This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2023.3308296

A Novel Enhancement Approach Following MVMD and NMF Separation of Complex Snoring Signals

Mariam Al Mawla, Kabalan Chaccour, *Member, IEEE*, and Hoda Fares, *Member, IEEE*

***Abstract*— Snoring is a prominent characteristic of sleep-disordered breathing, and its detection is critical for determining the severity of the upper airway obstruction and improving daily quality of life. Home snoring analysis is a highly invasive method, but it becomes challenging when a sleeping partner also snores, leading to distorted evaluations in such environments. In this paper, we tackle the problem of complex snore signal separation of multiple snorers. This article introduces two audio-based methods that efficiently extract an individual's snoring signal, allowing for the analysis of sleep-breathing disorders in a normal sleeping environment without isolating individuals. In the first method, Principal Component Analysis (PCA) identifies the source components from the finite number of modes generated by the decomposition of the snoring mixture using Multivariate Variational Mode Decomposition (MVMD). The second method applies Blind Source Separation (BSS) based on Non-Negative Matrix Factorization (NMF) to separate the single-channel snoring mixture. Furthermore, the decomposed signals are tuned using the iterative enhancement algorithm to adequately match the source snoring signals. These methods were evaluated by simulating various real-time snoring recordings of 7 subjects (2 men, 2 women, and 3 children). The correlation coefficient between the source and its separated signal was computed to assess the separation results, exhibiting good performance of the methods used. The enhancement approach also demonstrated its efficiency by increasing the correlation over to 80% in both methods. The experimental results show that the proposed algorithms are effective and practical for separating mixed snoring signals.**

***Index Terms*— Snoring, MVMD, NMF, single-channel mixture, signal Separation, signal enhancement.**

"This work was supported in part by the TICKET Research Lab. at the Antonine University"

Mariam Al Mawla is with the TICKET Lab., Dept. of Computer and Communications Engineering Antonine University, Hadat-Baabda, Lebanon (e-mail: mariamalmawla1@gmail.com).

Kabalan Chaccour is with the Nanomedicine, Imagery and Therapeutics (NIT Lab.), University of Technology Belfort-Montbeliard, Besançon, France and the TICKET Lab., Dept. of Computer and Communications Engineering, Antonine University, Hadat-Baabda, Lebanon (e-mail: kabalan.chaccour@gmail.com)

Hoda Fares is with the Dept. of Electrical, Electronic, and Telecommunications Engineering and Naval Architecture, University of Genoa, Italy (e-mail: hodafares1@gmail.com).

I. Introduction

SNORING is an important indicator of upper airway dysfunction, and it is directly associated with sleep-disordered breathing like Obstructive Sleep Apnea (OSA). In fact, snoring is defined by the American Sleep Diseases Association (ASDA) as "a noise that is in the upper airway" [1]. Detection of snoring is important for improving quality of daily human life, thereby helping to identify and reduce the cause of sleep disorders. In fact, according to the American Academy of Sleep Medicine (AASM), around 40% of adult men and 24% of adult women are habitual snorers [2]. Snoring levels vary between and within individuals. Every individual snores at a different frequency and the sounds produced are characterized as a spasmodic signal. Snoring could reach high noise-level in the range of 80-90 dB [3].

An extravagant clinical method using numerous sensors called Polysomnography (PSG) [4] is considered the most reliable diagnosis for snoring. It records oxygen levels in the blood, eye movements, breathing functions, and heart rhythm [5]. However, PSG is expensive, complex, and requires an overnight sleep in the laboratory, which delays diagnosis and treatment [6]. Despite its outstanding functionality, patients avoid the PSG test due to its significant expense, cumbersome procedure, and long overnight waiting time [7]. According to [7], more than 90% of people who are at risk of OSA go undetected and untreated. This life-threatening sleep disorder can lead to high blood pressure, coronary heart disease, pulmonary heart failure, and even nocturnal death [8]. Studies have found that snoring identifies the severity and the obstruction site in the upper airway of OSA patients [9]. This significant discovery has motivated several scientists to propose reasonable and non-invasive snoring monitoring and screening methods [9]. Researchers in [10]–[13] have recently addressed the issue of home snoring monitoring based on sleep audio recordings. In these studies [10]–[13], the sound is generally recorded under controlled conditions with careful management of the acoustic surroundings. However, home-based snoring analysis may be affected by the snoring emitted from two or more sleeping individuals, resulting in a mixture of snoring sounds.

This research strives to provide a comfortable, long-term, and reliable analysis of snoring in a normal sleep environment without requiring separate laboratory tests. Therefore, this study attempts to extract individual snoring from single-channel snoring mixtures using two methods. The first method consists of decomposing the snoring mixture into multiple band-limited modes by applying Multivariate Variational Mode Decomposition (MVMD) [14] and identifying the principal components for the reconstruction of the snoring sources using the Principal Component Analysis (PCA) [15]. In the second method, Non-negative Matrix Factorization (NMF) [16] is used to extract individual snoring from a set of complex snoring signals. Indeed, MVMD and NMF approaches are applied to mixed snoring signals for the first time, to the best of our knowledge. In addition, the separated signals generated by the BSS methods are subjected to an enhancement approach that leads to an accurate extraction of individual snoring signals with a correlation coefficient between 80% and 99%.

This paper is organized as follows. Section II gives an overview of the existing work on snoring and audio signal separation methods. Section III provides a background on MVMD and NMF audio separation methods. Section IV describes our proposed iterative approach to enhance snore signal decomposition in both MVMD and NMF. Section V illustrates the simulation scenarios taken into account for the proof of concept. Section VI discusses the results. Section VII concludes the paper.

II. Related Work

This section summarizes the recent work conducted to separate mixed snoring signals using analytical methods and the audio source separation methods that can be employed for different acoustic signals.

A. Snoring Separation Methods

In previous work [17], Nigam *et al.* addressed the separation of mixed snoring signals by designing a Blind Delayed Source Separation (BDSS) algorithm with a delayed mixing model considering the source propagation delays and the attenuation of the snore to reach the sensor. However, this snore separator is based on the manual assignment of the distance between the snorer and the sensor requiring the supervision of a physician during sleep rather than self-monitoring of snoring. Additionally, Perez-Macias *et al.*, in [18], proposed an automatic snore detection method using Non-Negative Matrix Factorization Deconvolution (NMFD) to extract snoring signals from a single-channel mixture of signals. The NMFD technique was applied to an Emfit (Electromechanical film transducer) signal, which included snoring signals, heartbeat and breathing signals, and noise. As well, a wavelet de-noising technique was employed to eliminate Gaussian noise from the decomposed signals. According to the findings reported in [18], the developed method demonstrated a sensitivity of 83% and a positive predictive value of 86% for snoring detection using the Emfit signal. However, it is important to note that the study by Perez-Macias *et al.* lacked the possibility of the presence of a second snoring signal that could potentially interfere with the snoring patterns of the tested individual. Furthermore, the separation of complex snoring signals was carried out by Romero *et al.* in [19] using Deep Neural Networks (DNNs).

