# Parameter Tuning-Free Missing-Feature Reconstruction for Robust Sound Recognition

Qi Liu, *Member, IEEE*, and Jibin Wu, *Member, IEEE*

***Abstract*—With the advent of the deep neural network, automatic speech recognition (ASR) has seen significant improvements in recent years. However, ASR performance degrades rapidly when the acoustic environment, such as communication channels or noise backgrounds, differ from those of training data. In the missing feature approach to speech processing, the unreliable feature components are identified and reconstructed to overcome signal degradation and the mismatch of the acoustic environment. To reduce the model dependency, we investigate the matrix completion technique in missing feature reconstruction tasks. However, most of the matrix completion techniques require $a\ priori$ tuning parameters, e.g., target rank, which is hard to determine in practice. In this work, we propose a matrix completion method based on matrix factorization for the missing-feature reconstruction task, that does not require model training nor parameter tuning. Experiments show superior feature reconstruction performance and computational efficiency in both speech recognition and environmental sound classification tasks.**

***Index Terms*—Missing-feature reconstruction, matrix factorization, deep neural networks (DNNs), automatic speech recognition (ASR), environmental sound classification.**

## I. Introduction

THE RECENT progress on deep neural networks (DNN) has improved audio signal processing by leaps and bounds, including sound classification [1], speech recognition [2], speaker verification [3], and speech synthesis [4].

The quality of audio features plays a pivotal role in DNN-based audio processing systems. Many studies have been devoted to acoustic feature representation which is the key in signal acquisition and processing [5]–[13]. However, the performance of DNNs degrades severely under mismatched conditions. Moreover, the noise and distortion in the communication channels also cause corruption to speech contents. The robustness of acoustic features, therefore, remains an important research topic in audio information processing tasks.

Recently, DNN-based speech enhancement and restoration [14], [15] methods have been studied to tackle the mismatch condition and signal degradation problems. These DNN-based methods can effectively model the characteristic of both the original audio signals and noise, such that they can restore the audio signals with high-fidelity. However, to allow good coverage of different acoustic environments and noise types, they require a notoriously large amount of training data and a gigantic DNN model. Moreover, the remarkable performance of DNNs also comes at the cost of computational resources and storage space during deployment, which prevents the large-scale deployment to pervasive low-power mobile and internet-of-things (IoT) devices.

In another vein of research, to reduce the complexity of model training and the computational burden at runtime, the model-free signal processing methods are resorted to restoring the original audio content from degraded observations. Inspired by the conceptual framework of image inpainting, whereby to restore the missing pixels, the audio inpainting framework [16] is formalized to address the audio degradation problems that caused by impulsive noise, clicks to old recordings of scratched CDs, clipping by insufficient dynamic range and packet loss in cordless phones or voice over IP (VoIP) [16]–[19]. Within this framework, the distorted samples are regarded as missing and their location indices are assumed to be known. Owing to the advances in compressed sensing (CS) [20], [21], numerous audio inpainting methods based on the sparse representation have been devised, including [16], [17], [19]. These early studies focus on the situation where short-period, continuous audio segments are missing. By leveraging the sparse regression techniques, the missing segments can be recovered with high quality.

Different from these early studies on audio inpainting, a general distribution of missing components in the feature space is considered in this work. That is, the components of the spectrogram feature are missing randomly. This phenomenon comes naturally in a variety of scenarios, instances include non-uniform/compressive sampling during signal acquisition [22], and noise corruptions during the acoustic feature dictionary transmission and storage [23]. Therefore, successfully detecting and restoring the distorted feature representation under these scenarios can significantly improve the robustness of subsequent speech processing systems.

Manuscript received April 12, 2020; revised July 19, 2020, September 14, 2020, November 9, 2020, and November 11, 2020; accepted November 11, 2020. Date of publication November 16, 2020; date of current version January 29, 2021. This work was supported in part by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG-GC-2019-002) and (AISG Award No: AISG-100E-2018-006), and its National Robotics Programme under Grant 192 25 00054, and in part by RIE2020 Advanced Manufacturing and Engineering Programmatic under Grants A1687b0033, A18A2b0046, and I2001E0053. The work of Jibin Wu was supported by Zhejiang Lab under Grant 2019KC0AB02. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of Singapore National Research Foundation, the Agency for Science, Technology and Research (A*STAR), and Zhejiang Lab. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Nicki Holighaus. *(Q. Liu and J. Wu contributed to this work equally.) (Corresponding author: Jibin Wu.)*

The authors are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 119077, Singapore (e-mail: elelqi@nus.edu.sg; jibin.wu@u.nus.edu).

Digital Object Identifier 10.1109/JSTSP.2020.3038054

The missing-feature approach [24] has been introduced earlier to tackle this task by explicitly modeling the audio signals and noise, such that the original audio signal can be effectively restored from the degraded observations. As such, the performance of the missing-feature approach highly depends on the capability to accurately measure the noise characteristics, therefore, limiting the effectiveness of this approach, especially for challenging non-stationary noise. To remove the dependence on audio signal and noise modeling as in the missing-feature approach, we investigate the matrix completion techniques to restore the missing feature components. Similar to ours, the work in [23], applies the singular value thresholding (SVT) [25], [26] to enhance the reliability of speech features for noise-robust speaker identification.

