposed modulation filtering approach to blind HS separation from LS recordings.

temporal trajectories. Moreover, it is observed with bandpass filtered signals that the removal of lowpass modulation spectral content may result in negative power spectral values. As with the spectral subtraction paradigm used in speech enhancement algorithms, a half-wave rectifier can be used. Rectification, however, may introduce unwanted perceptual artifacts to the separated HS signal. To avoid such artifacts, one can opt to filter the cubic-root compressed magnitude trajectories in lieu of the magnitude trajectories. In such instances, cubic power expansion must be performed prior to taking the IDFT. In our experiments, cubic compression-expansion of bandpass filtered signals is used and negligible rectification activation rates (<2%) are obtained.

IV. EXPERIMENTS

In this section, a description of the data used in our experiments is given and experimental results are presented.

A. Data: Breath Sound Recordings

The University of Manitoba breath sound recordings are used in our experiments; the data has been made publicly available by the Biomedical Engineering Laboratory. Data is obtained from two healthy subjects aged 25 and 30 years. Piezoelectric contact accelerometers were used to record the respiratory sounds from the subjects in sitting position. Accelerometers were secured with double-sided adhesive tape rings at the following five locations: (1) right and (2) left midclavicular, 2nd intercostal space, (3) right and (4) left 4th intercostal space, and (5) center of chest.

Subjects were asked to maintain their target breathing at low (7.5 ml/s/kg), medium (15 ml/s/kg), and high (22.5 ml/s/kg) flow rates. Subjects were instructed to breathe such that one full breath occurred every two to three seconds at every flow rate and had at least five breaths at each target flow. Three recordings were made per subject and each recording consisted of approximately 20 s at each target flow and concluded with an approximate 5 s of breath hold (total of ~ 65 s). During breath hold, subjects were asked to hold their breath with a closed glottis, thus allowing for a reference heartbeat signal and background noise characterization. Breath sound signals were digitized with 10240 Hz sample rate and 16-bit precision. In our experiments, data is downsampled to 5 kHz in order to reduce computational complexity. More detail about the data acquisition process can be found in [8].

B. Experimental Results

The proposed method is tested on breath sound signals captured at the five aforementioned locations. For the sake of brevity, the plots in Fig. 4 depict signals recorded at the center of the chest; similar performance is observed for signals captured at the other four locations. Subplot (a) depicts, from top to bottom, the airflow signal, breath sound signal, and the separated HS and LS signals. Note that HS signals are accurately separated even at high flow rates. Moreover, subplot (b) illustrates zoomed-in plots of a segment of low airflow breath sound along with the separated HS and LS signals. As observed, HS and LS signals are accurately separated even at low airflow rates.

Spectral plots are further depicted in Fig. 5. Subplot (a) illustrates the spectra of “HS-free” breath sounds and the separated LS signal. Power spectra are averaged over 5 s of HS-free breath sounds, which were randomly selected from segments of the breath sound recording between successive heartbeats (selected segments were within $\pm 20\%$ of the target low airflow rate). Similarly, subplot (b) depicts average power spectra of breath-hold sounds and the separated HS signal. Power spectra are averaged over the approximate 5 s breath-hold duration at the end of the recording session. In order to quantitatively assess the performance of the proposed method, the average log-spectral distance (LSD) between the aforementioned breath sound spectra $P(\omega)$ and separated signal spectra $\hat{P}(\omega)$ is used. The LSD, expressed in decibel, is given by

$$LSD = \sqrt{\frac{1}{2\pi} \int_{-\omega}^{\omega} \left[10 \log_{10} \frac{P(\omega)}{\hat{P}(\omega)} \right]^2 d\omega}. \quad (1)$$

In speech coding research, two signals with $LSD < 1$ dB are considered to be perceptually indistinguishable [9]. Using this same difference limen for spectral transparency, it is observed that average LSD of 0.61 dB and 0.31 dB are attained for the separated LS and HS signals, respectively. Hence, audible artifacts are not introduced by the proposed separation method; this is further corroborated by listening to the separated LS and HS signals.

C. A Note on Algorithm Execution Time

Algorithm execution time is also an important figure of merit. The proposed algorithm is implemented using Matlab version 7.2 Release 2006a. Simulations are run on a PC

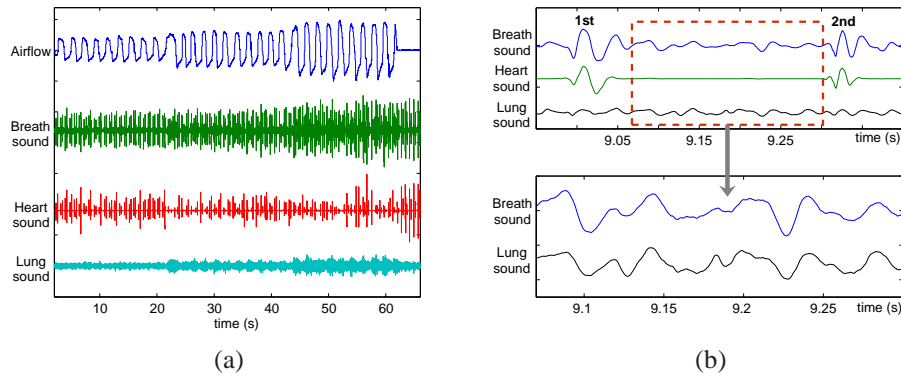


Fig. 4. Waveforms from top to bottom: (a) airflow signal, breath sound recording, separated HS, and separated LS; and (b) zoomed in plots of a segment of breath sound and the respective separated heart and lung sounds.