B. Audio Source Separation Methods

In [20], Guo *et al.* proposed a new method that combines Ensemble Empirical Mode Decomposition (EEMD), PCA, and ICA to decompose mixed speech, audio, and biomedical signals without prior knowledge of the sources. The extracted speech signal showed a correlation of 88% when combined with a sinusoidal signal. In contrast, the surface electromyography (sEMG) signal was retrieved from a real-life electrocardiography (ECG) signal with a correlation of 77%. This method was also used in the mechanical fault diagnosis in [21], resulting in 96% and 88% correlation of the outer and inner ring signals, respectively. Furthermore, Dey *et al.* developed in [22] a single-channel source separation method based on the Variational Mode Decomposition (VMD) and PCA. VMD is a decomposition method published by Dragomiretskiy *et al.* in [23]. It divides univariate data into a finite number of intrinsic modes. In [22], tests performed on a mixture of male speech and a sinusoidal signal showed a correlation of 81% for speech and 99% for the sinusoidal signal. MVMD, a generic extension of VMD, was suggested by Naveed *et al.* in [14] to decompose multivariate data into its inherent multivariate modulated oscillations. This extension of VMD was applied by Gavas *et al.* in [24] to real-world applications such as electroencephalogram (EEG) and cardiotocographic (CTG) signals. In addition, in a research carried out by Ranny *et al.* [25], NMF was applied to separate overlapping sounds. The resulting factorization was classified using Support Vector Machines (SVMs). That study demonstrated that NMF was effective in distinguishing sounds and improved sound identification performance. Paris Smaragdis has also targeted speech source separation with NMF in various studies [26]–[29].

III. Background on MVMD and NMF

This section gives a background on the MVMD and NMF separation methods for general sound signals.

A. Multivariate Variational Mode Decomposition (MVMD)

This subsection generalizes the VMD concept, as Naveed *et al.* described it in [14], to multivariate data by extracting K multivariate modulated oscillations $u_k(t)$ residing in the input signal $x(t)$ containing C data channels, as expressed by

$$x(t) = \sum_{k=1}^{K} u_k(t) \tag{1}$$

where u_k is the k^{th} narrow-band mode from a total of K modes. MVMD decomposes the input signal into K Intrinsic Mode Functions (IMFs) present in the input signal having specified scattered features. Each mode is assumed to be restricted around the central frequency. To estimate the bandwidth of each mode, the process explained in [14] and [23] must be followed.

First, the analytical signal of each mode u_k is computed using the Hilbert Transform (HT) to obtain the unilateral frequency spectrum of each mode. Second, the spectrum of each mode is shifted to the baseband by mixing with an exponential tuned to the estimated center frequency. Finally, the bandwidth of each spectrum is estimated through the H^1 Gaussian smoothness of the shifted signal, which is the squared L^2-norm of the gradient. Thus, the resulting constrained optimization problem in [14] is expressed by

$$\min_{\{u_{k,c}\},\{\omega_k\}} \left\{ \sum_k \sum_c \left\| \partial_t \left[u_+^{k,c}(t) e^{-j\omega_k t} \right] \right\|_2^2 \right\} \tag{2}$$

subject to $\sum_k u_{k,c}(t) = x_c(t)$, where $u_{k,c}$ represents the decomposed mode u_k of channel c and ω_k is the corresponding central frequency of each mode. The complex-valued signal $u_+^{k,c}(t)$ represents the analytic modulated signal of mode k and channel c [14]. Equation (2) is solved by introducing the Lagrange function generating the unconstrained variational problem [14] designated as follows:

$$\begin{aligned} &\mathcal{L}(\{u_{k,c}\},\{\omega_k\},\lambda_c) = \\ &\alpha \sum_k \sum_c \left\| \partial_t \left[u_+^{k,c}(t) e^{-j\omega_k t} \right] \right\|_2^2 \\ &+ \sum_c \left\| (x_c(t) - \sum_k u_{k,c}(t)) \right\|_2^2 \\ &+ \sum_c \left\langle \lambda_c(t), x_c(t) - \sum_k u_{k,c}(t) \right\rangle \end{aligned} \tag{3}$$

where α is the moderate bandwidth constraint, which determines the bandwidth of the reconstructed modes. λ_c represents the Lagrange multiplier of channel c to ensure that the constraints are strictly fulfilled [14].

The optimal solution of the above augmented Langrage function is obtained by the Alternate Direction Method of Multipliers (ADMM) [14]. ADMM is an algorithm that solves complex optimization problems by dividing them into multiple and simpler sections that can then be handled independently [30]. The unconstrained sub-problems are resolved in the Fourier domain to obtain the appropriate update of the mode and the central frequency in the frequency domain using the ADMM algorithm. Algorithm 1 shows the pseudo-code of MVMD [14] that is applied to decompose complex snoring signals in this paper.

Algorithm 1 Algorithm of MVMD

Initialize $\{\hat{u}^1_{k,c}\}$, $\{\omega^1_k\}$, $\hat{\lambda}^1_c$, $n \leftarrow 0$
repeat
  $n \leftarrow n+1$
  for $k \leftarrow 1 : K$ **do**
    for $c \leftarrow 1 : C$ **do** *Update mode* $\hat{u}_{k,c}$

$$\hat{u}^{n+1}_{k,c}(\omega) \leftarrow \frac{\hat{x}_c(\omega) - \sum_{i\neq k} \hat{u}_{i,c}(\omega) + \frac{\hat{\lambda}^n_c(\omega)}{2}}{1 + 2\alpha(\omega - \omega^n_k)^2}$$

    end for
  end for
  for $k \leftarrow 1 : K$ **do** *Update center frequency* ω_k

$$\omega^{n+1}_k \leftarrow \frac{\sum_c \int_0^\infty \omega |\hat{u}^{n+1}_{k,c}(\omega)|^2 d\omega}{\sum_c \int_0^\infty |\hat{u}^{n+1}_{k,c}(\omega)|^2 d\omega}$$

  end for
  for $c \leftarrow 1 : C$ **do** *Update* λ_c *for all* $\omega \geq 0$

$$\hat{\lambda}^{n+1}_c(\omega) \leftarrow \hat{\lambda}^n_c(\omega) + \tau(\hat{x}_c(\omega) - \textstyle\sum_k \hat{u}^{n+1}_{k,c}(\omega))$$

  end for
until convergence: $\sum_k \sum_c \frac{\|\hat{u}^{n+1}_{k,c} - \hat{u}^n_{k,c}\|_2^2}{\|\hat{u}^n_{k,c}\|_2^2} < \varepsilon$

B. Non-Negative Matrix Factorization (NMF)

This subsection describes NMF as a blind source separation technique that was first introduced by Paatero *et al.* in [31] and popularized in an article by Lee *et al.* in [32]. NMF has proved its efficiency in the statistical analysis of multivariate data where it automatically extracts sparse and meaningful features from a set of non-negative data vectors [16]. NMF is commonly used in feature recognition [33], and audio source separation [34]–[36]. It has also been used in the medical field for the separation of cardiac and respiratory sounds from a single mixture [37]. It approximates an M-by-N matrix V into the product of two matrices W and H having dimensions $M \times T$ and $T \times N$, respectively [16], such that

$$V \approx WH \tag{4}$$

where N is the number of samples in the data set and $T \leq \min(M, N)$. NMF can factorize V into two lower-rank matrices W and H resulting in a compressed version of the original matrix V. The decomposed matrix W contains the basis elements that are optimized for the linear approximation of the data in V, while H is the activation matrix that gives the coefficients or weights associated with each data point in the basis W [33].