The acoustic features typically exhibit spectro-temporal dynamics that can be represented as a matrix. Therefore, matrix completion techniques can be employed to restore the missing components in acoustic features. However, most of the existing matrix completion techniques require costly hyperparameter tuning that are difficult to determine in practice. For example, SVT [25], [26] and truncated nuclear norm regularization by the alternating direction method of multipliers (TNNR-ADMM) [27] are two well-known matrix completion methods that exhibit compelling recovery capabilities. Nevertheless, the SVT method may obtain sub-optimal performance in practical applications especially for real-world datasets with indeterminate rank information, which is due to the fact that the SVT method is sensitive to its hyperparameters. Additionally, the nuclear norm treats the singular values differently by adding them together, which may not be a good approximation to the rank function wherein all the nonzero singular values have equal contributions. As for the TNNR-ADMM method, its performance is vulnerable to the pre-specified rank information, which is usually hard to determine in practice due to the fact that the real audio signals have diversified spectral characteristics. Moreover, the rank information can be easily contaminated by the noise during transmission. In [28], to reduce the time and storage complexity, a rank-one matrix pursuit (R1MP) method based on the matrix format of orthogonal matching pursuit (OMP) is proposed for matrix completion problem, where any desired matrix is expressed as a linear combination of rank-one matrices generated by the singular value decomposition (SVD). Indeed, R1MP performs in a scalable manner and computationally efficient for large-scale datasets, yet at the cost of a performance overhead. Different from the proposed method, R1MP still requires to know the rank information in advance, and, hence, is not parameter tuning-free.

To overcome the above issues, we propose a novel parameter tuning-free matrix completion method, which does not require a pre-determined target rank information. Motivated by the matrix factorization techniques, the hard-to-handle rank function is approximated in the proposed method based on the rationale of alternately minimizing a convex function over one variable while fixing the others. Herein, different from the existing matrix factorization-based approaches, the decision matrix is decomposed as the sum of a set of rank-one matrices. Then, the iteratively reweighted least squares (IRLS) method is employed to solve the resulting simplified-version of weighted least squares (LS) problem with non-zeros. The main contributions of this work are summarized as follows:

- A parameter tuning-free missing-feature reconstruction method is proposed based on the matrix factorization using $\ell_p$-norm regression $(p \geq 1)$, where alternating minimization is employed to solve the resulting nonconvex optimization problem in a block coordinate descent (BCD) manner. Moreover, this method is also model training-free and computationally efficient in terms of CPU runtime.
- The proposed method has been compared thoroughly with well-known matrix completion techniques, i.e., SVT [25], [26], TNNR-ADMM [27], and R1MP [28], to reconstruct the missing features. A systematic analysis for clean signals and a wide range of noisy scenarios are provided.
- Numerical simulations are conducted for both speech and environmental sounds, that exhibit distinctive spectral characteristics, with applications to DNN-based speech recognition and environmental sound classification.

The rest of the paper is organized as follows. In Section II, the missing feature reconstruction task that studied in this work is first formulated. In Section III, the proposed parameter-tuning matrix completion method is developed to solve the missing feature reconstruction task and the preliminaries of the matrix completion are also provided. Section IV presents the numerical simulations on speech and environmental sound signals under both clean and noisy conditions with applications to speech recognition and environmental sound classification. Finally, conclusions are drawn in Section V.

## II. Problem Formulation

In speech and environmental sound recognition systems, audio signals are usually transformed into the spectro-temporal domain by applying Short-time Fourier Transform (STFT) to the subsampled frames. The resulting power spectrum is further processed by a mel-scaled or other perceptually motivated filter banks to extract a low-dimensional and discriminative feature representation, which we referred to as the spectrogram feature in this paper. As shown in Fig. 1, the two-dimensional spectrogram feature can be represented as a matrix. In the real acoustic environment, the clean audio signal is usually corrupted by noise and the resulting spectrogram feature $\mathbf{M}$ is approximately equal to the sum of the clean signal and the noise:[1]

$$\mathbf{M}(b,d) = \mathbf{X}(b,d) + \mathbf{N}(b,d) \tag{1}$$

where $\mathbf{X}(b,d)$ and $\mathbf{N}(b,d)$ denote the spectrogram component of the $d$-th frequency band of the $b$-th frame for the clean audio and the noise, respectively. In this work, we consider additive impulse noises as the major corruption to the clean audio signals, and such corruption is unevenly distributed across

[1]The finite-sized samples of uncorrelated processes are rarely perfectly orthogonal, and therefore the power spectra of two uncorrelated processes do not simply add within any particular analysis frame. An additional factor of $2\sqrt{\mathbf{X}(b,d)\mathbf{N}(b,d)}\cos(\theta)$ must be included in Equation (1), where $\theta$ is the angle between the $d$-th term of the complex spectra of the speech and the noise. However, this term is usually small and hence can be ignored [24].

