# **Open Access**

# An MMSE graph spectral magnitude estimator for speech signals residing on an undirected multiple graph



Tingting Wang<sup>1\*</sup>, Haiyan Guo<sup>1</sup>, Zirui Ge<sup>1</sup>, Qiquan Zhang<sup>2</sup> and Zhen Yang<sup>1\*</sup>

# Abstract

The paper uses the *K*-graphs learning method to construct weighted, connected, undirected multiple graphs, aiming to reveal intrinsic relationships of speech samples in the inter-frame and intra-frame. To benefit from the learned multiple graphs' property and enhance interpretability, we study the spectral property of speech samples in the joint vertex-frequency domain by using the new graph weight matrix. Moreover, we propose the representation of minimum mean-square error (MMSE) graph spectral magnitude estimator for speech signals residing on undirected multiple graphs. We use the MMSE graph spectral magnitude estimator to improve speech enhancement performance. The numerical simulation results show that the proposed method outperforms the existing methods in graph signal processing (GSP) and the baseline methods for speech enhancement in discrete signal processing (DSP) in terms of PESQ, LLR, output SNR, and STOI results. These results also demonstrate the validity of the learned multiple graphs.

**Keywords** Graph signal processing, *K*-graphs learning, Graph representation, MMSE, Speech enhancement

# **1** Introduction

Graph signal processing (GSP) [1, 2] explores the relationships among discrete signals residing on vertexes via graph models [3]. It has been developing a set of theories based on the traditional discrete signal processing to investigate, analyze, and process the data defined over arbitrary topologies [4, 5]. The scope of research has changed from fundamental GSP concepts [6–8] to practical applications including graph signals denoising and restoration [9–11], learning graphs from observed data [12–14], image processing [15, 16], and graph clustering [17, 18].

\*Correspondence: Tingting Wang 2018010215@njupt.edu.cn Zhen Yang yangz@njupt.edu.cn

<sup>1</sup> School of Communication and Information Engineering, Nanjing

University of Posts and Telecommunications, Nanjing, China

GSP is efficient in characterizing time series by using graph theories [19]. To be specific, in [20], dynamic visibility graphs (DVG) were constructed to describe time series, which studied the DVG dependence of different time series. In [21], the recurrence matrix of the time series was defined as the adjacency matrix of an associated complex graph to link different points in the case where the evolution of the considered states is similar. In [22, 23], a visibility graph based on the Gaussian kernel function was defined for electroencephalogram (EEG) signals, which provided a topology to capture sudden fluctuations happening in EEG during seizure activity.

Traditional discrete signal processing (DSP)-based speech enhancement algorithms use short-time spectrum estimation to suppress noises. Specifically, in [24], the authors proposed the traditional Wiener filtering for speech enhancement by leveraging the frequency-domain characteristics of noises and speech signals. In [25], the authors focused on the major importance of the shorttime spectral amplitude (STSA) of the speech signal and proposed a minimum mean-square error (MMSE) STSA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

<sup>&</sup>lt;sup>2</sup> Department of Electrical and Computer Engineering, National University of Singapore, Singapore

estimator for speech enhancement by modeling speech and noise spectral components as statistically independent Gaussian random variables. In [26], the authors proposed the optimal modified minimum mean-square error log-spectral (OMLSA) for robust speech enhancement by minimizing the mean-square error of the log-spectra as a weighted geometric mean of the hypothetical gains associated with the speech presence uncertainty. In addition, in [27], the authors proposed a statistical speech enhancement model using acoustic environment classification supported by a Gaussian mixture model. In [28], the authors proposed a joint-constrained dictionary learning method to solve the "cross projection" problem of signals in the joint dictionary for single-channel speech enhancement.

Unlike the traditional DSP-based speech enhancement methods, GSP-based speech enhancement methods first establish graphs for speech signals before performing enhancement. By establishing the graph adjacency matrix with different edges and weights, speech signals can then be flexibly mapped into different graph frequency domains with different graph Fourier bases. It is worth noting that finite (periodic) time series have been constructed as signals indexed by a directed cycle graph [1, 2, 6]. Speech signals are special time series. The current graph topology of finite time series is directly applied for unstructured speech signals, which explores sampled speech signals' time shifts and succession and fails to capture the potential relationship among speech samples.

Our previous work [29, 30] has made many processes for inferring a suitable graph representation of speech signals. To be specific, in [29], we first established a single undirected graph topology for unstructured speech signals, which successfully mapped time-domain speech signals into the vertex domain and viewed them as speech graph signals. In [30], we proposed a single digraph by using algebraic signal processing (ASP) [31] theories and then built graph wiener filters in the graph Fourier domain for speech enhancement. However, the designed static graph topology for speech signals in our previous work [29, 30] cannot capture the potential relationships between different speech frames. In [32], we proposed to learn a directed multilayer graph model for speech signals by using graph learning, which reveals both the intrinsic relationships of inter-frames and those among speech samples within a frame. But it aimed to learn a complex and large volume graph model for the total speech signals, which does not reveal the dynamic change characteristics in speech signals.

Against this backdrop, in this paper, we propose a *K*-graph learning method to learn multiple undirected graphs for framed noisy speech graph signals to better

match the dynamic nature of speech. Specifically, the framed noisy speech graph signals are partitioned into a set of clusters. For each cluster, an undirected graph is learned to reveal the potential relationships among noisy speech frames in the cluster. In this way, multiple graphs of a small size other than a large-volume graph are learned, which reveal the inter-frame relationships of the total speech signals in a more dynamic way. Additionally, as the size of each cluster is much smaller than that of the whole speech signals, the *K*-graphs learning method leads to multiple graphs of small volumes with a much lower learning complexity.

On the basis of the learn multiple undirected graphs, we propose the gain function representation for the MMSE graph spectral magnitude estimator by extending the classical MMSE-STSA to enhance noisy speech signals. The contributions of the paper are summarized as follows.

i) We propose the novel undirected multiple graphs by using the *K*-graphs learning method, which reveals potential relationships among noisy speech frames in real-time. On this basis, we construct a joint graph weight matrix and define the related graph Fourier basis.

ii) Based on the constructed graph Fourier basis, we investigate the gain function representation of MMSE graph magnitude spectral estimator for speech graph signals (SGSs). We propose an MMSE graph spectral magnitude estimator by extending the classical MMSE-STSA method in DSP to perform speech enhancement.

iii) Our numerical results show that the proposed method outperforms the benchmarks in terms of the output signal-to-noise ratio (SNR), perceptual evaluation of speech quality (PESQ), log-likelihood ratio (LLR), and short-time objective intelligibility (STOI). The numerical results also demonstrate the validity of the learned multiple graphs for speech signals. We use the classical MMSE-STSA [25] and OMLSA [26] methods in DSP and existing graph Wiener filtering methods [33] as benchmarks.

The remainder of the paper is organized as follows. Section 2 introduces the related work of GSP. Section 3 investigates the *K*-graphs learning method for speech signals. The details of the MMSE graph spectral magnitude estimator method are given in Section 4. Section 5 provides our experimental results. Section 6 concludes the paper.

Notation: The superscript  $^T$  represents the transpose. trace( $\cdot$ ),  $\|\cdot\|_F$  and  $\|\cdot\|_1$  represent the trace, F-norm, 1-norm, respectively, and 1 denotes the all-ones matrix. × represents the Cartesian product. I is the identity matrix.  $E(\cdot)$  and  $p(\cdot)$  represent the expectation operator and the probability density function, respectively.

# 2 Related work

## 2.1 Basics of GSP

Let  $G = (\mathcal{V}, \mathcal{E}, \mathbf{A})$  represent a weighted graph, where  $\mathcal{V}$  denotes the collection of vertices  $v_1, \dots, v_N$ , the set of edges  $\mathcal{E}$  satisfies  $(m, n) \in \mathcal{E}$  if and only if the vertex  $v_m$  is connected to the vertex  $v_n$  [34], and  $\mathbf{A}$  denotes the graph weight matrix. The element in  $\mathbf{A}$  represents the weight of the edge between vertices  $v_m$  and  $v_n$ , which intuitively and numerically shows the appropriate dependence or similarity between signals on vertices [35, 36]. This paper focuses on a connected undirected graph with finite vertices. We can define the variation operator by the combinatorial graph Laplacian matrix  $\mathbf{L} = \mathbf{D} - \mathbf{A}$ , where  $\mathbf{D}$  is a diagonal matrix with elements  $d_{mm} = \sum_{n=1}^{N} A_{mn}, (m = 1, \dots N)$  [37].

Following [38, 39], we have the smoothness of graph signals X given by

$$(\mathbf{XL})(\nu_m) = \sum_{\{\nu_m\} \in \mathcal{V}} A_{m,n}(x(m) - x(n)).$$
(1)

A small smoothness value means that signals indexed by adjacent vertices have similar values, that is, X is smooth [40]. From (1), we have

$$\mathbf{X}^{T}\mathbf{L}\mathbf{X} = \sum_{\{\nu_{m}\}\in\mathcal{V}} x(m) \sum_{\{\nu_{n}\}\in\mathcal{V}} A_{m,n}(x(m) - x(n))$$
  
$$= \sum_{\{\nu_{m}\}\in\mathcal{V}} \sum_{\{\nu_{n}\}\in\mathcal{V}} A_{m,n}(x^{2}(m) - x(m)x(n))$$
  
$$= \frac{1}{2} \sum_{\{\nu_{m}\}\in\mathcal{V}} \sum_{\{\nu_{n}\}\in\mathcal{V}} A_{m,n}(x(m) - x(n))^{2}.$$
 (2)

#### 2.2 Speech signals on graphs

After the framing operation, speech signals can be expressed as a matrix  $\mathbf{s} \in \mathbb{R}^{M \times N_s}$  where each row represents a frame speech signals, M represents the number of whole noisy speech frames, and  $N_s$  denotes the length of a speech frame. Let us view each discretized speech sample as a vertex. Assuming that the relationship between the adjacent samples is symmetrical, speech samples can be constructed as speech graph signals (SGSs) residing on a connected undirected graph. The one-to-one mapping between the  $i_{th}$  noisy speech sample  $f_i$  in a frame and the signal value of the  $i_{th}$  vertex  $v_i$  is given by

$$\mathbf{s}_G: \mathbb{R} \to \mathcal{V}_s, f_i \to \nu_i, \tag{3}$$

where  $V_s$  is the set of vertices cardinality  $|V_s| = N_s$ , in this way, noisy speech signals in each frame are mapped into the graph domain.