To perform the approximate factorization, cost functions that evaluate the quality of the approximation are defined. The first assessment is constructed using the Euclidean distance between V and WH shown in equation (5) [38] that is lower-limited at zero and vanishes when $V = WH$ [16]:

$$\|V - WH\|^2 = \sum_{i=1}^{M} \sum_{j=1}^{N} \big(V_{ij} - (WH)_{ij}\big)^2 \tag{5}$$

Furthermore, another measurement arises referred to the divergence between V and WH presented in equation (6) [38], which is also lower-bounded by zero and optimal if $V = WH$ [16]:

$$D(V\|WH) = \sum_{i=1}^{M} \sum_{j=1}^{N} \Big(V_{ij} \log \frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij} \Big) \tag{6}$$

NMF can be formulated as an optimization problem using equations (5) and (6), where the goal is to minimize $\|V - WH\|^2$ and $D(V\|WH)$ with respect to W and H under the constraint $W \geq 0$ and $H \geq 0$. These functions are convex in either W only or H only. Hence, Lee and Seung suggested in [16] the technique of multiplicative update rules to find the local minima. The multiplicative updates that converge the Euclidean distance to the local minima are given by

$$W \leftarrow W \odot \frac{VH^T}{WHH^T} \tag{7}$$

$$H \leftarrow H \odot \frac{W^TV}{W^TWH} \tag{8}$$

where $\odot$ denotes element-wise multiplication. The NMF approach is carried out by initializing the matrices W and H with random non-negative values [39]. Then, equations (7) and (8) are successively applied to update W and H until reaching the maximum number of iterations unless the multiplicative updates converge to a stationary point. NMF converges when the distance between two consecutive estimated spectrogram $\hat{V}$, after the multiplicative update of W and H, attains the stopping criterion threshold ε [40], given by

$$(\|V - \hat{V}_{\text{itr-1}}\|^2 - \|V - \hat{V}_{\text{itr}}\|^2) \leq \varepsilon \tag{9}$$

where the subscript “itr” denotes the number of iterations.

IV. Proposed Methodology

In this work, snoring separation requires the formation of overlapping data on which audio source separation techniques, such as MVMD and NMF can be applied. To our knowledge, these methods have not been applied to snoring signals yet. Then, the source separation methods assume a single-channel sleep audio signal with two or more snorers as input and perform an automated extraction of the embedded snore signals. Besides, the enhancing iterative approach will improve the decomposition result by adjusting the uncorrelated segments. The framework of the approach is illustrated in Fig. 1.

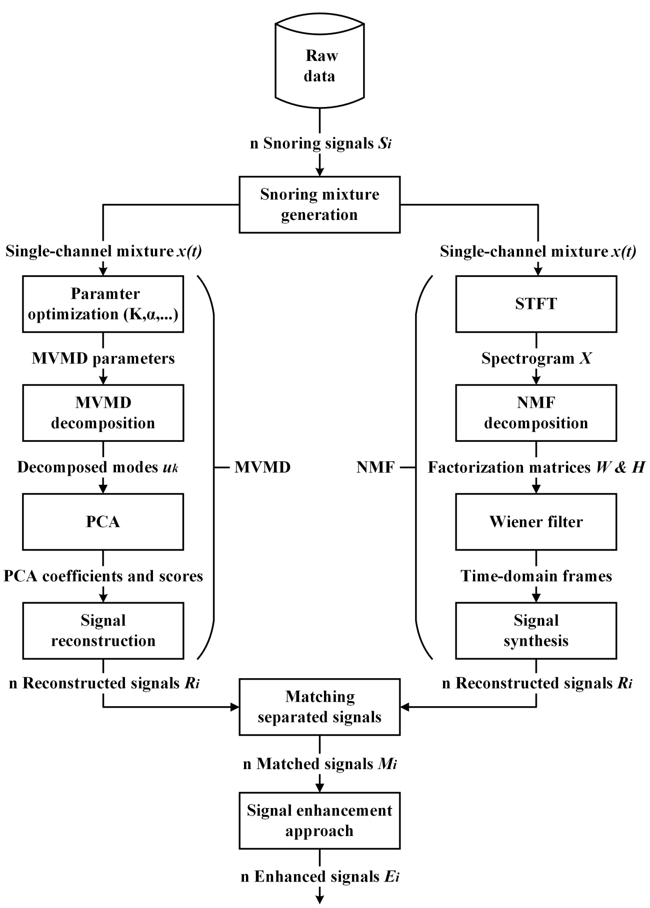

Fig. 1. Framework of the snoring separation methodology using MVMD and NMF

A. Snoring Mixture Generation

In a real-life context such as at home, the study of snoring can be disrupted by other acoustics, such as the snore of a sleeping partner, which results in a mixture of sounds. This mixture is composed of the snoring of both subjects. Consequently, the generation of a mixed signal can be computed by the summation of multiple snoring signals according to equation (10), resulting in a single-channel mixture $x(t)$.

$$x(t) = \sum_{i=1}^{n} S_i(t) \tag{10}$$

where S_i is the i^{th} snoring source given a total of n sources.

B. Decomposing the Snoring Mixture using MVMD

Before proceeding with the MVMD decomposition, various parameters must be initialized and assigned exact values, particularly the number of modes K and the penalty factor α. Other parameters usually have default values, tolerance ε is typically set to $1e^{-7}$, and noise tolerance τ to 0. In general, the Direct Component (DC) and the center frequency ω are initialized to 0. However, K is the most influential parameter that must be accurately pre-determined, where lower values of K lead to mode aliasing and under-segmented modes. Similarly, higher K values contribute to useless over-decomposition [41]. The resolved modes have a larger bandwidth when α is small. Wide bandwidth induces more background noise and interference elements in the output. Unfortunately, if the filter bandwidth is too narrow, MVMD findings can often be distorted [42]. According to MATLAB simulations, higher K and α values result in excessive processing time. Therefore, different values of α and K must be examined to determine the optimal range and accurately extract the signals.

The quality and the processing time of the decomposition determine the value of K. As such, K is set as a decreasing value initialized to a maximum and gradually decreased until it reaches the lowest convenient K while maintaining the integrity of the decomposition. The quality of the decomposition is evaluated by computing the correlation coefficient C_j between each source S_i and its decomposed mode u_k for every k. The value of K is assigned once the decomposed modes have the highest correlation C_j. The process of optimizing K is detailed in Algorithm 2.

Likewise, the bandwidth constraint α is assigned before it is processed to produce the expected signals. It specifies the bandwidth of the resulting modes. The aim is to generate a signal with almost the same bandwidth as that of the experimental sources. After each increase of $alpha$ by $value$, the frequency components $F(S_i)$ and $F(u_k)$ of each source

Algorithm 2 Optimization of the number of modes K

Initialize $K > n$
for $j \leftarrow K : n$ **do**
  Apply MVMD
  Select the mode u_k that corresponds to the source S_i
  $C_j(i) \leftarrow \text{corr}(S_i, u_k)$
end for
$K \leftarrow j$ where $C \leftarrow \max(C_j)$

S_i and its decomposed mode u_k, as well as the power levels $P(S_i)$ and $P(u_k)$ of each frequency, are examined. When the frequency components and their power levels are roughly related to those of the sources, optimal α is realized. The process of optimizing α is discussed in Algorithm 3.