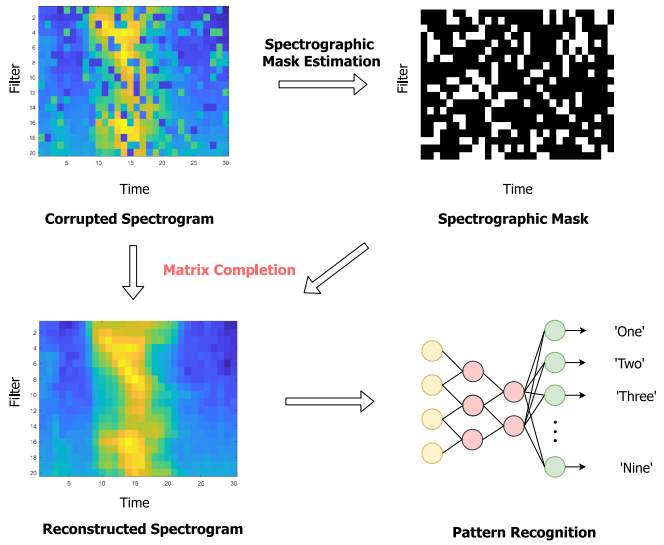


Fig. 1. Illustration of the proposed feature reconstruction system, wherein the matrix completion technique is employed to reconstruct the features from the corrupted spectrogram and the spectrographic mask. The reconstructed features are taken as input to the DNN-based classifier for pattern recognition tasks.

the spectrogram. More realistic environmental noises are also considered in Section IV-E. As shown in Fig. 1, for example, the resulting binary spectrographic mask $\Omega$ that distinguish reliable and unreliable components and the corrupted spectrogram are used jointly to reconstruct the clean spectrogram, known as the feature-vector imputation. In this work, we focus on reconstructing the clean spectrogram from noisy measurements and develop a novel parameter tuning-free matrix completion technique that will be introduced in the following section. Here, we assume the spectrographic mask is estimated or provided beforehand. It is worth noting that, independent from the matrix completion techniques studied in this paper, the effective spectrographic mask estimation is another important research topic that deserves more attention [29], [30].

## III. Main Contributions

### A. Preliminaries of Matrix Completion

For the ease of notation, we denote $\mathbf{M}(b,d)$ as $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$, whose compact SVD is given by $\mathbf{M} = \bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^T := \sum_{i \in \mathbb{N}_r^+} \sigma_i(\mathbf{M}) \bar{\mathbf{u}}_i \bar{\mathbf{v}}_i^T$ with column and row subspaces respectively being denoted as $\bar{\mathbf{U}}$ and $\bar{\mathbf{V}}$. They are spanned by the sets $\{\bar{\mathbf{u}}_i \in \mathbb{R}^{n_1 \times 1}\}_{i \in \mathbb{N}_r^+}$ and $\{\bar{\mathbf{v}}_i \in \mathbb{R}^{n_2 \times 1}\}_{i \in \mathbb{N}_r^+}$, respectively, where $r$ represents the target rank of desired matrix. Let $\mathbf{M}_\Omega \in \mathbb{R}^{n_1 \times n_2}$ be a data matrix with missing entries where $\Omega$ is a subset of the complete set of entries $[n_1] \times [n_2]$, with $[n]$ being the list $\{1, \ldots, n\}$. In this work, $\Omega$ is also referred to as the binary spectrographic mask. Throughout the paper, the subscript $(\cdot)_\Omega$ represents the projection on the known entries. The $(i,j)$ entry of $\mathbf{M}_\Omega$, denoted by $[\mathbf{M}_\Omega]_{ij}$, can be written as:

$$[\mathbf{M}_\Omega]_{ij} = \begin{cases} \mathbf{M}_{ij}, & \text{if } (i,j) \in \Omega, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The task of matrix completion is to find a matrix $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ given incomplete observation $\mathbf{M}_\Omega$ by incorporating the low-rank information. Mathematically, it is formulated as a rank minimization problem:

$$\min_{\mathbf{X}} \ \text{rank}(\mathbf{X}) \quad \text{s.t. } [\mathbf{X}]_{ij} = [\mathbf{M}]_{ij},\ (i,j) \in \Omega. \tag{3}$$

Unfortunately, the rank minimization is a combinatorial problem known to be NP-hard in general. To handle this issue, nuclear norm minimization is proposed to relax rank minimization [26], which is analogous to the strategy of approximation of $\ell_0$-norm replaced by $\ell_1$-norm in compressed sensing [31]. The nuclear norm is the convex envelope of rank. On the basis of that, Candés and Tao prove that one can deal with the matrix completion problem via minimizing nuclear norm with a high probability [32]. Therefore, it results in a nuclear norm optimization problem:

$$\min_{\mathbf{X}} \ ||\mathbf{X}||_* \quad \text{s.t. } [\mathbf{X}]_{ij} = [\mathbf{M}]_{ij},\ (i,j) \in \Omega \tag{4}$$

where $||\mathbf{X}||_* := \sum_r \sigma_r$ denotes the nuclear norm of matrix $\mathbf{X}$.