#### 2.3 The joint graph Fourier transform

Let we denote the time-vertex graph signal as  $\mathbf{X}_E = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T] \in \mathbb{R}^{N \times T}$ , where  $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T$  represent graph signals sampled at 1, 2, ... T successive regular intervals with length N [41]. To investigate the spectral properties of  $\mathbf{X}_E$ , following [42, 43], the joint time-vertex graph Fourier transform (JFT) is defined as

$$\mathbf{\hat{X}}_{E} = JFT\{\mathbf{X}_{E}\} = \Psi_{G}\mathbf{X}_{E}\Psi_{T},$$
(4)

where  $\Psi_T \in \mathbb{R}^{T \times T}$  is constructed as the normalized discrete Fourier transform matrix, with  $\Psi_T(t,k) = e^{-j\frac{2\pi(k-1)}{T}}/\sqrt{T}$  and  $\Psi_G \in \mathbb{R}^{N \times N}$  is the eigenvector matrix obtained by the eigendecomposition of  $X_E$ 's graph weight matrix. More specifically,  $\Psi_T$  is applied to analyze the time-frequency oscillations of  $X_E$  along the time domain,  $\Psi_G$  allows us to obtain the graph-frequency characters of  $X_E$  along the graph edges. Moreover, the corresponding inverse IJFT is defined as

$$IJFT\{\widehat{\mathbf{X}}_{E}\} = \Psi_{G}^{-1}\widehat{\mathbf{X}}_{E}\Psi_{T}^{T}.$$
(5)

The definitions of JFT and IJFT allow us to take into account the variation of the graph and the temporal aspects of time-vertex graph signals.

# 3 The K-graphs learning method for speech graph signals

This section uses the *K*-graphs learning method to infer a multiple-graphs model for SGSs without prior graph topologies. It is noted that the inter-frame relationship is not considered in traditional speech enhancement systems. Differing from noises, speech samples have a strong correlation between and within frames. In this paper, considering that when we learn a global graph for speech signals, it would be complex. Inspired by *K*-means and *K*-graphs learning [44], we partition the SGSs into clusters and use the *K*-graphs learning method to capture the potential properties of speech samples both the inter-frame and intra-frame.

Noisy speech signals are mapped into the graph domain by using Eq. (3) and are constructed as SGSs. We employ  $M_s$  speech frames in a cluster to learn a graph for capturing the relationships between  $M_s$  speech frames. The SGSs in the *k*th cluster,  $\mathbf{s}_G^k$  $(1 \le k \le K, K = M/M_s)$ , resides on the *k*th undirected multiple graph  $G_k = (\mathcal{V}_k, \mathcal{L}_k)$ , where  $\mathcal{V}_k$  indicates the vertex set and  $\mathcal{L}_k$  is the graph Laplacian matrix of  $G_k$ , and *M* is the total number of speech frames. Let us now investigate the graph weighted matrix  $\mathcal{L}_k$  of  $G_k$ , for the sake of revealing the intrinsic relationships among speech frames in real-time. Following the *K*-graphs learning framework in [44], we formulate the multiple graphs learning problem of noisy speech graph signals as

strong causality between adjacent speech samples within a frame. Upon denoting the graph weight matrix of  $O_i$  by  $\mathbf{W}_i \in \mathbb{R}^{N_s \times N_s}$ , we set  $w_i(m, n) = 1$  if there exists a strong

$$\min_{\substack{\boldsymbol{\mathcal{L}}_{1},\cdots,\boldsymbol{\mathcal{L}}_{K}\\\mathbf{s}_{G}^{1},\cdots,\mathbf{s}_{G}^{K} \in \mathbb{R}^{M_{S} \times N_{S}}}} \sum_{k=1}^{K} \sum_{\substack{\boldsymbol{s}_{G}^{k} \in \mathbf{s}_{G}^{K}\\\mathbf{s}_{G}^{k} \in \mathbb{R}^{M_{S} \times N_{S}}}} \operatorname{trace}(\boldsymbol{s}_{G}^{k} \boldsymbol{\mathcal{L}}_{k}(\boldsymbol{s}_{G}^{k})^{\mathrm{T}}) + \sum_{k=1}^{K} \beta \|\mathbf{1} + \boldsymbol{\mathcal{L}}_{k}\|_{F}^{2} + \sum_{k=1}^{K} \alpha \|\boldsymbol{\mathcal{L}}_{k}\|_{1},$$

$$s.t. \quad \operatorname{diag}(\boldsymbol{\mathcal{L}}_{k}) \geq 0,$$

$$\operatorname{trace}(\boldsymbol{\mathcal{L}}_{k}) = M_{s}.$$

$$\boldsymbol{\mathcal{L}}_{k} = (\boldsymbol{\mathcal{L}}_{k})^{T}, 1 \leq k \leq K,$$

$$(6)$$

where  $\mathbf{s}_G^k \mathcal{L}_k (\mathbf{s}_G^k)^T$  describes the smoothness of SGSs supported on  $G_k$ , the Frobenius norm of  $\mathbf{1} + \mathcal{L}_k$  is used to control the distribution of the edge weights and the sparsity, and the third term is the regularization function and ensures its positive value [45]. The first and second constraints are used to ensure the symmetricity and nonnegativity of  $\mathcal{L}_k$ . The third constraint is added to prevent trivial solutions and control the volume of the corresponding multiple graphs  $G_k$ .  $M_s$  controls the volume of the intra-graph.  $\alpha$  and  $\beta$  are non-negative regularization parameters. As the objective function (6) is convex, we can solve it with the CVX toolbox [46] in the experimental section. causality between the vertex  $v_m$  and its adjacent vertex  $v_n$ , and otherwise  $w_i(m, n) = 0$ . That is,

$$\mathbf{W}_{i} = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & 1 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & \cdots & 1 & 0 & 1 \\ 0 & \cdots & \cdots & \cdots & 1 & 0 \end{bmatrix}.$$
 (7)

Then, we have  $O_i = (\mathcal{V}_i, \mathbf{W}_i)$ , where  $\mathcal{V}_i$  represents the vertex set with cardinality  $|\mathcal{V}_i| = N_s$ . Hence,  $M_s$  speech frames **f** can be constructed as the speech graph signal  $\mathbf{s}_G$  indexed by the multiple graphs  $G_S = (\mathcal{V}, \mathcal{L}^*)$  as shown in Fig. 1.

$$p: \mathbf{f} \to \mathbf{s}_{\mathbf{G}} \in \mathbb{R}^{M_s \times N_s} \text{ indexed by } G_S = G_k \times O_i, 1 \le k \le K, 1 \le i \le M_s.$$
(8)

We infer an intra-graph topology to investigate the internal relationships between speech samples within a frame. We denote  $\mathbf{s}_g^i$  as the  $i_{th}$  row of  $\mathbf{s}_G^k$ , which is indexed by an intra-graph  $O_i$ . Here we focus on studying the

By applying the Cartesian product of the inter-graph Laplacian matrix  $\mathcal{L}_k$  and the intra-graph weight matrix  $\mathbf{W}_i, \mathcal{L}^*$  is constructed as



**Fig. 1** The visualization of the multiple graphs  $G_s$ 

$$\mathcal{L}^* = \mathcal{L}_k \times \mathbf{W}_i, 1 \le k \le K, 1 \le i \le M_s, \tag{9}$$

and the vertex set  $\mathcal{V}^*$  is given as

$$\mathcal{V}^* = \mathcal{V}_k \times \mathcal{V}_i. \tag{10}$$

# 4 The MMSE graph spectral magnitude estimator method

This section proposes a minimum mean-square error (MMSE) graph spectral magnitude estimator based on the learned multiple graphs above. Specifically, we first define the corresponding joint graph Fourier transform (JGFT) and the inverse JGFT (IJGFT). Then we investigate the representation of the MMSE graph spectral magnitude estimator in GSP by extending the classical MMSE short-time spectral amplitude (STSA) estimator.

### 4.1 The joint graph Fourier transform for SGSs

Differ from the joint graph Fourier transform definition in Section 2, by applying the singular value decomposition (SVD) on  $\mathbf{W}_i$ , we have

$$\mathbf{W}_i = \mathbf{F}_w \times \Lambda^w \times \mathbf{D}_w^T, \tag{11}$$

where  $\mathbf{F}_{w}$  and  $\mathbf{D}_{w}$  are the left unitary matrix and the right unitary matrix of  $\mathbf{W}_{i}$ , respectively, and  $\Lambda^{w} = \operatorname{diag}(\lambda_{1}^{w}, \lambda_{2}^{w}, ..., \lambda_{N_{s}}^{w})$  is the corresponding diagonal matrix and its element represents the graph frequency along the intra-graph edge. Similarly, we have

$$\mathcal{L}_k = \mathbf{F}_{\mathcal{L}} \times \Lambda^{\mathcal{L}} \times \mathbf{D}_{\mathcal{L}}^T, \tag{12}$$

where  $\mathbf{F}_{\mathcal{L}}$  and  $\mathbf{D}_{\mathcal{L}}$  respectively represent the left unitary matrix and the right unitary matrix of  $\mathcal{L}_k$  and the element  $\lambda_j^{\mathcal{L}} (1 \le j \le M_s)$  of  $\Lambda^{\mathcal{L}}$  represents the graph frequency along the inter-graph edges. The joint graph Fourier transform (JGFT) for  $\mathbf{s}_G^k$  can be defined as

$$\mathbf{S}_{\mathcal{F}}^{k} = \mathrm{JGFT}\{\mathbf{s}_{\mathrm{G}}^{k}\} = (\mathbf{F}_{\mathrm{w}})^{-1}\mathbf{s}_{\mathrm{G}}^{k}(\mathbf{F}_{\mathcal{L}})^{-1},$$
(13)

where  $\mathbf{S}_{\mathcal{F}}^k$  is the graph Fourier version of  $\mathbf{s}_G^k$ . Moreover, the inverse IJGFT of  $\mathbf{S}_{\mathcal{F}}^k$  is defined as

$$\mathbf{s}_{G}^{k} = \mathrm{IJGFT}\{\mathbf{S}_{\mathcal{F}}^{k}\} = \mathbf{F}_{\mathrm{w}}\mathbf{S}_{\mathcal{F}}^{k}\mathbf{F}_{\mathcal{L}}.$$
(14)

It should be noted that by using the defined JGFT, we can get the graph magnitude spectra of speech signals belonging to the real field by mapping speech signals into the graph frequency domain.