Algorithm 3 Optimization of the bandwidth constraint α

Initialize: $alpha, value, flag \leftarrow FALSE$
while $flag \neq TRUE$ **do**
  Apply MVMD
  Select the mode u_k that corresponds to the source S_i
  Find the power spectrum of each u_k and S_i
  if $F(S_i) \simeq F(u_k)$ and $P(S_i) \simeq P(u_k)$ **do**
    $\alpha \leftarrow alpha$
    $flag \leftarrow TRUE$
  else
    $alpha \leftarrow alpha + value$
  end if
end while

Subsequently, the built-in MATLAB function "PCA()" is used to calculate the principal components of the decomposed modes for the extraction of the snoring signals. It is applied for identifying the decomposed modes that efficiently provide the source components. Karl Pearson introduced PCA in [15] to separate linear combinations of signals. It transforms a set of observed correlated vectors X into a set of uncorrelated variables Z, called principal components [15]. The principal components Z are the linear combination of the original data X with maximized variance under the constraint that these principal component vectors are orthogonal to each other. The orthogonal transformation is given by

$$Z = XV \tag{11}$$

where X is the data matrix of K decomposed modes, and V is the orthogonal transform matrix of n eigenvectors.

The first step in PCA is to standardize the dataset. The centered data is obtained after subtracting the mean vector μ from each column of the original data X. PCA computes the eigenvectors ("principal directions") of the covariance matrix XX^T and sorts them by their eigenvalues in descending order. The first n eigenvectors of the n source signals having the largest eigenvalues are selected as a result of the matrix decomposition. The centered data is then projected onto a subspace along the principal directions to yield principal components, also known as scores.

The MATLAB PCA function returns the eigenvector coefficients and the scores that correspond to the principal components, as well as the estimated mean when the input is centered in the pre-processing phase. It takes the modes as input as well as several parameters that can be changed according to the requirements, like the number of components n to be selected. After extracting the principal components required for generating n snoring signals from K modes, equation (12) is utilized to reconstruct the separated signals.

$$ReconstructedSignals = Z * V^T + \mu \tag{12}$$

C. Decomposing the Snoring Mixture using NMF

As a pre-processing step of the NMF decomposition, Short-Time Fourier Transform (STFT) is performed to convert the time-domain signal $x(t)$ into a magnitude spectrogram X. NMF visualizes X as the product of the basis elements W and the activation matrix H as in equation (4). Following the factorization, the spectrogram is computed using the same approximation as in (4). A Wiener filter is employed to depict the relative energy contribution of each source to the total energy of the mixture $x(t)$ to ensure a conservative reconstruction of the signals. The separated snoring signals are calculated using the time-domain frames, which are recovered by the Inverse Short-Time Fourier Transform (ISTFT) of the estimated magnitude spectrograms with the phase spectrogram of the input mixture $x(t)$.

D. Matching Separated Signals

An identification phase is required for the reconstructed signals R_i obtained following the signal reconstruction of MVMD modes and the signal synthesis of NMF factorization matrices. Accordingly, the selection of the signal that best matches the original sources is based on the correlation between each source and each reconstructed signal. The reconstructed signal R_i with highest correlation coefficient corresponding to each source S_i is selected to be assigned as a matched snoring signal M_i.

E. Enhancing Separated Snore Signals in MVMD and NMF

Even though the antecedent phases enabled the decomposition of the snoring sources, the separated signals may involve several uncorrelated segments that should be discarded. Thus, our approach revealed in Algorithm 4 is intended to fine-tune the decomposed signals.

Enhancing the decomposed signals consists of two phases. In the first phase, the segments of size seg of the matched signal M_i that have a lower correlation coefficient than the segments of the other matched signal are normalized to remove the uncorrelated parts. Once all the signals are tested, the correlation coefficient C_i between each source S_i and its enhanced separated signal E_i is examined. If C_i is still below a certain threshold (i.e. 80%), the enhanced signal E_i proceeds to the second phase of the enhancement method. In the second phase, another segmentation size seg can be assigned to examine the similarity between the segments of the sources

S_i and those of the reconstructed signals R_i. The uncorrelated sections of the enhanced signals E_i are replaced by the matching sections of the reconstructed signals R_i resulting from the MVMD reconstruction or the NMF synthesis. When the correlation coefficient C_{S-R} between the segments of the source S_i and the reconstructed signals R_i is greater than the correlation C_{S-E} between the segments of the source S_i and the enhanced signal E_i, R_i replaces these segments of E_i. The enhanced signals E_i are then normalized and processed with the first stage to ensure the exact separation of multiple snoring signals.

Algorithm 4 Approach for enhancing separated snoring signals

Input: Matched signals M_i, Source signals S_i, Reconstructed signals R_i
Output: Enhanced Separated Signals E_i
Initialize: segmentation size *seg*
$E_i \leftarrow M_i$
for $i \leftarrow 1 : n$ **do**
  for $j \leftarrow 1 : seg : length(E_i)$ **do**
    $C_i(:) \leftarrow corr(S_i(j : j + seg), E_i(j : j + seg))$
    if $C_i(j : j + seg) < C_{i+1}(j : j + seg)$ **do**
      Normalize $E_i(j : j + seg)$
    end if
  end for
end for
$C_i \leftarrow corr(S_i, E_i)$
if $C_i < threshold$ **do**
  (Optional) $seg \leftarrow [\ ]$
  for $j \leftarrow 1 : seg : length(E_i)$ **do**
    $C_i(:) \leftarrow corr(S_i(j : j + seg), R_i(j : j + seg))$
    $col \leftarrow$ i where $C \leftarrow max(C_i)$
    $C_{S-R} \leftarrow corr(S_{col}(j : j + seg), R_{col}(j : j + seg))$
    $C_{S-E} \leftarrow corr(S_{col}(j : j + seg), E_{col}(j : j + seg))$
    if $C_{S-R} > C_{S-E}$ **do**
      $E_{col}(j : j + seg) \leftarrow R_{col}(j : j + seg)$
    end if
  end for
  Normalize E_i
  for $j \leftarrow 1 : seg : length(E_i)$ **do**
    $C_i(:) \leftarrow corr(S_i(j : j + seg), E_i(j : j + seg))$
    if $C_i(j : j + seg) < C_{i+1}(j : j + seg)$ **do**
      Normalize $E_i(j : j + seg)$
    end if
  end for
end if

V. SIMULATIONS

A. Experimental Setup

In this section, the performance of the implemented methods are evaluated using actual recordings while eliminating experimental noise. Data merging was conducted due to the necessity of having separated data. Given that a mixed snore is recorded, the identity of each signal is lost, and the snoring structure will not be identical if recorded separately as in a mixed condition.