In [33], the *incoherence property* has been introduced to derive conditions under which the solution of (4) coincides with $\mathbf{M}$, where two assumptions with respect to the subspace $\bar{\mathbf{U}}$ and $\bar{\mathbf{V}}$ are stated as follows:

- $\max\{\rho(\bar{\mathbf{U}}), \rho(\bar{\mathbf{V}}))\} \leq \rho_0$, $\rho_0 \in \mathbb{R}^+$ is a constant;
- $||\sum_{i \in \mathbb{N}_r^+} \bar{\mathbf{u}}_i \bar{\mathbf{v}}_i^T||_\infty \leq \rho_1 \sqrt{\frac{r}{n_1 n_2}}$, $\rho_1 \in \mathbb{R}^+$ is a constant.

Herein, $\rho(\bar{\mathbf{X}}) := \frac{n}{r} \text{sup}_{i \in \mathbb{N}_n^+} ||\text{P}_{\bar{\mathbf{X}}} \cdot \mathbf{e}_i||_2^2$, where $\text{P}_{\bar{\mathbf{X}}} := \bar{\mathbf{X}}\bar{\mathbf{X}}^T$ is the orthogonal projection onto a general subspace $\bar{\mathbf{X}}$ (viz. $\bar{\mathbf{U}}$ and $\bar{\mathbf{V}}$) and $\{\mathbf{e}_i\}_{i \in \mathbb{N}_n^+}$ are the standard basis. Meanwhile, $\rho(\bar{\mathbf{U}})$ is also expanded to $\rho(\bar{\mathbf{U}}) = \frac{n}{r} \text{sup}_{i \in \mathbb{N}_n^+} \sum_{j \in \mathbb{N}_r^+} |\bar{\mathbf{U}}(i,j)|^2 \in [1, \frac{n}{r}]$. If $\rho_0$ and $\rho_1$ associated with the singular vectors of $\mathbf{M}$ are known to be bounded and sufficiently small, it is proved in [33] that $\mathcal{O}(\overline{n}^{6/5} r \log \overline{n})$ randomly sampled elements with $r = \mathcal{O}(n^{1/5})$ and $(\overline{n} := \max\{n_1, n_2\})$ suffices to exactly complete $\mathbf{M}$ with high probability. Therefore, it is shown that the error term $||\mathbf{M} - \hat{\mathbf{M}}||_F$ is bounded by [34]:

$$||\mathbf{M} - \hat{\mathbf{M}}||_F \leq 4 \sqrt{\frac{1}{k}(2+k)\underline{n}\xi} + 2\xi \tag{5}$$

in which $\underline{n} := \min\{n_1, n_2\}$, $\xi = \sqrt{(m + \sqrt{8m})\sigma^2}$ and $k = \frac{m}{n_1 n_2}$. Additionally, $\mathbf{X} = \hat{\mathbf{M}}$ is the estimate of $\mathbf{M}$ from (4). In the presence of noise, the resulting noise matrix is bounded by $||\mathcal{P}(\mathbf{M} - \mathbf{X})||_F \leq \xi$, where $\mathcal{P}(\mathbf{M})$ represents an element-wise sampling operator. When $m \geq C\rho^2 \overline{n} r \log^6 \overline{n}$ with a positive numerical constant $C$, the minimizer of problem (4) is unique and equals to $\mathbf{M}$ with probability at least $1 - \overline{n}^{-3}$.

A variety of state-of-the-art approaches have been proposed to deal with the nuclear norm optimization problem in (4), including SVT, TNNR-ADMM, and $\ell_p$-reg [35], to name just a few. Nevertheless, most of them require the rank information and full SVD operation. It is impractical to know *a priori* rank information. Moreover, the full SVD operation will result in a

demanding computational complexity for data-driven systems with large-scale datasets.

### B. The Proposed Method

To reduce the computational burden of performing a full SVD operation, matrix factorization has been employed, corresponding to the following optimization problem [35]:

$$\min_{\mathbf{U},\mathbf{V}} ||\mathbf{M}_\Omega - (\mathbf{UV})_\Omega||_F^2 \tag{6}$$

where $\mathbf{U} \in \mathbb{R}^{n_1 \times r}$ and $\mathbf{V} \in \mathbb{R}^{r \times n_2}$ with $r$ being the rank of target matrix. Nevertheless, (6) is difficult to address because of its nonconvexity. One may solve the problem (6) in an iterative manner as a bi-convex problem [36], $\mathbf{U}$ and $\mathbf{V}$ are alternately minimized as follows:

$$\begin{cases} \mathbf{V}^{t+1} = \arg\min\limits_{\mathbf{V}} ||\mathbf{M}_\Omega - (\mathbf{U}^t\mathbf{V})_\Omega||_F^2 \\ \mathbf{U}^{t+1} = \arg\min\limits_{\mathbf{U}} ||\mathbf{M}_\Omega - (\mathbf{U}\mathbf{V}^{t+1})_\Omega||_F^2 \end{cases}. \tag{7}$$

Then, the target matrix can be computed by $\mathbf{M} = \mathbf{U}^{t+1}\mathbf{V}^{t+1}$ after determining $\mathbf{U}$ and $\mathbf{V}$, which is due to the fact that the low-rank property of $\mathbf{X}$ is automatically fulfilled with $a$ $priori$ rank in each iteration. Matrix factorization has been widely-used in the topics of recommender system [37], image restoration [35], [38], data mining and machine learning [39], direction-of-arrival (DOA) estimation [40], etc. However, it is worth noting that both of them require a pre-specified rank, which is challenging to determine in practice. Moreover, the model in (6), which employs the Frobenius norm, is not robust against the non-Gaussian noise.