# 4.2 The MMSE graph spectral magnitude estimator for SGSs

Let us now investigate the MMSE graph spectral magnitude estimator. We denote  $\mathbf{s}_G^k = \mathbf{x}_G + \mathbf{n}_G$  where  $\mathbf{x}_G$  is clean SGSs,  $\mathbf{n}_G$  is the additive graph noise signal which is independent of  $\mathbf{x}_G$ . By performing the defined JGFT in (13), we have

$$\mathbf{S}_{\mathcal{F}}^{k} = \mathbf{X}_{\mathcal{F}} + \mathbf{N}_{\mathcal{F}} \in \mathbb{R}$$
(15)

where  $\mathbf{S}_{\mathcal{F}}^{k}$ ,  $\mathbf{X}_{\mathcal{F}}$  and  $\mathbf{N}_{\mathcal{F}}$  are the JGFT coefficient of  $\mathbf{s}_{\mathcal{F}}^{k}$ ,  $\mathbf{x}_{\mathcal{G}}$ and  $\mathbf{n}_{G}$  respectively. Upon denoting the  $i_{th}$  row of  $\mathbf{S}_{\mathcal{F}}^{k}$ ,  $\mathbf{X}_{\mathcal{F}}$ , and  $\mathbf{N}_{\mathcal{F}}$  by  $\mathbf{Y}^{i}$ ,  $\mathbf{X}^{i}$ , and  $\mathbf{N}^{i}$ , respectively, a graph speech sample on a vertex can be donated as  $\mathbf{Y}_{j}^{i} = \mathbf{X}_{j}^{i} + \mathbf{N}_{j}^{i}$  where  $i = 1, 2, 3, ..., K, j = 0, 1, 2, ..., N_{s} - 1$ . Let us denote  $R_{j}^{i} = |\mathbf{Y}_{j}^{i}|$  and  $\mathbf{Z}_{j}^{i} = |\mathbf{X}_{j}^{i}|$ . Based on the work

Let us denote  $R_j^i = |Y_j^i|$  and  $Z_j^i = |X_j^i|$ . Based on the work in [46], the MMSE estimator for the graph magnitude spectrum  $X_i^i$  can be obtained as

$$\widehat{Z}_{j}^{i} = E(Z_{j}^{i}|Y_{j}^{i}) = \frac{\int_{0}^{\infty} p(Y_{j}^{i}|Z_{j}^{i})p(Z_{j}^{i})(Z_{j}^{i})dZ_{j}^{i}}{\int_{0}^{\infty} p(Y_{j}^{i}|Z_{j}^{i})p(Z_{j}^{i})dZ_{j}^{i}}.$$
 (16)

In case of the Gaussian statistical model for spectral components, we have

$$p(Y_{j}^{i}|Z_{j}^{i}) = \frac{1}{\pi(\lambda_{n})_{j}^{i}} \exp\left\{-\frac{1}{(\lambda_{n})_{j}^{i}}|Y_{j}^{i} - Z_{j}^{i}|^{2}\right\}, \quad (17)$$

$$p\left(Z_{j}^{i}\right) = \frac{Z_{j}^{i}}{\pi\left(\lambda_{x}\right)_{j}^{i}} \exp\left\{-\frac{\left(Z_{j}^{i}\right)^{2}}{\left(\lambda_{x}\right)_{j}^{i}}\right\},$$
(18)

where  $(\lambda_x)_j^i = E\left[\left(X_j^i\right)^2\right]$  and  $(\lambda_n)_j^i = E\left[\left(N_j^i\right)^2\right]$  are the  $j_{th}$ SGS and the graph noise variance for  $X_j^i$  and  $N_j^i$  respectively. By substituting (17) and (18) into  $\int_0^\infty Z_j^i p(Y_j^i|Z_j^i) p(Z_j^i) dZ_j^i$  and  $\int_0^\infty p(Y_j^i|Z_j^i) p(Z_j^i) dZ_j^i$ , we have

$$\xi = \int_{0}^{\infty} Z_{j}^{i} p(Y_{j}^{i} | Z_{j}^{i}) p(Z_{j}^{i}) dZ_{j}^{i}$$

$$= \int_{0}^{\infty} \frac{1}{\pi (\lambda_{n})_{j}^{i}} \exp\left\{-\frac{1}{(\lambda_{n})_{j}^{i}} |Y_{j}^{i} - Z_{j}^{i}|^{2}\right\} \frac{(Z_{j}^{i})^{2}}{\pi (\lambda_{x})_{j}^{i}} \exp\left\{-\frac{\left(Z_{j}^{i}\right)^{2}}{(\lambda_{x})_{j}^{i}}\right\} dZ_{j}^{i},$$
(19)

and

$$\begin{aligned} \zeta &= \int_{0}^{\infty} p(Y_{j}^{i}|Z_{j}^{i})p(Z_{j}^{i})dZ_{j}^{i} \\ &= \int_{0}^{\infty} \frac{1}{\pi(\lambda_{n})_{j}^{i}} \exp\left\{-\frac{1}{(\lambda_{n})_{j}^{i}}|Y_{j}^{i} - Z_{j}^{i}|^{2}\right\} \frac{Z_{j}^{i}}{\pi(\lambda_{x})_{j}^{i}} \exp\left\{-\frac{\left(Z_{j}^{i}\right)^{2}}{\left(\lambda_{x}\right)_{j}^{i}}\right\} dZ_{j}^{i}. \end{aligned}$$

$$(20)$$

Then, we can rewrite (16) equivalently as

$$\hat{Z}_{j}^{i} = \frac{\xi}{\zeta} = \frac{\int_{0}^{\infty} \exp\left\{-\frac{1}{(\lambda_{n})_{j}^{i}}|Y_{j}^{i} - Z_{j}^{i}|^{2}\right\} \exp\left\{-\frac{(Z_{j}^{i})^{2}}{(\lambda_{n})_{j}^{i}}\right\} (Z_{j}^{i})^{2} dZ_{j}^{i}}{\int_{0}^{\infty} \exp\left\{-\frac{1}{(\lambda_{n})_{j}^{i}}|Y_{j}^{i} - Z_{j}^{i}|^{2}\right\} \exp\left\{-\frac{(Z_{j}^{i})^{2}}{(\lambda_{n})_{j}^{i}}\right\} Z_{j}^{i} dZ_{j}^{i}}.$$
(21)

And, we arrive at

$$\widehat{Z_{j}^{i}} = \frac{\xi}{\zeta} = \frac{\int_{0}^{\infty} \exp\left\{-\frac{(\lambda_{n})_{j}^{i}+(\lambda_{n})_{j}^{i}}{(\lambda_{n})_{j}^{i}(\lambda_{n})_{j}^{i}}\left(Z_{j}^{i}-\frac{Y_{j}^{i}(\lambda_{n})_{j}^{i}}{(\lambda_{n})_{j}^{i}+(\lambda_{n})_{j}^{i}}\right)^{2}\right\} \exp\left\{-\frac{(Z_{j}^{i})^{2}}{(\lambda_{n})_{j}^{i}}\right\} (Z_{j}^{i})^{2} dZ_{j}^{i}}{\int_{0}^{\infty} \exp\left\{-\frac{(\lambda_{n})_{j}^{i}+(\lambda_{n})_{j}^{i}}{(\lambda_{n})_{j}^{i}(\lambda_{n})_{j}^{i}}\left(Z_{j}^{i}-\frac{Y_{j}^{i}(\lambda_{n})_{j}^{i}}{(\lambda_{n})_{j}^{i}+(\lambda_{n})_{j}^{i}}\right)^{2}\right\} \exp\left\{-\frac{(Z_{j}^{i})^{2}}{(\lambda_{n})_{j}^{i}}\right\} (Z_{j}^{i}) dZ_{j}^{i}} (22)$$

Upon denoting  $\sigma^2 = \frac{(\lambda_n)_j^i + (\lambda_x)_j^i}{(\lambda_n)_j^i (\lambda_x)_j^i}$  and  $\tau = \frac{Y_j^i (\lambda_x)_j^i}{(\lambda_n)_j^i + (\lambda_x)_j^i}$ , we can rewrite (21) equivalently as

$$\widehat{Z}_{j}^{i} = \frac{\xi}{\zeta} = \frac{\int_{0}^{\infty} \exp\left\{-\sigma^{2} (Z_{j}^{i} - \tau)^{2}\right\} (Z_{j}^{i})^{2} dZ_{j}^{i}}{\int_{0}^{\infty} \exp\left\{-\sigma^{2} (Z_{j}^{i} - \tau)^{2}\right\} (Z_{j}^{i}) dZ_{j}^{i}}.$$
 (23)

Let us now analyze  $\xi$ . Upon denoting  $y = Z_j^i \sigma - \sigma \tau$ , we have

Let us introduce the Gauss error function  $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\eta^2} d\eta$ . Substituting erf(x) into (24) (see Appendix 1) gives

$$\xi = \frac{\tau}{2\sigma^2} \exp\{-\sigma^2 \tau^2\} + \frac{1}{\sigma^3} (\frac{\pi}{4} + \frac{\sqrt{\pi}}{2} \sigma^2 \tau^2) - \frac{1 + 2\sigma^2 \tau^2}{2\sigma^3} \frac{\sqrt{\pi}}{2} erf(\sigma\tau).$$
(25)

Similarly, we have

$$\zeta = \frac{1}{2\sigma^2} \exp\left\{-\sigma^2 \tau^2\right\} + \frac{\sqrt{\pi}}{2} \frac{\tau}{\sigma} + \frac{\sqrt{\pi}}{2} \frac{\tau}{\sigma} \operatorname{erf}(\sigma \tau).$$
(26)

By combining (21), (25) and (26) (see Appendix 2), we arrive at

$$\hat{Z}_{j}^{i} = \frac{\frac{\tau}{2\sigma^{2}} \exp\left\{-\sigma^{2}\tau^{2}\right\} + \frac{1}{\sigma^{3}} (\frac{\pi}{4} + \frac{\sqrt{\pi}}{2}\sigma^{2}\tau^{2}) - \frac{1+2\sigma^{2}\tau^{2}}{2\sigma^{3}} \frac{\sqrt{\pi}}{2} erf(\sigma\tau)}{\frac{1}{2\sigma^{2}} \exp\left\{-\sigma^{2}\tau^{2}\right\} + \frac{\sqrt{\pi}}{2} \frac{\tau}{\sigma} + \frac{\sqrt{\pi}}{2} \frac{\tau}{\sigma} erf(\sigma\tau)} R_{j}^{i}.$$
(27)

Let us introduce the notation of the prior signal-to-noise ratio  $\varphi_j^i = \frac{(\lambda_x)_j^i}{(\lambda_n)_j^i}$  and the posterior signal-to-noise ratio  $\delta_j^i = \frac{(\lambda_n)_j^i}{(\lambda_x)_j^i}$ . Due to  $\sigma^2 = \frac{(\lambda_n)_j^i + (\lambda_x)_j^i}{(\lambda_n)_j^i (\lambda_x)_j^i}$  and  $\tau = \frac{Y_j^i (\lambda_x)_j^i}{(\lambda_n)_j^i + (\lambda_x)_j^i}$ , we have

$$\gamma_{j}^{i} = \sqrt{\frac{(\lambda_{n})_{j}^{i} + (\lambda_{x})_{j}^{i}}{(\lambda_{n})_{j}^{i}(\lambda_{x})_{j}^{i}}} = \sqrt{\frac{1 + \varphi_{j}^{i}}{\varphi_{j}^{i}((\lambda_{n})_{j}^{i})^{2}}}, \\ \theta_{j}^{i} = \frac{Y_{j}^{i}(\lambda_{x})_{j}^{i}}{(\lambda_{n})_{j}^{i} + (\lambda_{x})_{j}^{i}} = \frac{Y_{j}^{i}(\varphi_{j}^{i})^{2}}{1 + (\varphi_{j}^{i})^{2}}.$$
(28)