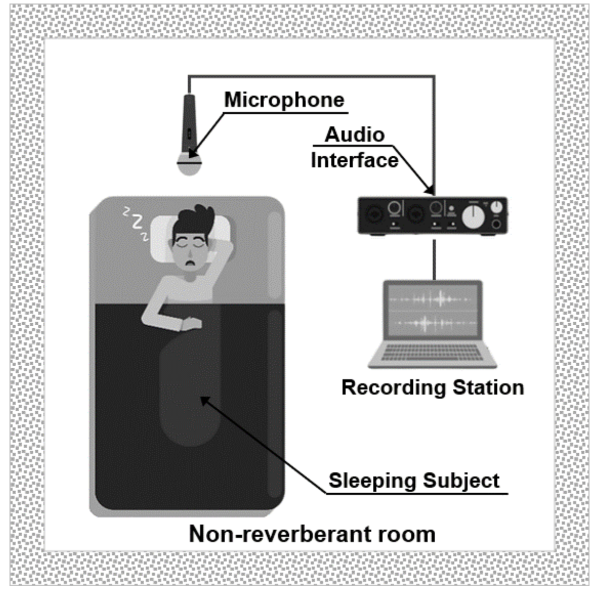

Fig. 2. Illustration of the recording setup

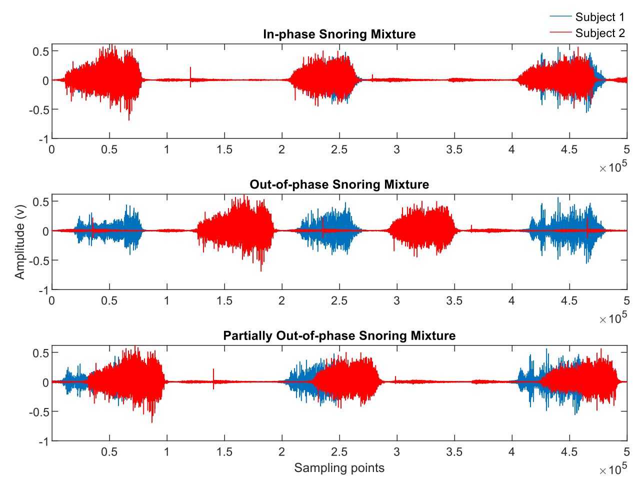

Fig. 3. Illustration of snoring mixtures for the three scenarios. top to bottom: in-phase, out-of-phase and partially out-of-phase

In addition, the recording setup of the mixture cannot be controlled to ensure the occurrence of the three scenarios illustrated in Fig. 3. In contrast to previous research, the snoring mixes were implemented in three expected scenarios namely, in-phase, out-of-phase, and partially out-of-phase.

Seven monophonic snoring signals were captured at a sampling frequency of f_s = 48 kHz, divided as follows: two men, two women, and three children. Fig. 2 illustrates the recording setup employed in this study, where snoring signals were individually recorded for each subject in a reverb-free room. To ensure high-quality recordings, a microphone was positioned adjacent to the sleeping subject, while a computer served as the recording station. The ages of the adults ranged from 22 to 50 years, and the ages of the children ranged from $1\frac{1}{2}$ to $3\frac{1}{2}$ years. Snore recordings are composed of consecutive periodic signals that are hard to be simulated in their original size, which is approximately 1 minute each. Thus, the number of samples is assumed to be 500000. The

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2023.3308296

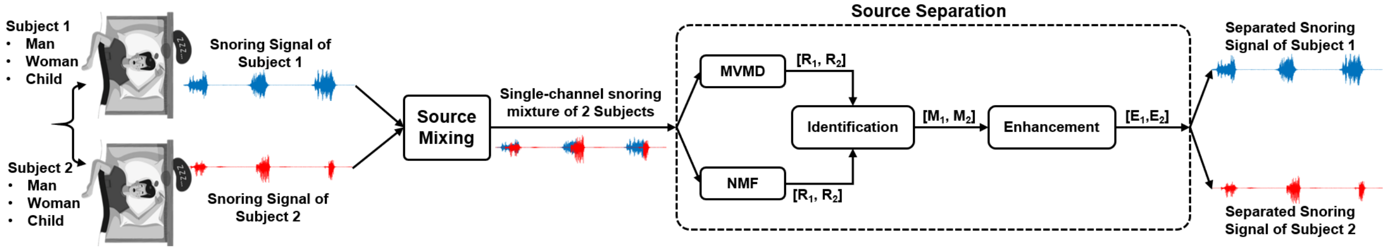

Fig. 4. Source separation of a snoring mixture of two subjects

separation methods were applied to a set of single-channel mixtures generated by simulating the addition of the recorded signals. Fig. 4 illustrates the simulated process for separating a snoring mixture of two subjects. All the scenarios were carried out using MATLAB.

B. MVMD Parameters Optimization

MVMD parameters, particularly K and α, have a significant impact on the separation. They were assigned using Algorithms 2 and 3, respectively. For example, separating the partially out-of-phase snoring mix of two children at $K = 10$ produced the same signals as at $K = 5$, as shown in Fig. 5. However, it took 6 minutes and 55 seconds to produce 10 modes while it took 3 minutes and 52 seconds to decompose the signal into 5 modes. Similarly, the MVMD decomposition was conducted on multiple values of α to determine the most appropriate value. Fig. 6 shows the power spectrum for the children's snoring signals and the separated signals with $\alpha = 2000, 5000$, and 7000. While the spectrum of $\alpha = 2000$ missed the frequency components required to generate the snoring sources, penalty factors of 5000 and 7000 were able to decompose the snoring mixture and recover some of those frequencies. Whereas, the power spectrum of $\alpha = 5000$ demonstrates that it is more compatible with the spectrum of the snoring sources, almost reaching the same power levels. The MVMD parameters used are listed in table I.

C. Source Separation of Snoring Signals

Fig. 7 represents a case study of the experiments carried out. Fig. 7a displays the partially phase-shifted snoring combination of two children. MVMD decomposed this mixture into five IMFs after discovering optimal values of K and α. IMFs were then processed by PCA to extract the principal components for the reconstruction of the snoring sources. In addition, NMF converted the single-channel snoring mixture into a product of the activation matrix H and the basis elements W, used to retrieve the sources. MVMD and NMF separated signals in Figs. 7c, 7c', 7e and 7e', respectively, present the results after the decomposition and the reconstruction phases. These signals indicates the presence of uncorrelated segments in each source. The enhancement approach was then employed to adjust the separated signals to meet the snoring sources. Apparently, the red signals in Figs. 7d, 7d', 7f and 7f' show the improvements that occurred for the enhanced separated signals relevant to each snoring subject.

VI. Results and Discussion

The validation of the algorithms is tested with the assessment of various scenarios. In certain scenarios, both MVMD and NMF exhibited poor performance primarily due to the presence of a low-intensity snoring signal combined with other

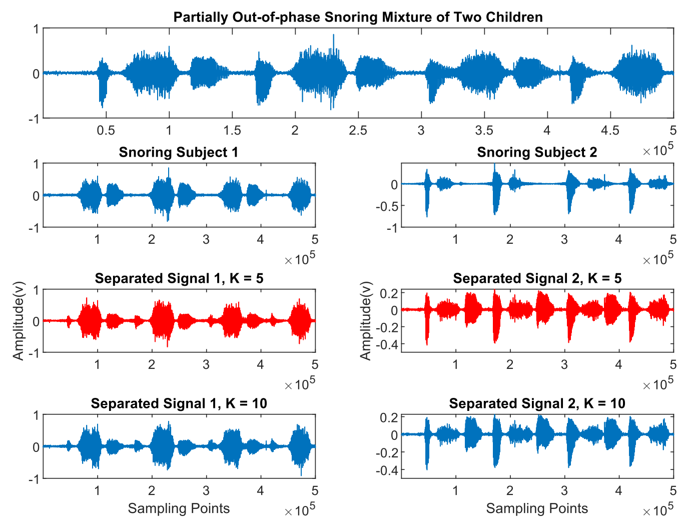

Fig. 5. MVMD Separated signals for the partially out-of-phase snoring mixture of two children for $K = 5$ and $K = 10$