To address that, we devise a simple yet efficient matrix completion method based on $\ell_p$-norm ($p \geq 1$), where the $\ell_p$-norm of matrix $\mathbf{E}$ is defined as $||[\mathbf{E}]_{ij}||_p^p = \sum_{i,j} |[\mathbf{E}]ij|^p$, $(i,j) \in \Omega$. The $\ell_p$-norm with $p \geq 1$ is convex. At first, we consider that the decision matrix $\mathbf{X}$ is decomposed as a summation of a set of rank-one matrices, that is:

$$\mathbf{X} = \sum_{\ell=1}^{q} \mathbf{X}_\ell \tag{8}$$

where $\mathbf{X}_\ell = \mathbf{u}_\ell \mathbf{v}_\ell^T$. Therefore, the rank information is unnecessary for our method, and we just need to find the optimal number of $q$. Note that $q$ is not equal to the desired rank of the approximated matrix $\mathbf{X}$. For a parameter tuning-free method, $q$ is automatically determined by the number of iterations when the proposed method converges. Combining with the Equation (8), the optimization problem in (6) is equivalently transformed into:

$$\min_{\mathbf{X}} \left|\left| \mathbf{M}_\Omega - \left( \sum_{\ell=1}^{q} \mathbf{X}_\ell \right)_\Omega \right|\right|_p^p, \; p \geq 1. \tag{9}$$

To be more specific, it results in the objective function $\mathcal{G}$, as shown in the following problem:

$$\mathcal{G}: \min_{\mathbf{u},\mathbf{v}} \left|\left| [\mathbf{M}]_{ij} - \left[ \sum_{\ell=1}^{q} \mathbf{u}_\ell \mathbf{v}_\ell^T \right]_{ij} \right|\right|_p^p, \; (i,j) \in \Omega, \; p \geq 1. \tag{10}$$

This formulation is motivated by the matrix factorization method. However, the difference between the proposed method and the existing non-negative matrix factorization (NMF) is that the existing NMF methods require to know the pre-specified rank, while ours employ a set of rank-one matrices in (8) and combine with matrix factorization methodology. Therefore, the proposed model is parameter tuning-free compared with the existing NMF methods. Let us define the block $\boldsymbol{\Upsilon} := (\boldsymbol{\Upsilon}_1, \cdots \boldsymbol{\Upsilon}_{2q}) = (\{\mathbf{u}\}_{\ell=1}^q, \{\mathbf{v}\}_{\ell=1}^q)$. From the analysis in [41], BCD method can converge to a critical point when the following conditions are satisfied:

$$\tilde{\mathcal{G}}_\iota \left( \boldsymbol{\Upsilon}_\iota^\ell | \boldsymbol{\Upsilon}_{-\iota}^{\ell-1} \right) = \mathcal{G} \left( \boldsymbol{\Upsilon}_\iota^\ell | \boldsymbol{\Upsilon}_{-\iota}^{\ell-1} \right) \tag{11}$$

$$\tilde{\mathcal{G}}_\iota \left( \boldsymbol{\Upsilon}_\iota^\ell | \boldsymbol{\Upsilon}_{-\iota} \right) \leq \mathcal{G} \left( \boldsymbol{\Upsilon}_\iota^\ell | \boldsymbol{\Upsilon}_{-\iota} \right), \forall \boldsymbol{\Upsilon}_{-\iota} \tag{12}$$

$$\nabla \tilde{\mathcal{G}}_\iota \left( \boldsymbol{\Upsilon}_\iota^\ell | \boldsymbol{\Upsilon}_{-\iota}^{\ell-1} \right) = \nabla \mathcal{G} \left( \boldsymbol{\Upsilon}_\iota^\ell | \boldsymbol{\Upsilon}_{-\iota}^{\ell-1} \right) \tag{13}$$

$$\tilde{\mathcal{G}}_\iota \left( \boldsymbol{\Upsilon}_\iota^\ell | \boldsymbol{\Upsilon}_{-\iota}^{\ell-1} \right) \text{ is continuous in } \boldsymbol{\Upsilon}. \tag{14}$$

where $\bar{\mathcal{G}}(\{\mathbf{u}\}_{\ell=1}^q, \{\mathbf{v}\}_{\ell=1}^q)$ is a surrogate function of $\mathcal{G}(\{\mathbf{u}\}_{\ell=1}^q, \{\mathbf{v}\}_{\ell=1}^q)$. Following the rationale of BCD method, we define the surrogate function as:

$$\bar{\mathcal{G}}: \min_{\mathbf{u}_q, \mathbf{v}_q} \left|\left| \mathbf{R}_q - \left[ \mathbf{u}_q \mathbf{v}_q^T \right]_{ij} \right|\right|_p^p, \; (i,j) \in \Omega, \; p \geq 1 \tag{15}$$

for each $q$-th iteration, where $\mathbf{R}_q := \mathbf{M}_\Omega - (\sum_{\ell=1}^{q-1} \mathbf{u}_\ell \mathbf{v}_\ell^T)_\Omega$ with $q \geq 2$ and $\mathbf{R}_1 = \mathbf{M}_\Omega$.