By combining (23), (27) and (28), we can obtain the gain function of the MMSE graph spectral magnitude estimator given as

$$G_{\text{GMMSE}}(\gamma_{j}^{i},\vartheta_{j}^{i}) = \frac{\widehat{Z}_{j}^{i}}{R_{j}^{i}}$$

$$= \frac{\frac{\vartheta_{j}^{i}}{2(\gamma_{j}^{i})^{2}} \exp\left\{-(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}\right\} + \frac{1}{(\gamma_{j}^{i})^{3}}(\frac{\pi}{4} + \frac{\sqrt{\pi}}{2}(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}) - \frac{1+2(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}}{2\gamma_{i}^{3}}\frac{\sqrt{\pi}}{2} \operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{1}{2(\gamma_{j}^{i})^{2}} \exp\left\{-(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}\right\} + \frac{\sqrt{\pi}}{2}\frac{\vartheta_{j}^{i}}{\gamma_{j}^{i}} + \frac{\sqrt{\pi}}{2}\frac{\vartheta_{i}}{\gamma_{i}}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{1}{2(\gamma_{j}^{i})^{2}} + \frac{1}{(\gamma_{j}^{i})^{3}}(\frac{\pi}{4} + \frac{\sqrt{\pi}}{2}(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}) - \frac{1+2(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}}{2(\gamma_{j}^{i})^{3}}\frac{\sqrt{\pi}}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{C}{2(\gamma_{j}^{i})^{2}} + \frac{\sqrt{\pi}}{2}\frac{\vartheta_{j}^{i}}{\gamma_{j}^{i}} + \frac{\sqrt{\pi}}{2}\frac{\vartheta_{j}^{i}}{\gamma_{j}^{i}}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})},$$

$$(29)$$

$$\xi = \int_{0}^{\infty} \exp\left\{-\sigma^{2}(Z_{j}^{i}-\tau)^{2}\right\} (Z_{j}^{i})^{2} dZ_{j}^{i}$$

$$= \frac{1}{\sigma^{3}} \int_{0}^{\infty} \exp\left\{-(Z_{j}^{i}\sigma-\sigma\tau)^{2}\right\} (Z_{j}^{i}\sigma)^{2} dZ_{j}^{i}\sigma^{y=Z_{j}^{i}\sigma-\sigma\tau}$$

$$= \frac{1}{\sigma^{3}} \int_{-\sigma\tau}^{0} y^{2} \exp\left\{-y^{2}\right\} dy + \frac{1}{\sigma^{3}} \int_{0}^{\infty} y^{2} \exp\left\{-y^{2}\right\} dy$$
(24)

where *C* is a constant. Following the classical decision directed approach in [25], we have  $\varphi_i^i$  as

$$\varphi_{j}^{i} = (1 - \upsilon^{2})\varphi_{j-1}^{i} + \upsilon^{2} \max((\delta_{j}^{i} - 1), 0),$$
(30)

where  $\upsilon \in (0, 1)$  is the gain parameter. For the sake of convenience, we donate GMMSE-KGL to name the

Page 7 of 16

proposed MMSE graph magnitude spectral estimatorbased *K*-graphs learning in the following sections.

#### 5 Numerical results and discussions

In this section, we present the output signal-to-noise ratio (SNR), perceptual evaluation of speech quality (PESQ) [47], log-likelihood ratio (LLR) [48], and short-time objective intelligibility (STOI) [49] measure results of the proposed MMSE graph spectral magnitude estimator. The traditional MMSE-STSA method in [25], the optimal modified minimum mean-square error log-spectral method (OMLSA) in [26], the improved graph Wiener filtering method (GWF-SGS) in [30], the vertex-frequency graph Wiener filtering (VFGWF) in [32], the graph Wiener filtering method for directed cyclic time series (GWF-DCGS), and the graph Wiener filtering for arbitrary graph signals (GWF-AGS) in [33] are used as the benchmarks. The output LLR is defined as

$$LLR = \log\left(\overrightarrow{d_p} \mathbf{R_c} \overrightarrow{d_p}^{\mathrm{T}} / \overrightarrow{d_c} \mathbf{R_c} \overrightarrow{d_c}^{\mathrm{T}}\right), \qquad (31)$$

where  $\overrightarrow{d_p}$  and  $\overrightarrow{d_c}$  represent the LPC vector of the enhanced speech frame and original speech signals, respectively. **R**<sub>c</sub> represents the autocorrelation matrix of the original speech signals [50]. In our numerical simulations, the noisy speech signals are generated by mixing pure speech signals from the TIMIT database [51] with noise signals at the input signal-to-noise ratios (SNRs) from -15 to 5 dB. Two hundred sentences consisting of 20 speakers (10 females and 10 males) are used as the clean speech signal. White noise, Gaussian color noise, and Babble noise from NOISES-92 library [52] are used as noise signals. The sampling frequency is 16 kHz. The speech signals are framed by the Hamming window with a length of 256 points and an overlap of 50%.

Figure 2 shows the traditional spectrogram obtained by the discrete Fourier transform (DFT) and the graph spectrogram obtained by the proposed graph Fourier



Fig. 2 Clean speech signals and its spectrograms. a The clean speech signals. b The traditional spectrogram. c The graph spectrogram

basis based on the graph weight matrix  $\mathbf{W}_i$  of the intergraph, respectively. Observe from Fig. 2 that the graph spectrogram is mainly distributed in the high graph frequency regions, while the traditional spectrogram is mainly distributed in the low-frequency regions. The reason is that considering the Theorem 1 for the graph frequency ordering in [53], the smallest eigenvalue of  $\mathbf{W}_i$ represents the lowest frequency, and its largest eigenvalue is the highest frequency. These are different from the traditional frequencies. Although the graph spectrogram is similar to that of the conventional spectrogram, this graph spectrogram is utterly different from that of the traditional spectrum. In addition, the proposed graph Fourier basis can map speech signals into the real graph frequency field by applying the eigenvector matrix of  $\mathbf{W}_i$ .

Figure 3 shows the output SNR of the proposed GMMSE-KGL method in the case of white noise versus the frame number  $M_s$  where  $N_s = 256$ . Observe from Fig. 3, to achieve a high output SNR,  $M_s$  should be neither too small nor too large. When it takes a small value, the relationships among non-adjacent frames cannot be well described as the designed small multiple graphs. In contrast, when  $M_s$  takes a large value, the boundaries between sub-multiple graphs might have a similar tendency, which would degrade the output SNR. Because the larger multiple graphs will lose some details of speech samples, resulting in estimating the inability to the graph spectral magnitude of speech samples well. Hence, the

range of  $M_s$  can be  $M_s \in [20, 30]$  in the case of  $N_s = 256$ , and we use  $M_s = 30$  in our numerical simulations below.

Figure 4 shows the output SNR results of the proposed GMMSE-KGL method under white noise with the different vertex number  $N_s$  of the intra-graph topology  $O_i$  where  $M_s = 30$ . We can see from Fig. 4 that the performance of the proposed GMMSE-KGL method decreases with the increase of  $N_s$ . The reason for this is that the intra-graph topology becomes more complex and larger as  $N_s$  increases, resulting in the gain function of the proposed GMMSE-KGL method is not accurately estimated by using the SGS's graph power spectrum on  $O_i$ . Moreover, the proposed GMMSE-KGL method in the case of  $N_s = 64$  would obtain a better performance. Considering the fairness of comparison, we use the frame with length 256 and build our intra-graph with  $N_s = 256$ , that is,  $O_i = (\mathcal{V}_i, \mathbf{W}_i^{256 \times 256})$ .

Table 1 shows the PESQ results of the *K*-graphs learning (KGL) followed by the GMMSE method, the graph k-shift operator (GKSO) [30] followed by the GMMSE method and the graph learning (GL) [32] followed by the GMMSE method in the case of white noise. For writing convenience, we denote the methods above as the GMMSE-KGL, GMMSE-KGS, and GMMSE-GL, respectively. We can observe from Table 1 that the PESQ of the proposed GMMSE-KGL outperforms that of the GMMSE-GKSO and GMMSE-GL, which illustrates the effectiveness of the *K*-graphs learning part in speech



Fig. 3 The output SNR versus the cluster number K of the proposed GMMSE-KGL method



**Fig. 4** The output SNR versus the sample number  $N_s$  of the proposed GMMSE-KGL method

 Table 1
 The PESQ of the different graph models followed by the GMMSE method

PESQ							
	Input SNR						
Graph models	-15 dB	-10 dB	-5 dB	0 dB	5 dB		
GMMSE-GKSO	1.028	1.283	1.058	1.465	1.420		
GMMSE-GL	0.984	1.128	1.408	1.798	2.109		
GMMSE-KGL	0.995	1.162	1.532	1.934	2.277		

**Table 2** The PESQ of different graph speech enhancement

 methods followed by the K-graphs learning method
 Image: Comparison of the second secon

PESQ							
	Input SNR						
Algorithms	-15 dB	-10 dB	—5 dB	0 dB	5 dB		
Noisy	0.954	1.079	1.253	1.560	1.924		
GWF-KGL	0.984	1.065	1.304	1.633	2.001		
GMMSE-KGL	0.995	1.162	1.532	1.934	2.277		

enhancement. Moreover, the PESQ of the proposed GMMSE-KGL is 0.5 higher than that of GMMSE-GKSO and is 0.2 higher than that of GMMSE-GL, when the input SNR is larger than -5 dB.

	Input SNR						
Algorithms	-15 dB	-10 dB	-5 dB	0 dB	5 dB		
Noisy	0.954	1.079	1.253	1.560	1.924		
MMSE-STSA	0.916	1.063	1.413	1.834	2.201		
OMLSA	0.920	0.874	1.255	1.757	2.246		
GWF-SGS	0.984	1.065	1.304	1.633	2.001		
VFGWF	1.002	1.067	1.347	1.711	2.078		
GWF-AGS	0.849	0.959	1.109	1.318	1.628		
GWF-DCGS	0.885	1.081	1.166	1.244	1.349		
GMMSE-KGL	0.995	1.162	1.532	1.934	2.277		

Table 2 shows the PESQ results of the proposed GMMSE and the graph Wiener filtering (GWF) method combined with the *K*-graphs learning method in the case of white noise, respectively. For writing convenience, we denote the two methods above as the GMMSE-KGL and GWF-KGL. From Table 2, we can see that the PESQ of the proposed GMMSE-KGL is 0.2 higher than that of GWF-KGL. The PESQ results of the proposed GMMSE-KGL demonstrate the effectiveness of the GMMSE part in speech enhancement. Tables 1 and 2 illustrate that the proposed *K*-graphs learning part and the GMMSE-based

Table 3 shows the PESQ results of the proposed GMMSE-KGL, GWF-DCGS, GWF-AGS, GWF-SGS, VFGWF, MMSE-STSA, and OMLSA methods for the case of white noise. Observe from Table 3 that though the proposed GMMSE-KGL method leads to a slightly lower PESQ value as compared to the VFGWF method in the case of -15 dB input SNR, the proposed GMMSE-KGL method outperforms all the benchmarks when the input SNR is more than -10dB in terms of PESQ, which shows the advantage of the proposed GMMSE-KGL method.