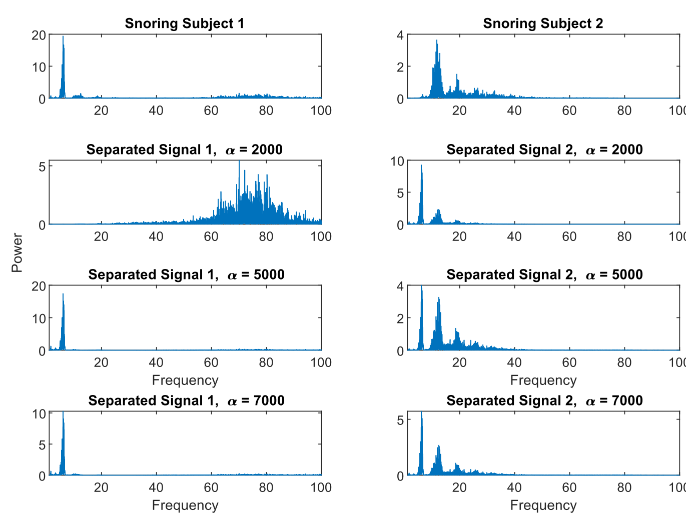

Fig. 6. Power spectrum for each snoring signal and its separated signal for $\alpha = 2000, 5000, 7000$

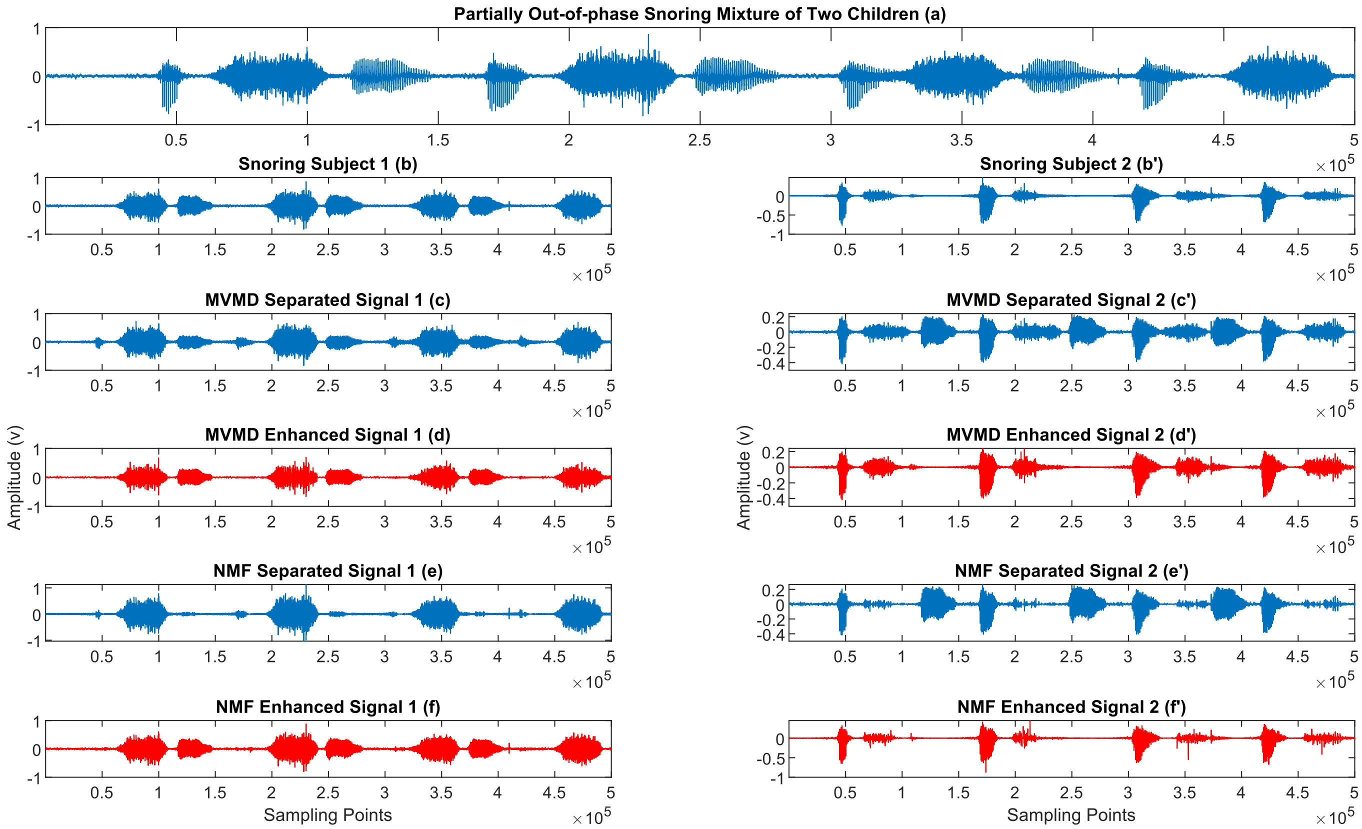

Fig. 7. Snoring separation example of a mixture of the two children "Child 1 - Child 2". (a) single-channel partially out-of-phase mixture, (b) snoring source signal of subject 1, (b') snoring source signal of subject 2, (c) and (c') separated signals after MVMD decomposition, (d) and (d') MVMD separated signals after enhancement, (e) and (e') separated signals after NMF decomposition, (f) and (f') NMF separated signals after enhancement

high-intensity signal. As table I shows, this discrepancy is more apparent in the mixtures where children's snoring is considered. Therefore, the enhancement approach demonstrates its effectiveness in such cases. The correlation coefficient for each separated and enhanced signal in each mixing scenario is presented in table I. It is worth mentioning that the correlation corresponding to each subject is listed in the same order as their mixture.

The integration of a man's snoring with a child's snoring in M4 dissipated the snoring structure of the child. The snoring signals of the child, after MVMD decomposition, showed correlation results between 51% and 59%, indicating that the signal is highly incompatible with the original source. However, these separated signals were adjusted after applying the enhancement algorithm, yielding a correlation ranging from 85.9% to 87.7%. As well, mixtures M6 and M7 experienced the weakest extraction of the snoring signals. In these scenarios, the enhancement approach often led to an increase in the similarity between the separated signal and its source by 30-40%. Fig.7 illustrates the process of separating the partially out-of-phase snoring mixture in M6. It exhibits the improvement in the enhanced signals, where the correlation of the MVMD separated signals increased from [75%, 64.58%] to [83.6%, 88.76%] and that of the NMF separated signals boosted to [98.41%, 94.14%]. Similar improvements were also observed in the NMF decomposition of the in-phase and out-of-phase mixtures of M6. Furthermore, the results in M7 demonstrate the efficiency of the enhancing approach by adding at least 10% compatibility with the source snoring signals. For example, the MVMD decomposition of the out-of-phase mixture resulted in a signal corresponding to 46.66% of child 3 snoring. This separated signal was refined to obtain 85.44% similarity. Special mixtures were also considered in testing the developed algorithms. M8 and M9 were designed to combine three and four sources, respectively. The evaluation results of these scenarios state the impact of the enhancement approach, which raises the correlation of some separated signals from 30% and 40% to an extremely high correlation approaching 90%.