*Proposition1:* The surrogate function $\bar{\mathcal{G}}$ in (15) satisfies the conditions from (11) to (14).

*Proof:* See Appendix. ■

At the $\ell$-th iteration, variable $\boldsymbol{\Upsilon}_\iota, \iota = 1, \ldots, 2q$, is updated by solving the following problem:

$$\boldsymbol{\Upsilon}_\iota^\ell = \arg\min_{\boldsymbol{\Upsilon}_\iota} \tilde{\mathcal{G}}_\iota \left( \boldsymbol{\Upsilon}_\iota | \boldsymbol{\Upsilon}_{-\iota}^{\ell-1} \right) \tag{16}$$

with respect to block variable $\boldsymbol{\Upsilon}_\iota$, and $\boldsymbol{\Upsilon}_{-\iota}^{\ell-1}$ represents the rest of the variables obtained at the $(\ell-1)$-th iteration except for $\boldsymbol{\Upsilon}_\iota$. Instead of optimizing the original objective function, we alternatively optimize a surrogate function $\tilde{\mathcal{G}}_\iota(\boldsymbol{\Upsilon}_\iota | \boldsymbol{\Upsilon}_{-\iota}^{\ell-1})$ which satisfies certain requirements such that the original problem can be easily tackled.

Toward this end, we utilize the greedy pursuit manner to search for the best rank-one basis matrix of the current residual $\mathbf{R}_q$ and the IRLS method [42] is employed to tackle the problem (15) in a BCD manner [41]:

*1) Updating* $\mathbf{u}_q^\ell$*:* To be specific, based on the rationale of alternating minimization, we fix variable $\mathbf{v}$ and then optimize $\mathbf{u}$, resulting in:

$$\mathbf{u}_q^\ell = \arg\min_{\mathbf{u}_q} \left|\left| \mathbf{R}_q - \left[ \mathbf{u}_q (\mathbf{v}_q^{\ell-1})^T \right]_{ij} \right|\right|_p^p, \; (i,j) \in \Omega. \tag{17}$$

Support that $\mathbf{r}_i$ and $u_i$ are the $i$-th row of $\mathbf{R}_q$ and $i$-th entry of $\mathbf{u}_q$, respectively. As $\{\mathbf{r}_i\}_{i=1}^{n_1}$ are independent for each $u_i$, (17) is equivalent to tackling the following $n_1$ independent subproblems:

$$u_i{}^\ell = \arg\min_{u_i} \left|\left| \mathbf{r}_i - u_i (\mathbf{v}_i^{\ell-1})^T \right|\right|_p^p, \; (i,j) \in \Omega, p \geq 1. \tag{18}$$

Since $\mathbf{R}_q = \mathbf{M}_\Omega - (\sum_{\ell=1}^{q-1} \mathbf{u}_\ell \mathbf{v}_\ell^T)_\Omega$ and $\mathbf{r}_i = [\mathbf{R}_q]_{i\cdot}$ is the $i$-th row of $\mathbf{R}_q$, it is easily observed that the residual error in (18) is only affected by all non-zero elements in $\mathbf{r}_i$ and $u_i(\mathbf{v}_i^{\ell-1})^T$. Therefore, (18) is further simplified as:

$$u_i{}^\ell = \arg\min_{u_i} \left|\left|\ddot{\mathbf{r}}_i - u_i(\ddot{\mathbf{v}}_i^{\ell-1})^T\right|\right|_p^p,\ p \geq 1 \tag{19}$$

where $\ddot{\mathbf{r}}_i$ and $\ddot{\mathbf{v}}_i^{\ell-1}$ stand for the $\mathbf{r}_i$ and $\mathbf{v}_i^{\ell-1}$ with non-zero entries inside, respectively. The problem (19) with $p = 2$ is easily handled as it is a LS problem. To guarantee the solution of (19) to perform good performance, IRLS is utilized, which can provide global convergence.

*Case 1:* For $p = 2$, the solution of LS is reweighted by $w_i^t = 1/(\max\{\delta, |(\ddot{\mathbf{r}}_i - (u_i^\ell)^t \ddot{\mathbf{v}}_i^{\ell-1})^T|\})$, and $(u_i^\ell)^t$ is initialized at $(u_i^\ell)^0 = (\ddot{\mathbf{v}}_i^{\ell-1})^T \ddot{\mathbf{r}}_i^T / ((\ddot{\mathbf{v}}_i^{\ell-1})^T \ddot{\mathbf{v}}_i^{\ell-1})$, where $\delta$ is a small regularization value, e.g., $10^{-3}$, to avoid dividing by zero.