Table 4 shows the LLR results of the proposed GMMSE-KGL, GWF-DCGS, GWF-AGS, GWF-SGS, VFGWF, MMSE-STSA, and OMLSA methods for the case of white noise. We can see from Table 4 that the proposed GMMSE-KGL method performs better than the GWF-DCGS, GWF-AGS, VFGWF, and MMSE-STSA methods in terms of LLR results. However, when the input SNR is between -10 and -5 dB, the proposed GMMSE-KGL method performs worse than the OMLSA method. Because OMLSA could estimate the noise very well by applying the MCRA method, estimating the log-spectral amplitude of clean speech signals.

Table 5 shows the output SNR results of the proposed GMMSE-KGL method, GWF-DCGS, GWF-AGS, GWF-SGS, VFGWF, MMSE-STSA, and OMLSA methods for the case of white noise. From Table 5, we can see that the output SNR of the proposed GMMSE-KGL method is higher than that of the GWF-DCGS and the GWF-AGS methods, regardless of the input SNR. The proposed GMMSE-KGL method is also higher than that of the GWF-SGS, VFGWF, and OMLSA methods for the cases where the input SNR is lower than -5 dB. However, when the input SNR is higher than -5 dB, the proposed GMMSE-KGL method leads to a lower output SNR as compared to the MMSE-STSA, OMLASL, GWF-SGS,

 Table 4
 The LLR of different methods for white noise

Wang et al. EURASIP Journal on Audio, Speech, and Music Processing	(2023) 2023:7

 Table 5
 The output SNR of different methods for white noise

Output SNR							
	Input SNR						
Algorithms	-15 dB	-10 dB	-5 dB	0 dB	5 dB		
MMSE-STSA	-0.633	1.488	3.358	5.225	6.763		
OMLSA	-0.046	0.961	2.356	3.869	4.842		
GWF-SGS	-7.533	-3.275	0.794	3.897	7.865		
VFGWF	-7.686	-3.386	0.736	4.732	8.746		
GWF-AGS	0.108	0.116	0.091	0.066	0.050		
GWF-DCGS	0.022	0.114	0.035	0.011	0.040		
GMMSE-KGL	0.304	0.414	2.694	3.384	3.876		

and VFGWF methods. The reason for this is that the proposed GMMSE-KGL method may misestimate the graph spectra of clean speech signals, so it cannot estimate the graph magnitude of pure graph speech signals.

Table 6 shows the STOI results of the proposed GMMSE-KGL method and the benchmarks in the case of white noise. Observe from Table 6 that the proposed GMMSE-KGL method outperforms the OMLSA, MMSE-STSA, GWF-AGA, and GWF-DCGS methods overall. However, the proposed GMMSE-KGL method performs worse than the GWF-SGS, and VFGWF methods in terms of STOI. The reason for this is that when we learn the dynamic graphs using the *K*-graphs learning method, the graph spectrum of some noises is regarded as that of clean speech details, leading to estimating accurately the graph spectrum of clean speech signals.

Table 7 reports the PESQ results of the proposed GMMSE-KGL, GWF-DCGS, GWF-AGS, VFGWF, MMSE-STSA, and OMLSA methods in a comparative study on suppressing Gaussian color noise. Observe from Table 7 that the proposed GMMSE-KGL method outperforms all the benchmarks in terms of PESQ overall. Moreover, from Table 7 we can see that when the input

LLR							
	Input SNR						
Algorithms	-15 dB	-10 dB	-5 dB	0 dB	5 dB		
Noisy	1.638	1.631	1.617	1.580	1.514		
MMSE-STSA	1.627	1.616	1.591	1.532	1.430		
OMLSA	1.677	1.670	1.646	1.579	1.468		
GWF-SGS	1.638	1.631	1.617	1.578	1.510		
VFGWF	1.643	1.634	1.611	1.565	1.485		
GWF-AGS	1.642	1.635	1.620	1.581	1.514		
GWF-DCGS	1.652	1.630	1.630	1.594	1.517		
GMMSE-KGL	1.745	1.649	1.634	1.626	1.569		

Table 6 The STOI of different methods for white noise

STOI							
	Input SNR						
Algorithms	-15 dB	-10 dB	-5 dB	0 dB	5 dB		
Noisy	0.388	0.464	0.571	0.697	0.817		
MMSE-STSA	0.405	0.461	0.538	0.630	0.719		
OMLSA	0.360	0.404	0.489	0.627	0.756		
GWF-SGS	0.406	0.506	0.626	0.743	0.837		
VFGWF	0.396	0.494	0.613	0.730	0.829		
GWF-AGS	0.511	0.535	0.562	0.586	0.631		
GWF-DCGS	0.528	0.543	0.565	0.555	0.560		
GMMSE-KGL	0.372	0.461	0.567	0.675	0.772		

5 dB

6.318

4.642

7.817

8 2 8 9

0.016

0111

4.591

<b>Table 7</b> The PESQ of different methods for Gaussian color noise
---

DECO

250							
	Input SNR						
Algorithms	_15 dB	-10 dB	-5 dB	0 dB	5 dB		
Noisy	0.937	1.058	1.181	1.437	1.797		
MMSE-STSA	0.940	0.966	1.281	1.692	2.097		
OMLSA	1.150	0.884	1.125	1.586	2.100		
GWF-SGS	0.981	1.006	1.195	1.500	1.870		
VFGWF	0.939	1.038	1.223	1.558	1.934		
GWF-AGS	0.836	0.960	1.049	1.231	1.356		
GWF-DCGS	0.838	0.940	1.104	1.253	1.471		
GMMSE-KGL	0.971	1.132	1.424	1.850	2.233		

 Table 8
 The LLR of different methods for Gaussian color noise

LLR	LLR						
	Input SNR						
Algorithms	-15 dB	-10 dB	-5 dB	0 dB	5 dB		
Noisy	1.998	1.986	1.938	1.815	1.611		
MMSE-STSA	1.967	1.907	1.810	1.662	1.454		
OMLSA	1.917	1.857	1.724	1.537	1.335		
GWF-SGS	1.998	1.986	1.938	1.815	1.609		
VFGWF	1.992	1.972	1.920	1.800	1.584		
GWF-AGS	1.889	1.859	1.814	1.748	1.553		
GWF-DCGS	1.985	1.969	1.912	1.789	1.595		
GMMSE-KGL	1.891	1.918	1.828	1.664	1.463		

 Table 10 The STOI of different methods for Gaussian color noise

 STOI

Table 9 The output SNR of different methods for Gaussian color

-10 dB

1.314

0.673

-3.521

-4.352

0.049

0 2 9 9

1.023

-5 dB

2,929

1.898

0.791

0.028

0.026

0.275

2.893

0 dB

4.693

3.474

3.698

3.997

0.032

0.185

3.565

Input SNR

-15 dB

-0.692

-0.037

-7.364

-8.846

0.024

0.232

0.292

noise

**Output SNR** 

Algorithms

MMSE-STSA

OMI SA

VFGWF

GWF-SGS

GWF-AGS

GWF-DCGS

GMMSE-KGL

	Input SNR						
Algorithms	-15 dB	-10 dB	-5 dB	0 dB	5 dB		
Noisy	0.371	0.435	0.531	0.650	0.775		
MMSE-STSA	0.391	0.436	0.507	0.593	0.685		
OMLSA	0.356	0.396	0.461	0.582	0.719		
GWF-SGS	0.386	0.476	0.594	0.709	0.807		
VFGWF	0.377	0.463	0.578	0.695	0.799		
GWF-AGS	0.513	0.543	0.579	0.597	0.609		
GWF-DCGS	0.520	0.530	0.540	0.543	0.528		
GMMSE-KGL	0.351	0.437	0.533	0.641	0.740		

SNR is –15dB, the PESQ of the proposed GMMSE-KGL method is lower than that of the OMLSA and GWF-SGS methods.

Table 8 shows the LLR results in a comparative study on suppressing Gaussian color noise. Observe from Table 8 that the LLR results of the proposed GMMSE-KGL method are slightly higher than the traditional MMSE-STSA and OMLSA methods from 0 to 5 dB and are lower than those of the GWF-DCGS, GWF-AGS, GWF-SGS, and VFGWF methods when the input SNR is no less than 0 dB. This illustrates that the proposed GMMSE-KGL method provides a slightly better spectral envelope than the traditional MMSE-STSA and OMLSA methods. However, the proposed GMMSE-KGL method cannot estimate the graph spectrum reasonably as compared to the existing GSP-based methods.

Table 9 shows the output SNR results in a comparative study on suppressing Gaussian color noise. From Table 9, we observe that in terms of its output SNR, the proposed GMMSE-KGL method outperforms the GWF-DCGS, GWF-AGS, GWF-SGS, VFGWF, and OMLSA methods in cases of negative input SNRs. In the positive input SNRs, the relationship among noise frames could not be captured by the *K*-graphs learning method, leading to not estimating noise power spectra in real-time. The performance of the proposed GMMSE-KGL method is lower than the GWF-SGS and VFGWF methods. Additionally, compared to the MMSE-STSA method, the proposed GMMSE-KGL method leads to a much smaller improvement in the output SNR results. The reason for this is that the proposed GMMSE-KGL method reduces more details of clean speech signals and has a slight improvement in terms of the output SNR.