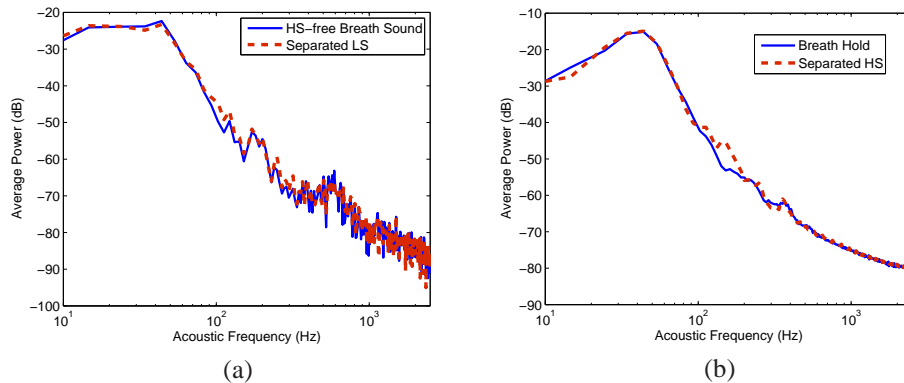


Fig. 5. Spectral plots of breath sounds and (a) separated LS and (b) HS signals.

with a 2.8 GHz Pentium 4 processor and 2 GB of RAM. The execution time for heart and lung sound separation of a 65 s breath sound recording, downsampled to 5 kHz, is approximately 5.04 s (i.e., $0.08 \times \text{realtime}$); execution time doubles if the original sampling rate is used. Furthermore, if only bandstop filtering is performed (equivalent to HS cancellation), algorithm execution time is reduced to 3.16 s for 5 kHz sample rate. Hence, the computational load of the proposed algorithm is several orders of magnitude lower than that reported in [4] for HS cancellation algorithms based on adaptive filtering (12 h), wavelet denoising (7 min), and time-frequency filtering (2 min) using a similar 60 s breath sound recording with 10240 Hz sample rate.

V. CONCLUSION

Conventional approaches to heart and lung sound separation are based on temporal or spectral signal processing techniques. A major disadvantage of such approaches, however, is that both sound signals overlap in the time and frequency domains. In this paper, we have proposed an alternate spectro-temporal signal representation which introduces an additional dimension (termed modulation frequency) at which improved signal separability can be attained. Visual inspection suggests successful signal separation. Quantitative analysis is used to demonstrate that audible artifacts are not introduced to separated heart and lung sounds; informal listening tests corroborate such findings. Moreover, algorithm

execution time is reduced by approximately one order of magnitude relative to a state-of-art HS cancellation algorithm based on time-frequency filtering.

REFERENCES

- [1] L. Hadjileontiadis and S. Panas, "Adaptive reduction of heart sounds from lung sounds using fourth-order statistics," *IEEE Trans. on Biomedical Eng.*, vol. 44, no. 7, pp. 642–648, July 1997.
- [2] D. Flores-Tapia, Z. Moussavi, and G. Thomas, "Heart sound cancellation based on multiscale products and linear prediction," *IEEE Trans. on Biomedical Eng.*, vol. 54, no. 2, pp. 234–243, Feb. 2007.
- [3] C. Ahlstrom, O. Liljefeldt, P. Hult, and P. Ask, "Heart sound cancellation from lung sound recordings using recurrence time statistics and nonlinear prediction," *IEEE Signal Proc. Letters*, vol. 12, no. 12, pp. 812–815, Dec. 2005.
- [4] M. Pourazad, Z. Moussavi, and G. Thomas, "Heart sound cancellation from lung sound recordings using time-frequency filtering," *Journal of Medical and Biological Eng. and Computing*, vol. 44, no. 3, pp. 216–225, March 2006.
- [5] J.-C. Chien, M.-C. Huang, Y.-D. Lin, and F.-C. Chong, "A study of heart sound and lung sound separation by independent component analysis technique," in *Proc. EMBS Intl. Conf.*, Sept. 2006, pp. 5708–5711.
- [6] J. Gnitecki, I. Hossain, H. Pasterkamp, and Z. Moussavi, "Qualitative and quantitative evaluation of heart sound reduction from lung sound recordings," *IEEE Trans. on Biomedical Eng.*, vol. 52, no. 10, pp. 1788–1792, Oct. 2005.
- [7] M. Vinton and L. Atlas, "A scalable and progressive audio codec," in *Proc. Intl. Conf. Audio, Speech, and Signal Proc.*, May 2001, pp. 3277–3280.
- [8] M. Pourazad, Z. Moussavi, F. Farahmand, and R. Ward, "Heart sounds separation from lung sounds using independent component analysis," in *Proc. IEEE Intl. Conf. Eng. in Medicine and Biology*, Sept. 2005.
- [9] W. B. Kleijn and K. K. Paliwal, Eds., *Speech Coding and Synthesis*. Amsterdam: Elsevier, 1995.