*Case 2:* In the optimization problem of IRLS at $1 \leq p < 2$, it addresses a weighted LS problem as follows:

$$(u_i{}^\ell)^{t+1} = \arg\min_{(u_i)^\ell} \left|\left|(\ddot{\mathbf{r}}_i - (u_i^\ell)^t \ddot{\mathbf{v}}_i^{\ell-1})^T \mathbf{w}^t\right|\right|_2^2 \tag{20}$$

where $w_i^t = 1/(\epsilon + |(\ddot{\mathbf{r}}_i - (u_i^\ell)^t \ddot{\mathbf{v}}_i^{\ell-1})^T|^2)^{\frac{1-p/2}{2}},\ \ 1 \leq p < 2$. Additionally, $\epsilon$ is a small positive parameter, e.g., $10^{-8}$.

Towards this end, the solution of block $\{\mathbf{u}\}_{\ell=1}^q$ has been achieved.

*2) Updating* $\mathbf{v}_q^\ell$*:* Taking in similar manner, we update $\mathbf{v}$ by fixing $\mathbf{u}$.

$$\mathbf{v}_q^\ell = \arg\min_{\mathbf{v}_q} \left|\left|\mathbf{R}_q - \left[\mathbf{u}_q^\ell \mathbf{v}_q^T\right]_{ij}\right|\right|_p^p,\ (i,j) \in \Omega,\ p \geq 1. \tag{21}$$

Then, we have its simplified version with non-zeros, namely:

$$(v_j{}^\ell)^{t+1} = \arg\min_{(v_j)^\ell} \left|\left|(\mathbf{w}^t)^T(\ddot{\mathbf{r}}_j - (\ddot{\mathbf{u}}_j^\ell)^{t+1}(v_j^\ell)^t)\right|\right|_2^2,\ (i,j) \in \Omega \tag{22}$$

where $\ddot{\mathbf{r}}_j$ and $\ddot{\mathbf{u}}_j^\ell$ denote the $j$-th column of $\mathbf{R}_q$ and $\mathbf{u}_j^\ell$ after removing missing entries, respectively. The pseudocode of the proposed method is summarized in Algorithm 1.

*Remark 1:* Different from the existing matrix completion approaches with *a priori* rank information, the proposed method is parameter tuning-free. Only $q$ is not available in (8). However, it is worth noting that the selection of $q$ is automatically determined by the number of iterations when the proposed method converges.

*Remark 2:* As the update rules of the proposed method, viz., (20) and (22), they are similar with the vertex least squares in *graph* signal processing, it has been demonstrated for $p = 2$ and can be extended for $1 \leq p < 2$ that we have $||\mathbf{R}_q - [\mathbf{u}_q \mathbf{v}_q^T]_{ij}||_p^p \leq \xi,\ (i,j) \in \Omega$, with $\mathcal{O}(\overline{n}^\gamma \log \overline{n})\ (\gamma > 0)$ number of iterations, under the assumptions of that the bipartite undirected graph on the vertex set $\mathcal{S} = \mathcal{S}_R \cup \mathcal{S}_C$ is connected and has $c\log\overline{n}$ diameter for some fixed constant $c$ and maximum degree $\Delta(\overline{n})$. $\mathcal{S}_R$ and $\mathcal{S}_C$ stand for the sets of rows and columns of $\mathbf{M}_\Omega$, respectively, and the maximum degree $\Delta$ denotes the maximum number of neighbors among all nodes of the graph. Hence, the convergence of the proposed method is guaranteed, details referred to [43].

Algorithm 1

**Require:** $\mathbf{M}_\Omega$ and $\Omega$.
  **Initialize:** Randomize $\mathbf{v}_0$ and $\mathbf{R}_1 = \mathbf{M}_\Omega$
  **for** $t = 1, 2, \cdots$ **do**
    **for** $q = 1, 2, \cdots$ **do**
      $\widetilde{\mathcal{G}}:\ \min_{\mathbf{u}_q, \mathbf{v}_q} ||\mathbf{R}_q - [\mathbf{u}_q \mathbf{v}_q^T]_{ij}||_p^p,\ (i,j) \in \Omega.$
      1) Updating $\mathbf{u}_q^\ell$ via (17) and (20)
      2) Updating $\mathbf{v}_q^\ell$ via (21) and (22)
      3) $\mathbf{R}_{q+1} = \mathbf{R}_q - (\mathbf{u}_\mathbf{q} \mathbf{v}_q^T)_\Omega$
    **end for**
    **Stop** if the stopping criterion is satisfied.
  **end for**
**Ensure:** $\mathbf{X} = \sum_{\ell=1}^q \mathbf{u}_\ell \mathbf{v}_\ell^T$

## IV. Experimental Evaluation

In this section, we evaluate the performance of the proposed parameter tuning-free matrix completion technique in missing-feature reconstruction tasks. We first introduce the experimental setups, which are designed to test the effectiveness of the proposed technique in restoring distorted spectrograms of speech and environmental sound signals. Then, we investigate the applicability and robustness of the proposed method to the additive Gaussian noise that is a common model for the communication channel and environment. Finally, we evaluate the effectiveness of the proposed technique under more realistic noisy environments. The source code is public available.[2]

### A. Experimental Details

*1) Datasets:* To provide a comprehensive evaluation of audio signals with distinctive spectral-temporal characteristics, we evaluate the proposed matrix completion technique on both speech and environmental sound reconstruction tasks.