Table 10 shows the STOI results of the proposed GMMSE-KGL, GWF-DCGS, GWF-AGS, GWF-SGS, VFGWF, MMSE-STSA, and OMLSA methods in the case of Gaussian color noise. Observe from Table 10 that the STOI results of the proposed GMMSE-KGL method are higher than that of the MMSE-STSA, OMLSA, GWF-DCGS, and GWF-AGS methods when the input SNR is between -15 and 5dB. Note that the proposed GMMSE-KGL method estimates the graph spectrum of some Gaussian color noise as that of clean speech signals,

	Input SNR					
Algorithms	_15 dB	-10 dB	—5 dB	0 dB	5 dB	
Noisy	1.472	1.460	1.633	1.920	2.240	
MMSE-STSA	0.981	1.196	1.582	1.993	2.374	
OMLSA	1.068	1.263	1.624	1.978	2.354	
GWF-SGS	1.174	1.264	1.514	1.671	2.122	
VFGWF	1.306	1.511	1.692	1.977	2.284	
GWF-AGS	0.975	1.204	1.456	1.683	1.799	
GWF-DCGS	1.084	1.321	1.179	1.474	1.475	
GMMSE-KGL	1.594	1.626	1.939	2.230	2.492	

# Table 11 The PESQ of different methods for Babble noise

PESO

Table 12 The LLR of different methods for Babble noise

LLR						
	Input SNR					
Algorithms	-15 dB	-10 dB	-5 dB	0 dB	5 dB	
Noisy	1.423	1.318	1.149	0.934	0.709	
MMSE-STSA	1.559	1.440	1.260	1.050	0.835	
OMLSA	1.585	1.465	1.287	1.074	0.850	
GWF-SGS	1.998	1.986	1.938	1.814	1.610	
VFGWF	1.471	1.356	1.204	0.997	0.773	
GWF-AGS	1.415	1.315	1.156	0.954	0.749	
GWF-DCGS	1.395	1.324	1.209	1.074	0.930	
GMMSE-KGL	1.885	1.887	1.838	1.810	1.784	

resulting in obtaining low STOI results as compared to the GWF-SGS and VFGWF methods.

Table 11 shows the PESQ results of the proposed GMMSE-KGL, GWF-DCGS, GWF-AGS, GWF-SGS, VFGWF, MMSE-STSA, and OMLSA methods for Babble noise. Observe from Table 11 that the proposed GMMSE-KGL method outperforms all the benchmarks in terms of PESQ with input SNR ranging from -15 to 5 dB.

Table 12 shows the LLR results of the proposed GMMSE-KGL, GWF-DCGS, GWF-AGS, GWF-SGS, VFGWF, MMSE-STSA, and OMLSA for Babble noise. Observe from Table 12 that the proposed GMMSE-KGL method performs better than the traditional MMSE-STSA and OMLSA methods when the input SNR is between -10 and 5dB. However, the proposed GMMSE-KGL method cannot outperform the GWF-SGS, VFGWF, GWF-DCGS, and GWF-AGS methods when the input SNR is between -15 and 5 dB. The reason for this is that the proposed GMMSE-KGL method may not estimate the graph spectrum of the nonstationary noise

Table 13 The output SNR of different methods for Babble noise

Output SNR						
	Input SNR					
Algorithms	-15 dB	-10 dB	-5 dB	0 dB	5 dB	
MMSE-STSA	-7.691	-3.262	0.830	4.741	8.442	
OMLSA	-2.119	0.678	2.604	3.873	4.753	
GWF-SGS	-10.816	-6.075	-1.547	2.359	6.599	
VFGWF	-9.100	-8.986	-4.057	0.830	5.727	
GWF-AGS	0.021	0.160	0.243	0.225	0.147	
GWF-DCGS	0.046	0.073	0.023	0.019	0.010	
GMMSE-KGL	-0.521	-0.243	0.211	1.022	2.304	

Table 14	The STOL of	different	methods	for Babble	noise
		unerent	I I I E I I U U S		: HUBC

STOI						
	Input SNR					
Algorithms	-15 dB	-10 dB	-5 dB	0 dB	5 dB	
Noisy	0.342	0.418	0.524	0.644	0.758	
MMSE-STSA	0.253	0.342	0.471	0.610	0.735	
OMLSA	0.265	0.356	0.481	0.614	0.736	
GWF-SGS	0.283	0.372	0.506	0.640	0.764	
VFGWF	0.255	0.407	0.519	0.645	0.760	
GWF-AGS	0.506	0.533	0.567	0.586	0.573	
GWF-DCGS	0.520	0.540	0.546	0.527	0.494	
GMMSE-KGL	0.315	0.391	0.498	0.603	0.697	

as compared to the GWF-SGS, VFGWF, GWF-DCGS, and GWF-AGS methods.

Table 13 shows the output SNR results of the proposed GMMSE-KGL, GWF-DCGS, GWF-AGS, GWF-SGS, VFGWF, MMSE-STSA, and OMLSA methods for Babble noise. Observe from Table 13 that the proposed GMMSE-KGL method outperforms the GWF-SGS, VFGWF, and MMSE-STSA methods when the input SNR is lower than -5dB. Though the GWF-AGS and GWF-DCGS methods could be better than GMMSE-KGL in negative input SNR values, our proposed GMMSE-KGL method performs better than the GWF-AGS and GWF-DCGS methods when input SNRs are positive.

Table 14 shows the STOI results of the proposed GMMSE-KGL, GWF-DCGS, GWF-AGS, GWF-SGS, VFGWF, MMSE-STSA, and OMLSA methods for Babble noise. It can be observed from Table 14 that when the input SNR is less than 0dB, the STOI results of the proposed GMMSE-KGL method are higher than those of MMSR, OMLSA, and GWF-SGS methods and are lower than those of the VFGWF, GWF-AGS, and GWF-DCGS methods. The above situation is reversed when the input

Algorithms	The complexity		
MMSE-STSA	$\mathcal{O}(N_{s}\log_{2}N_{s}+N_{s}\log N_{s})$		
OMLSA	$\mathcal{O}(N_{s}\log_{2}N_{s}+N_{s}\log N_{s})$		
GWF-SGS	$\mathcal{O}(N_s + N_s^2)$		
VFGWF	$\mathcal{O}(M^2N_s)$		
GWF-AGS	$\mathcal{O}(LN_s)$		
GWF-DCGS	$\mathcal{O}(LN_s)$		
GMMSE-KGL	$\mathcal{O}(M_s^2 N_s)$		

Table 15 The computational complexity of different methods

SNR is not less than 0dB. This implies that the proposed GMMSE-KGL method may not estimate a graph spectrum of Babble noise, and even some speech details may be regarded as Babble noise and removed in higher input SNR cases.

Table 15 gives the computational complexity of the proposed GMMSE-KGL, GWF-DCGS, GWF-AGS, GWF-SGS, VFGWF, MMSE-STSA, and OMLSA methods. Here we discuss the computational complexity of methods caused by the graph/discrete Fourier transform and its inverse transform. To be specific, let M, M<sub>s</sub> N<sub>s</sub>, L represent the number of whole noisy speech frames, the number of noisy speech frames in a cluster, the length of a noisy speech frame, and the tap number of the graph Wiener filtering, respectively. Based on the computational complexity discussion in [30, 32] and the matrix multiplication theory, we can observe from Table 15 that the computational complexity of the proposed GMMSE-KGL method is higher than that of traditional methods. The reason for this is that the proposed GMMSE-KGL method does not apply the fast graph Fourier transform, while all the traditional methods use the FFT operation. Meanwhile, the proposed GMMSE-KGL method based on the Fourier basis of the sparse matrix  $\mathcal{L}^*$  has a lower computational complexity compared to both the GWF-DCGS method and that of the GWF-AGS method. Note that  $M_s$  is smaller than M, and the computational complexity of the proposed GMMSE-KGL method is lower than that of the VFGWF method.

# 6 Conclusions

This paper used the *K*-graphs learning method to learn multiple graphs for speech signals, which can investigate intrinsic relationships among inter-frames and the relationship between speech samples within a frame in real-time. On this basis, we developed a representation of the MMSE graph spectral magnitude estimator and used different input SNRs to evaluate the performance of the proposed GMMSE-KGL method on speech enhancement. The experimental results showed that the proposed GMMSE-KGL method outperformed the graph Wiener filtering methods in GSP on the PESQ and was comparable to some of the well-performing traditional baseline methods in DSP in terms of the LLR, STOI, and output SNR.

# Appendixes

# Appendix 1

In this Appendix, we derive the molecular fraction  $\xi$  in (23)

$$\begin{aligned} \xi &= \int_0^\infty \exp\left\{-\sigma^2 (Z_j^i - \tau)^2\right\} (Z_j^i)^2 dZ_j^i \\ &= \frac{1}{\sigma^3} \int_0^\infty \exp\left\{-(Z_j^i \sigma - \sigma \tau)^2\right\} (Z_j^i \sigma)^2 dZ_j^i \sigma. \end{aligned}$$
(32)

Upon denoting  $y = Z_i^i \sigma - \sigma \tau$ , we have

$$\xi = \frac{1}{\sigma^3} \int_{-\sigma\tau}^0 \left( y^2 + 2y\sigma\tau + (\sigma\tau)^2 \right) \exp\{-y^2\} dy + \frac{1}{\sigma^3} \int_0^\infty \left( y^2 + 2y\sigma\tau + (\sigma\tau)^2 \right) \exp\{-y^2\} dy.$$
(33)

Using the representation of the combinations of exponentials and arbitrary power [[54], eq.3.381.4]

$$\int_0^\infty x^{\nu-1} \exp\{-\mu \mathbf{v}\} \mathrm{d}\mathbf{x},\tag{34}$$

we obtain

$$\begin{split} \xi &= \frac{1}{\sigma^3} \left( \int_{-\sigma\tau}^0 y^2 \exp\{\{-y^2\} dy + \int_{-\sigma\tau}^0 2y\sigma\tau \exp\{\{-y^2\} dy + \int_{-\sigma\tau}^0 (\sigma\tau)^2 \exp\{\{-y^2\} dy \right) \right. \\ &+ \frac{1}{\sigma^3} \left( \int_0^\infty y^2 \exp\{\{-y^2\} dy + \int_0^\infty 2y\sigma\tau \exp\{\{-y^2\} dy + \int_0^\infty (\sigma\tau)^2 \exp\{\{-y^2\} dy \right) \right. \\ &= \frac{1}{\sigma^3} \left( -\frac{1}{2} \int_{-\sigma\tau}^0 y d\exp\{\{-y^2\} - \sigma\tau \int_{-\sigma\tau}^0 d\exp\{\{-y^2\} + (\sigma\tau)^2 \int_{-\sigma\tau}^0 \exp\{\{-y^2\} dy \right) \right. \\ &+ \left( \frac{1}{\sigma^3} \Gamma\left(\frac{3}{2}\right) + \frac{\tau}{\sigma^2} \Gamma(1) + \frac{\tau^2}{\sigma} \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \right) \\ &= \frac{1}{2\sigma^3} \left( -y \exp\{\{-y^2\} \right) \Big|_{-\sigma\tau}^0 + \frac{\tau^2}{\sigma^2} \exp\{\{-y^2\} \Big|_{-\sigma\tau}^0 + \frac{\tau^2}{\sigma} \int_{-\sigma\tau}^0 \exp\{\{-y^2\} dy \right. \\ &+ \left( \frac{\pi}{4\sigma^3} + \frac{\tau}{\sigma^2} + \frac{\tau^2 \sqrt{\pi}}{2\sigma} \right) \\ &= \frac{-\tau}{2\sigma^2} \exp\{\{-\sigma^2\tau^2\} + \frac{\tau}{\sigma^2} (\exp\{\{-\sigma^2\tau^2\} - 1) + \frac{\tau^2}{\sigma} \int_{-\sigma\tau}^0 \exp\{\{-y^2\} dy \right. \\ &+ \left( \frac{\pi}{4\sigma^3} + \frac{\tau}{\sigma^2} + \frac{\tau^2 \sqrt{\pi}}{2\sigma} \right) . \end{split}$$