Depending on what is mentioned in section II.A., our paper focuses on the issue of complex snoring separation, in contrast to Perez *et al.* [18] where the snoring mixture was not considered. Even though the deep learning utilized in [19] was successful in separating a single-channel snoring mixture, our work takes a different perspective. Unlike deep learning, we present practical methods based on the fundamentals of digital signal processing to demonstrate the applicability of these methods in separating snoring mixtures. Additionally, this research paper addresses a vulnerability highlighted in [17]. It introduces a novel approach for effectively separating

TABLE I
TABLE SHOWING THE DECOMPOSITION AND ENHANCEMENT RESULTS OF ALL THE PERFORMED MIXING SCENARIOS

Mixing Scenarios				MVMD				NMF	
				MVMD Parameters		Correlation (%)		Correlation (%)	
	M#	Mixtures	Cases	K	Alpha	Before enhancement	After enhancement	Before enhancement	After enhancement
Two Sources Mixtures	M1	Man - Man	In	10	5000	[90.18, 78.12]	**[90.46, 80.33]**	[92.52, 83.22]	**[92.88, 83.40]**
			Out	10	5000	[92.39, 77.37]	**[94.53, 88.99]**	[97.99, 93.18]	**[98.73, 97.33]**
			Partially	10	5000	[90.13, 78.60]	**[91.57, 81.92]**	[93.65, 86.16]	**[94.26, 86.83]**
	M2	Woman - Woman	In	5	2000	[83, 98.04]	**[83.88, 98.14]**	[88.27, 98.40]	**[88.49, 98.53]**
			Out	5	2000	[83.82, 97.46]	**[86.44, 99.34]**	[98.16, 99.63]	**[98.28, 99.89]**
			Partially	5	2000	[82.85, 97.87]	**[84.69, 98.57]**	[93.78, 99.08]	**[93.76, 99.21]**
	M3	Man - Woman	In	10	2000	[91.10, 82.97]	**[91.54, 83.35]**	[94.20, 90.25]	**[94.53, 90.38]**
			Out	5	2000	[92.65, 81.05]	**[96.37, 87.41]**	[98.97, 97.71]	**[99.54, 98.10]**
			Partially	5	2000	[91.24, 82.16]	**[93.69, 83.40]**	[96.35, 93.91]	**[96.73, 93.96]**
	M4	Man - Child	In	5	2000	[80.85, 59.82]	**[91.66, 87.77]**	[90.04, 80.19]	**[93.49, 85.57]**
			Out	5	2000	[83.34, 53.31]	**[95.22, 87.17]**	[94.01, 84.57]	**[98.35, 90.69]**
			Partially	10	2000	[85.43, 51.03]	**[93.56, 85.94]**	[91.71, 77.36]	**[94.63, 85.48]**
	M5	Woman - Child	In	5	2000	[85.46, 91.29]	**[86.15, 94.04]**	[92.08, 93.78]	**[92.70, 93.64]**
			Out	5	2000	[85.90, 90.57]	**[87.57, 97.48]**	[98.51, 98.70]	**[98.48, 99.35]**
			Partially	5	2000	[85.97, 90.70]	**[86.67, 94.98]**	[95.79, 96.36]	**[95.75, 95.81]**
	M6	Child 1 - Child 2	In	5	5000	[89.17, 70.21]	**[89.46, 83.20]**	[69.14, 44.34]	**[96.24, 83.10]**
			Out	5	5000	[73.11, 40.31]	**[80.07, 80.54]**	[77, 53.33]	**[87.75, 93.95]**
			Partially	5	5000	[75.02, 64.58]	**[83.60, 88.76]**	[75.56, 58.78]	**[98.41, 94.14]**
	M7	Child 2 - Child 3	In	10	2000	[74.69, 63.56]	**[87.21, 80.21]**	[81.22, 67.28]	**[84.79, 82.54]**
			Out	5	2000	[85.46, 46.66]	**[95.67, 85.44]**	[81.88, 63.26]	**[85.94, 89.09]**
			Partially	10	2000	[77.52, 58.21]	**[94.16, 90.30]**	[83.27, 65.30]	**[86.87, 84.43]**
Special Mixtures	M8	Man - Woman - Child	In	10	2000	[85.34, 79.34, 51.11]	**[93.74, 80.91, 87.61]**	[88.89, 81.37, 75.39]	**[90.42, 81.99, 83.69]**
			Out	5	2000	[75.49, 79.31, 55.91]	**[91.55, 86.52, 90.44]**	[88.39, 96.78, 83.82]	**[92.6, 97.31, 84.72]**
	M9	Man - Woman - Child - Child	In - Out	5	2000	[42.57, 78.79, 60.72, 30.17]	**[92.03, 94.99, 88.13, 84.21]**	[44.24, 69.50, 69.26, 47.27]	**[82.38, 83.73, 95.59, 88.19]**

complex snoring signals without the requirement of night supervision to monitor the movements of the sleeping subjects. Typically, night supervision is essential in the context of BDSS to analyze sound propagation delays involved in the recording scenario for accurately estimating the snoring of each subject. Unlike traditional methods that require supervision for snoring separation, the proposed method in this paper utilizes MVMD and NMF to achieve reliable snoring separation results without the need for such supervision. Hence, BDSS performance depends on the accuracy of the time-delay estimation. On the other hand, MVMD's effectiveness relies on the choice of parameters, which were determined based on the optimization algorithms represented in Algorithms 2 and 3. BDSS is specifically designed for blind source separation tasks involving time-delayed sources, while MVMD and NMF are more general mode decomposition methods. MVMD is particularly well-suited for signals with varying frequency components (i.e.,like snoring signals), enabling it to effectively capture mode variations over time. Moreover, NMF is employed to separate mixed audio signals into their individual source components by decomposing the audio spectrogram into spectral components. This approach is computationally efficient and easier to implement compared to some BDSS methods that involve time-domain processing. By operating in the frequency domain, both MVMD and NMF can handle signals with time-varying spectral characteristics and provide a more accurate decomposition compared to time-domain methods that assume stationarity. Although MVMD and NMF have been widely developed for audio source separation in [24]–[29], it is worth noting that these techniques are being applied to snoring signals for the first time. Furthermore, our implemented approaches exhibit some significant distinctions. First, our research marks the first instance of introducing these comprehensive scenarios of snoring mixtures, including in-phase, out-of-phase, and partially out-of-phase. Additionally, our enhancement approach precisely fine-tunes the separated signals to closely resemble the snoring sources, resulting in a remarkably high similarity rate that frequently surpasses 85%.

VII. CONCLUSION

This comprehensive study introduced the problem of separating complex audio signals originating from the home sleep environment while examining sleep audio recordings realized when both sleeping partners snore. In this paper, MVMD and NMF algorithms were applied to extract individual snoring signals presented in the single-channel snoring mixtures arising from the following scenarios: in-phase, out-of-phase, and partially out-of-phase. In addition, the enhancement algorithm has contributed to the fine-tuning of the separated signals generated after MVMD and NMF decomposition. The designed approaches were assessed by simulating real-time snoring recordings and calculating the correlation coefficient between each source and its separated signal. Experimental results of the underlying methods conducted in this study attained a correlation in the range of 80% to 99%, demonstrating the effectiveness and reliability of the developed methodologies in separating snoring mixtures. Individual snoring extraction can be discussed for an early diagnosis as well as replacing costly and time-consuming clinical procedures such as PSG. It allows doctors to monitor people's snoring while maintaining a normal sleep environment and not isolating subjects from their sleeping partners.