Using the representation of the Gauss error function  $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\eta^2} d\eta$  [[54], 3.321.1] and upon denoting -y = t, we obtain

$$\begin{split} \xi &= \frac{\tau}{2\sigma^2} \exp\{-\sigma^2 \tau^2\} + \frac{1}{\sigma^3} \left(\frac{\pi}{4} + \frac{\sqrt{\pi}}{2} \sigma^2 \tau^2\right) + \frac{\tau^2}{\sigma} \int_{-\sigma\tau}^{0} \exp\{-y^2\} dy \\ &= \frac{\tau}{2\sigma^2} \exp\{-\sigma^2 \tau^2\} + \frac{1}{\sigma^3} \left(\frac{\pi}{4} + \frac{\sqrt{\pi}}{2} \sigma^2 \tau^2\right) + \frac{\sqrt{\pi}}{2} \cdot \frac{2}{\sqrt{\pi}} \int_{0}^{\sigma\tau} \exp\{-t^2\} dt \\ &= \frac{\tau}{2\sigma^2} \exp\{-\sigma^2 \tau^2\} + \frac{1}{\sigma^3} \left(\frac{\pi}{4} + \frac{\sqrt{\pi}}{2} \sigma^2 \tau^2\right) + \frac{\sqrt{\pi}}{2} \operatorname{erf}(\sigma\tau). \end{split}$$

$$(36)$$

## Appendix 2

In this Appendix, we derive the gain function (28) of MMSE graph magnitude estimator, the denominator fraction  $\zeta$  in (22) is given as

$$\zeta = \int_0^\infty \exp\left\{-\sigma^2 (Z_j^i - \tau)^2\right\} Z_j^i dZ_j^i.$$
(37)

Upon denoting  $y = Z_i^i \sigma - \sigma \tau$ , we have

Using the representation of the Gauss error function  $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\eta^2} d\eta$  [[54], 3.321.1] and upon denoting *-y*=t, we have

$$\zeta = \frac{1}{2\sigma^2} \exp\left\{-\sigma^2 \tau^2\right\} + \frac{\sqrt{\pi}}{2} \frac{\tau}{\sigma} + \frac{\tau}{\sigma} \int_0^{\sigma\tau} \exp\left\{-t^2\right\} dt$$
$$= \frac{1}{2\sigma^2} \exp\left\{-\sigma^2 \tau^2\right\} + \frac{\sqrt{\pi}}{2} \frac{\tau}{\sigma} + \frac{\tau}{\sigma} \frac{\sqrt{\pi}}{2} \cdot \frac{2}{\sqrt{\pi}} \int_0^{\sigma\tau} \exp\left\{-t^2\right\} dt$$
$$= \frac{1}{2\sigma^2} \exp\left\{-\sigma^2 \tau^2\right\} + \frac{\sqrt{\pi}}{2} \frac{\tau}{\sigma} + \frac{\tau}{\sigma} \frac{\sqrt{\pi}}{2} \operatorname{erf}(\sigma\tau).$$
(39)

Submitting the (A.5) and (B.3) into (23), we obtain

$$\widehat{Z_{j}^{i}} = \frac{\xi}{\zeta} = \frac{\frac{\tau}{2\sigma^{2}} \exp\{-\sigma^{2}\tau^{2}\} + \frac{1}{\sigma^{3}}(\frac{\pi}{4} + \frac{\sqrt{\pi}}{2}\sigma^{2}\tau^{2}) - \frac{1+2\sigma^{2}\tau^{2}}{2\sigma^{3}}\frac{\sqrt{\pi}}{2}erf(\sigma\tau)}{\frac{1}{2\sigma^{2}} \exp\{-\sigma^{2}\tau^{2}\} + \frac{\sqrt{\pi}}{2}\frac{\tau}{\sigma} + \frac{\sqrt{\pi}}{2}\frac{\tau}{\sigma}erf(\sigma\tau)}R_{j}^{i}.$$
(40)

Hence, the gain function of the MMSE graph spectral magnitude estimator can be described as

$$G_{\text{GMMSE}}(\gamma_{j}^{i},\vartheta_{j}^{i}) = \frac{\widehat{Z}_{j}^{i}}{R_{j}^{i}} = \frac{\frac{\vartheta_{j}^{i}}{2(\gamma_{j}^{i})^{2}} \exp\left\{-(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}\right\} + \frac{1}{(\gamma_{j}^{i})^{3}}(\frac{\pi}{4} + \frac{\sqrt{\pi}}{2}(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}) - \frac{1+2(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}}{2\gamma_{j}^{3}}\frac{\sqrt{\pi}}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{1}{2(\gamma_{j}^{i})^{2}} \exp\left\{-(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}\right\} + \frac{\sqrt{\pi}}{2}\frac{\vartheta_{j}^{i}}{\gamma_{j}^{i}} + \frac{\sqrt{\pi}}{2}\frac{\vartheta_{i}}{\gamma_{i}}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{1}{2(\gamma_{j}^{i})^{2}} + \frac{1}{(\gamma_{j}^{i})^{3}}(\frac{\pi}{4} + \frac{\sqrt{\pi}}{2}(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}) - \frac{1+2(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}}{2(\gamma_{j}^{i})^{3}}\frac{\sqrt{\pi}}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{2}{2(\gamma_{j}^{i})^{3}} - \frac{1}{2}\frac{(\gamma_{j}^{i}\vartheta_{j}^{i})^{2}}{2(\gamma_{j}^{i})^{3}} - \frac{1}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{1}{2(\gamma_{j}^{i})^{3}} - \frac{1}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}}{\frac{1}{2(\gamma_{j}^{i})^{3}} - \frac{1}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{1}{2(\gamma_{j}^{i})^{3}} - \frac{1}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{1}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}}{\frac{1}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{1}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}}{\frac{1}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{1}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}}{\frac{1}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{1}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}}{\frac{1}{2}\operatorname{erf}(\gamma_{j}^{i}\vartheta_{j}^{i})}{\frac{1}{2}\operatorname{erf}(\gamma_{$$

$$\begin{split} \zeta &= \frac{1}{\sigma^2} \int_{-\sigma\tau}^0 (y + \sigma\tau) \exp\{-y^2\} dy + \frac{1}{\sigma^2} \int_0^\infty (y + \sigma\tau) \exp\{-y^2\} dy \\ &= \frac{1}{\sigma^2} \int_{-\sigma\tau}^0 y \exp\{-y^2\} dy + \frac{\tau}{\sigma} \int_{-\sigma\tau}^0 \exp\{-y^2\} dy + \frac{1}{\sigma^2} \int_0^\infty y \exp\{-y^2\} dy \\ &+ \frac{\tau}{\sigma} \int_0^\infty \exp\{-y^2\} dy \\ &= \frac{-1}{2\sigma^2} \int_{-\sigma\tau}^0 d \exp\{-y^2\} + \frac{\tau}{\sigma} \int_{-\sigma\tau}^0 \exp\{-y^2\} dy + \frac{1}{\sigma^2} \frac{1}{2} \Gamma(1) + \frac{\tau}{\sigma} \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \\ &= \frac{-1}{2\sigma^2} \exp\{-y^2\} \Big|_{-\sigma\tau}^0 + \frac{1}{2\sigma^2} \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \\ &= \frac{1}{2\sigma^2} \exp\{-\sigma^2\tau^2\} \frac{1}{2\sigma^2} + \frac{\tau\sqrt{\pi}}{2\sigma} + \frac{\tau}{\sigma} \int_{-\sigma\tau}^0 \exp\{-y^2\} dy. \end{split}$$
(38)

#### Acknowledgements

Not applicable.

#### Authors' contributions

(1) We proposed a novel global graph topology by using the *K*-graphs learning method, which consists of the multiple graphs to describe the potential connections among inter-frames and intra-frames. (2) We proposed the graph representation of minimum mean-square error (MMSE) graph spectral magnitude estimator for SGSs in the vertex-frequency domain by extending the classical MMSE-STSA estimator in DSP to perform speech enhancement. (3) The performance of the MMSE graph spectral magnitude estimator method outperforms the benchmarks in terms of PESQ, LLR, output SNR and STOI. The classical MMSE-STSA and OMLSA methods in DSP and the existing graph Wiener filtering methods are used as benchmarks. The authors read and approved the final manuscript.

#### Authors' information

Tingting Wang is pursuing her doctorate in signal and information processing from the Nanjing University of Posts and Telecommunications, Nanjing, China. Her current research interests include graph signal processing, classical speech signal processing.

Haiyan is currently an Associate Professor with NJUPT, Nanjing, China. Her research interests include speech signal processing and B5G/6G wireless transmission.

Zirui Ge is pursuing his doctoral degree in signal and information processing from the Nanjing University of Posts and Telecommunications, Nanjing, China. Qiquan Zhang's research interests are digital speech and audio signal processing, speech enhancement algorithms, and microphone array signal processing.

Zhen Yang's research interests include various aspects of signal processing and communication, such as communication systems and networks, cognitive radio, spectrum sensing, speech and audio processing, compressive sensing, and wireless communication. He has published more than 200 papers in academic journals and conferences. Prof. Yang served as Vice Chairman of Chinese Institute of Communications, Chairman of Jiangsu Institute of Communications from 2010 to 2015, the Chair of APCC(Asian Pacific Communication Conference) Steering Committee from 2013 to 2014. He is currently the Fellow of the Chinese Institute of Communications, Vice Director of the Editorial Board of the Journal on Communications. He is also a Member of the Editorial Board for several other journals such as Chinese Journal of Electronics et al.

#### Funding

This work is supported by the National Natural Science Foundation of China (No.62071242, No.61901227), the Graduate Innovation Program of Jiangsu Province (No.KYCX19-0897), and the China Scholarship Council.

#### Availability of data and materials

Not applicable.

### Declarations

**Ethics approval and consent to participate** Not applicable.

#### **Consent for publication**

Not applicable.

#### **Competing interests**

The authors declare that they have no competing interests.

Received: 26 March 2022 Accepted: 6 January 2023 Published online: 03 February 2023

#### References

 A. Ortega, P. Frossard, J. Kovačević, J.M.F. Moura, P. Vandergheynst, Graph signal processing: Overview, challenges, and applications. Proc. IEEE 106(5), 808–828 (2018). https://doi.org/10.1109/JPROC.2018.2820126

- D.I. Shuman, S.K. Narang, P. Frossard, A. Ortega, P. Vandergheynst, The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Process. Mag. **30**(3), 83–98 (2013). https://doi.org/10.1109/MSP.2012. 2235192
- Q.S.Q.S. Junzheng, J. David, B. Tay, Design of non-subsampled graph filter banks via lifting schemes. IEEE Signal Process. Lett. 27, 441–445 (2020).
- B. Girault, A. Ortega, S.S. Narayayan, Graph vertex sampling with arbitrary graph signal hilbert spaces, IEEE Int. Conf. Acoust., Speech, Signal Processing, Spain, 2020, 5670–5674 (2020).
- Y. Tanaka, Y.C. Eldar, A. Ortega, G. Cheung, Sampling signals on graphs: From theory to applications. IEEE Signal Process. Mag. 37(6), 14–30 (2020). https://doi.org/10.1109/MSP.2020.3016908
- A. Sandryhaila, J.M.F. Moura, Discrete signal processing on graphs. IEEE Trans. Signal Process. 61(3), 1644-1656 (2013). https://doi.org/10.1109/ TSP2013.2238935
- J. Domingos, J.M.F. Moura, Graph fourier transform: A stable approximation. IEEE Trans. Signal Process. 68, 4422–4437 (2020). https://doi.org/10. 1109/TSP.2020.3009645
- M.J.M.F. Shi J., Graph signal processing: Modulation, convolution, and sampling, 2019, https://arxiv.org/abs/1912.06762.
- S. Chen, A. Sandryhaila, J.M.F. Moura, J. Kovacevic, Signal denoising on graphs via graph filtering, 872–876 (2014). https://doi.org/10.1109/Globa ISIP.2014.7032244
- M. Onuki, S. Ono, M. Yamagishi, Y. Tanaka, Graph signal denoising via trilateral filter on graph spectral domain. IEEE Trans. Signal Inf. Process. Over Netw. 2(2), 137–148 (2016). https://doi.org/10.1109/TSIPN.2016.2532464
- 11. S. Ono, I. Yamada, I. Kumazawa, Total generalized variation for graph signals , 5456–5460 (2015). https://doi.org/10.1109/ICASSP.2015.7179014
- 12. V. Kalofolias, How to learn a graph from smooth signals **51**, 920–929 (2016). http://proceedings.mlr.press/v51/kalofolias16.html
- K. Yamada, Y. Tanaka, A. Ortega, Time-varying graph learning based on sparseness of temporal variation, 5411–5415 (2019). https://doi.org/10. 1109/ICASSP.2019.8682762
- 14. K. Yamada, Y. Tanaka, A. Ortega, Time-varying graph learning with constraints on graph temporal variation. CoRR, **abs/2001.03346** (2020). https://arxiv.org/abs/2001.03346
- G. Cheung, E. Magli, Y. Tanaka, M.K. Ng, Graph spectral image processing. IEEE, **106** (5), 907–930 (2018). https://ediss.sub.uni-hamburg.de/handle/ ediss/9268
- H. Sadreazami, A. Asif, A. Mohammadi, A late adaptive graph-based edgeaware filtering with iterative weight updating process, 1581–1584 (2017). https://doi.org/10.1109/MWSCAS.2017.8053239
- L.J. Kondor R. I., Diffusion kernels on graphs and other discrete structures, International Conference on Machine Learning, 315–322 (2002).
- A.J. Smola, R. Kondor, Kernels and regularization on graphs. 2777, 144–158 (2003). https://doi.org/10.1007/978-3-540-45167-9\_12
- B.F.e.a. Lacasa L., Luque B, From time series to complex networks: the visibility graph, 105 (13), 4972–4975 (2008).
- S.A. Bezsudnov I.V., Gavrilov S.V., From time series to complex networks: the dynamical visibility graph. Phys. A Stat. Mech. Appl. 414, 1-13 (2012). https://arxiv.org/abs/1208.6365v1
- D.J.e.a. Donner R.V., Zou Y., Recurrence networks-a novel paradigm for nonlinear time analysis, New Journal of Physics, 12(3), 129-132 (2010).
- 22. C.V.K. Mathur P., .Graph signal processing of eeg signals for detection of epilepsy, 7th International Conference on Signal Processing and Information Networks, 839–843 (2020).
- M.S.e.a. Roy S. S., Chatterjee S., Detection of focal eeg signals employing weighted visibility graph, International Conference on Computer, Electrical & Communication Engineering, India, 2020, pp. 1–5(2020).
- P. Scalart, J. Filho, Speech enhancement based on a priori signal to noise estimation, IEEE Int. Conf. Acoust., Speech, Signal Processing, USA, 1996, 629–632 (1996).
- Y. Ephraim, D. Malah, Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Sig. Process. **32**(6), 1109–1121 (1984).
- I. Cohen, B. Berdugo, Speech enhancement for non-stationary noise environments. Signal Process. 81(11), 2403–2418 (2001)
- J.H. Choi, J.H. Chang, On using acoustic environment classification for statistical model-based speech enhancement. Speech Comm. 54(3), 477–490 (2012)

- L. Sun, Y. Bu, P. Li, Z. Wu, Single-channel speech enhancement based on joint constrained dictionary learning. EURASIP J. Audio Speech Music. Process. 2021(1), 29 (2021). https://doi.org/10.1186/s13636-021-00218-3
- B.L.Z.Y. Tingting W., Haiyan G., Speech signal processing on graphs: Graph topology, graph frequency analysis and denoising. Chin. J. Elect. 29(5), 926–936 (2020)
- T. Wang, H. Guo, X. Yan, Z. Yang, Speech signal processing on graphs: The graph frequency analysis and an improved graph wiener filtering method. Speech Commun. **127**, 82–91 (2021). https://doi.org/10.1016/j. specom.2020.12.010
- M. Puschel, J.M. Moura, Algebraic signal processing theory: Foundation and 1-d time. IEEE Trans. Signal Process. 56(8–1), 3572–3585 (2008)
- T. Wang, H. Guo, Q. Zhang, Z. Yang, A new multilayer graph model for speech signals with graph learning. Digit. Signal Process. **122**, 103360 (2022). https://doi.org/10.1016/j.dsp.2021.103360
- A. Gavili, X. Zhang, On the shift operator, graph frequency, and optimal filtering in graph signal processing. IEEE Trans. Signal Process. 65(23), 6303–6318 (2017). https://doi.org/10.1109/TSP.2017.2752689
- G. Yang, L. Yang, C. Huang, An orthogonal partition selection strategy for the sampling of graph signals with successive local aggregations. Signal Process. 188, 108211 (2021). https://doi.org/10.1016/j.sigpro.2021.108211
- J. Miettinen, S.A. Vorobyov, E. Ollila, Modelling and studying the effect of graph errors in graph signal processing. Signal Process. 189, 108-256 (2021). https://doi.org/10.1016/j.sigpro.2021.108256
- H. Sevi, G. Rilling, P. Borgnat, Modeling signals over directed graphs through filtering, IEEE Global Conference on Signal and Information Processing, USA, 2018, 718–722 (2018). https://doi.org/10.1109/GlobalSIP. 2018.8646534
- F. Wang, Y. Wang, G. Cheung, A-optimal sampling and robust reconstruction for graph signals via truncated neumann series. IEEE Signal Process. Lett. 25(5), 680–684 (2018). https://doi.org/10.1109/LSP.2018. 2818062
- B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, M.G. Rabbat, Characterization and inference of graph diffusion processes from observations of stationary signals. IEEE Trans. Signal Inf. Process. Netw. 4(3), 481–496 (2018). https://doi.org/10.1109/TSIPN.2017.2742940
- X. Dong, D. Thanou, P. Frossard, P. Vandergheynst, Learning laplacian matrix in smooth graph signal representations. IEEE Trans. Sig. Process. 64(23), 6160–6173 (2016).
- Y. Yankelevsky, M. Elad, Finding GEMS: multi-scale dictionaries for highdimensional graph signals. IEEE Trans. Signal Process. 67(7), 1889–1901 (2019). https://doi.org/10.1109/TSP.2019.2899822
- F. Grassi, A. Loukas, N. Perraudin, B. Ricaud, A time-vertex signal processing framework: Scalable processing and meaningful representations for time-series on graphs. IEEE Trans. Signal Process. 66(3), 817–829 (2018). https://doi.org/10.1109/TSP.2017.2775589
- A. Loukas, D. Foucard, Frequency analysis of temporal graph signals, CoRR abs/1602.04434 (2016). http://arxiv.org/abs/1602.04434
- J. Yu, X. Xie, H. Feng, B. Hu, On critical sampling of time-vertex graph signals, IEEE Global Conference on Signal and Information Processing, Canada, 1–5 (2019). https://doi.org/10.1109/GlobalSIP45357.2019.89691 08
- H. Araghi, M. Sabbaqi, M. Babaie-Zadeh, K-graphs: An algorithm for graph signal clustering and multiple graph learning. IEEE Signal Process. Lett. 26(10), 1486–1490 (2019). https://doi.org/10.1109/LSP.2019.2936665
- X. Dong, D. Thanou, M.G. Rabbat, P. Frossard, Learning graphs from data: A signal representation perspective. IEEE Signal Process. Mag. 36(3), 44–63 (2019). https://doi.org/10.1109/MSP.2018.2887284
- B.S. Grant M., CVX: matlab software for disciplined convex programming 2012-2019 CVX Research, Inc., Austin. http://cvxr.com
- I.T. Recommendation, Perceptual evaluation of speech quality (pesq): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. Rec. ITU-T P (2001)
- C.M.A. Quackenbush S. R., Barnwell T. P., Objective measures of speech qualit (Prentice Hall Advanced Reference Series, Englewood Cliffs, 1986), ISBN: 0-13-629056-6
- C.H. Taal, R.C. Hendriks, R. Heusdens, J. Jensen, A short-time objective intelligibility measure for time-frequency weighted noisy speech, IEEE Int. Conf. Acoust., Speech, Signal Processing, USA, 2010, 4214–4217(2010). https://doi.org/10.1109/ICASSP.2010.5495701

- Y. Hu, PC. Loizou, Evaluation of objective quality measures for speech enhancement. IEEE Trans. Speech Audio Process. 16(1), 229–238 (2008). https://doi.org/10.1109/TASL.2007.911054
- S.F. Boll, DARPA TIMIT acoustic continous speech corpus CD-ROM. NIST speech disc 1-1.1. NASA STI/Recon Tech. Rep. (1993)
- A. Varga, H.J. Steeneken, Assessment for automatic speech recognition: li. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems. Speech Comm. 12(3), 247–251 (1993).
- A. Sandryhaila, J.M.F. Moura, Discrete signal processing on graphs: Frequency analysis. IEEE Trans. Signal Process. 62(12), 3042–3054 (2014). https://doi.org/10.1109/TSP.2014.2321121
- 54. R.I. Gradsh teyn I.S., Table of integrals, series and product (Academic, New York, 1980). https://doi.org/10.1109/TSP.2014.2321121

#### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- ► High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com