This study addressed the issue of decomposing complex snoring signals, whereas snoring is a primordial symptom of sleep disorders that needs to be monitored and diagnosed. As the next phase of our work, we are dedicated to obtaining real-time snoring recordings from an expanded dataset, enabling us to perform systematic testing and analysis. Future work will also look into screening of snoring individuals to aid in a comfortable and everlasting diagnostic analysis of sleep-related breathing disorders. In addition, we look forward to design this system as a portable device for home-based monitoring and diagnosis of sleep disorders.

References

[1] J. Fiz and R. Jane, "Snoring analysis. A complex question," *J Sleep Disord Treat Care.*, vol. 1, no. 1, pp. 1–3, 2012.
[2] I. Shaikh and S. Khosla, "Snoring," Sleep Education, November 2020. [Online]. Available: https://sleepeducation.org/sleep-disorders/snoring/. [Accessed 5 October 2022].
[3] D. G. Clark, "Snoring disorders and sleep disordered breathing for dentists," Online Dental Programs, July 2022. [Online]. Available: https://ostrowonline.usc.edu/snoring-disorders-for-dentists/. [Accessed 5 October 2022].
[4] K. H. Levin and P. Chauvel, "Polysomnography," in *Clinical Neurophysiology: Basis and Technical Aspects: Handbook of Clinical Neurology Series,* Elsevier, Ed. Cambridge, MA, USA: 2019.
[5] J. V. Rundo and R. Downey III, "Polysomnography," *Handb Clin Neurol.*, vol. 160, pp. 381–392, 2019.
[6] N. Scalzitti et al, "Comparison of home sleep apnea testing versus laboratory polysomnography for the diagnosis of obstructive sleep apnea in children," *Int. J. Pediatr. Otorhinolaryngol.*, vol. 100, pp. 44–51, 2017.
[7] M. Shokrollahi et al, "Snoring sound classification from respiratory signal," in *Proc. IEEE-EMBC*, Orlando, FL, USA, 2016, pp. 3215—3218.
[8] K. Banno and M. H. Kryger, "Sleep apnea: clinical investigations in humans," *Sleep Med.*, vol. 8, no. 4, pp. 400—426, 2007.
[9] T. Young et al, "Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women," *Sleep.*, vol. 20, no. 9, pp. 705—706, 1997.
[10] C. Wang and J. Peng, "The methods of acoustical analysis of snoring for the diagnosis of OSAHS," *J Sleep Med Disord.*, pp. 1—7, 2017.
[11] K. Qian et al, "Automatic detection, segmentation and classification of snore related signals from overnight audio recording," *IET Signal Process.*, vol. 9, no. 1, pp. 21—29, 2015.
[12] K. Nishijima et al, "Snore activity detection using smartphone sensors," in *Proc. IEEE-ICCE*, Taiwan, 2015, pp. 128—129.
[13] V. R. Swarnkar et al, "Automatic picking of snore events from overnight breath sound recordings," in *Proc. IEEE-EMBC*, Jeju Island, Korea, 2017, pp. 2822—2825.
[14] N. ur Rehman and H. Aftab, "Multivariate variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 67, no. 23, pp. 6039—6052, 2019.
[15] K. Pearson, "Liii. On lines and planes of closest fit to systems of points in space," *Lond. Edinb. Dublin philos. mag. j. sci.*, vol. 2, no. 11, pp. 559—572, 1901.
[16] D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Adv. Neural Inf. Process. Syst.*, vol. 13, 2000.
[17] V. Nigam and R. Priemer, "A snore extraction method from mixed sound for a mobile snore recorder," *J. Med. Syst.*, vol. 30, no. 2, pp. 91—99, 2006.
[18] J. M. Perez-Macias et al, "Detection of snores using source separation on an Emfit signal," *IEEE J Biomed Health Inform.*, vol. 22, no. 4, pp. 1157—1167, 2017.
[19] H. E. Romero et al, "Snorer diarisation based on deep neural network embeddings," in *Proc. IEEE-ICASSP*, 2020, pp. 876—880.
[20] Y. Guo et al, "Single-mixture source separation using dimensionality reduction of ensemble empirical mode decomposition and independent component analysis," *Circuits, Syst. Signal Process.*, vol. 31, no. 6, pp. 2047—2060, 2012.
[21] W. Xu and X. Yan, "Application of single channel blind separation algorithm based on EEMD-PCA-robustICA in bearing fault diagnosis," *Int. J. Commun. Syst.*, vol. 10, no. 8, pp. 138—147, 2017.
[22] P. Dey et al, "Single channel blind source separation based on variational mode decomposition and PCA," in *Proc. IEEE-INDICON*, New Delhi, India, 2015, pp.1—5.
[23] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531—544, 2013.
[24] R. Gavas et al, "Multivariate variational mode decomposition based approach for blink removal from EEG signal," in *Proc. IEEE-PerCom Workshops*, Austin, TX, USA, 2020, pp. 1—6.
[25] R. Ranny et al, "Separation of Overlapping Sound using Nonnegative Matrix Factorization," in *Proc. IEEE ISRITI*, Yogyakarta, Indonesia, 2019, pp. 424—429.
[26] P. Smaragdis, "Convolutive speech bases and their application to supervised speech separation," *IEEE Trans Audio Speech Lang Process.*, vol. 15, no. 1, pp. 1—12, 2006.
[27] N. Mohammadiha et al, "Supervised and unsupervised speech enhancement using nonnegative matrix factorization," *IEEE Trans Audio Speech Lang Process.*, vol. 21, no. 10, pp. 2140—2151, 2013.
[28] P. Smaragdis et al, "Static and dynamic source separation using nonnegative factorizations: A unified view," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 66—75, 2014.
[29] M. Kim and P. Smaragdis, "Single channel source separation using smooth nonnegative matrix factorization with Markov random fields," in *Proc. IEEE-MLSP*, Southampton, UK, 2013, pp. 1—6.
[30] S. Boyd et al, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1—122, 2011.
[31] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111—126, 1994.
[32] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
[33] N. Gillis, "The why and how of nonnegative matrix factorization," *Connections*, vol. 12, no. 2, 2014.
[34] T. Shakhnovsky and A. Lischke, "Non-negative Matrix Factorization for Audio Source Separation," 2019.
[35] C. Fevotte et al, "Single-channel audio source separation with NMF: divergences, constraints and algorithms," *Audio Source Separation* , pp. 1–24, 2018.
[36] J. Park et al, "Separation of instrument sounds using non-negative matrix factorization with spectral envelope constraints," *arXiv preprint arXiv:1801.04081*, 2018.
[37] G. Shah et al, "On the blind recovery of cardiac and respiratory sounds," *IEEE J Biomed Health Inform.*, vol. 19, no. 1, pp. 151—157, 2014.
[38] C. J. Lin, "On the convergence of multiplicative update algorithms for nonnegative matrix factorization," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1589–1596, 2007.
[39] M. Berry et al, "Algorithms and applications for approximate nonnegative matrix factorization," *Comput. Stat. Data Anal.*, vol. 52, no. 1, pp. 155–173, 2007.
[40] J. Kim et al, "Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework," *J Glob Optim.*, vol. 58, no. 2, pp. 285–319, 2014.
[41] T. Liang et al, "Application of parameter optimized variational mode decomposition method in fault feature extraction of rolling bearing," *Entropy*, vol. 23, no. 5, p. 520, 2021.
[42] W. Yang et al, "Superiorities of variational mode decomposition over empirical mode decomposition particularly in time–frequency feature extraction and wind turbine condition monitoring," *IET Renew. Power Gener.*, vol. 11, no. 4, pp. 443–452, 2017.