The speech samples are taken from the TIDIGITS [44] dataset that managed by the Linguistic Data Consortium. This dataset consists of reading digit sequences of variable lengths from 21 dialectical regions of the United States. We use the subset of isolated spoken digits from 11 classes (i.e., 'zero' to 'nine' and 'oh'), which consists of 2,464 train and 2,486 test utterances. The utterances are spoken by 111 male and 114 female speakers at a sampling rate of 20 kHz. This dataset has been used as a common benchmark for different speech recognition algorithms.

The environmental sound samples are taken from the Real World Computing Partnership (RWCP) [45] dataset. This dataset consists of high-fidelity natural sound samples recorded in the real acoustic environment at a sampling rate of 16 kHz. Following the experimental setup of [46], we use the same 10 environmental sound classes, including 'cymbals,' 'horn,' 'phone4,' 'bells5,' 'kara,' 'bottle1,' 'buzzer,' 'metal15,' 'whistle1' and 'ring'. We randomly selected 40 samples from each class, wherein 20 samples are used to train a DNN-based sound classifier and the rest are used for evaluation.

[2][Online]. Available: https://github.com/deepspike/audioImpainting

The utterances are segmented into frames of 100 ms and 50 ms for the RWCP and TIDIGITS datasets respectively, with 50% overlap between the neighboring frames. The 20-dimensional mel-scaled filter bank (FBANK) features are extracted before input to the DNN-based classifier. The feature masking and reconstruction studies are performed on the FBANK features. To ensure a consistent temporal dimension, we first determine the maximal time duration $T_{max}$ (total number of frames) from the training set of each dataset. Then, we zero-pad all of the training and testing samples along the time dimension to $T_{max}$. For those testing samples that may have a time duration longer than $T_{max}$ (rarely happen in the testing set), we discard the rest of the frames beyond $T_{max}$. Specifically, the input feature dimensions for the RWCP and TIDIGITS datasets are $20 \times 61$ and $20 \times 101$, respectively.

As mentioned earlier, different sound classes exhibit distinctive spectral-temporal dynamics. Hence, we use a convolutional neural network that is inspired by the AlexNet [47] to classify different speech and environmental sound samples. The network has a structure of 24c3s1-48c3s2-48c3s1-96c3s2-128c3s2-256-$N_{class}$, wherein the numbers before and after 'c' refers to the number of convolution kernels and the corresponding kernel size (same size for both frequency and time dimensions) at each layer. The number after 's' refers to the stride of the convolution operation at each layer. Instead of using pooling layers to reduce the dimensionality of feature maps, we apply convolution kernels at a stride of 2 at layer 2, 4, 5 to achieve the dimensionality reduction.

To prevent overfitting, we add dropout layers with a probability of 15% after the last convolutional layer and the first fully-connected layer. The models are trained with the Adam optimizer and the cross-entropy loss function for 50 epochs. We initialize the learning rate at a value of 0.01. A batch size of 16 and 64 are used for the RWCP and TIDIGITS dataset, respectively.

To simulate impulsive noise which causes spectrogram feature components missing, we generate random binary masks at different probabilities, ranging from 0% to 40% at an interval of 10%, to corrupt the original contents of the spectrogram features. We compare the proposed parameter tuning-free technique with other existing matrix completion techniques that require extensive parameter tuning, including SVT [23], [26], TNNR-ADMM [27], and R1MP [28]. Since rank cannot be fixed in SVT, the thresholding parameter $\tau$ is chosen as the tuning parameter to obtain the desired rank solution, which here is set as proposed in [25], [26]. For a fair comparison with the proposed method under noisy conditions, we use the noisy version of SVT as proposed in [26]. It is also worth noting that the TNNR-ADMM method is designed for eliminating Gaussian noise, however, its performance is dependent on the pre-specified rank of the data. In this work, the rank is determined by the number of the largest singular values,[3] viz., rank $= 2$ for RWCP and rank $= 3$ for TIDIGITS datasets, respectively. In addition, the penalty parameter $\beta$ is equal to 0.01, and the remaining parameters are chosen as in [27]. Similar to our method, R1MP is developed from a set of rank-one matrices, but R1MP requires to pre-specify the rank. We use the same rank as for TNNR-ADMM. For the proposed method, we use $p = 2$ for the case of Gaussian noise, and $p = 1.2$ for non-Gaussian noises. These values were determined empirically.

*2) Evaluation Metrics:* We use the signal-to-noise ratio (SNR) to evaluate the quality of the reconstructed spectrogram features on the test set of the two datasets. Moreover, to study the interaction between the spectrogram feature and DNN-based classifier for sound and speech recognition tasks, we evaluate the classifier performance on both the corrupted and reconstructed features. Without any prior knowledge about the feature degradation, we train these classifiers with clean spectrogram features. We report the classification accuracies for both corrupted and reconstructed features under different mask ratios. The experimental results are summarised from 5 independent runs. Similarly, to compare the computational efficiency of different matrix completion techniques, we calculate the average CPU time required per sample across the 5 independent runs.

*3) Spectrographic Mask Estimation:* To study the effectiveness of the proposed matrix completion technique in reconstructing spectrogram features corrupted by the impulsive noises (Sections IV-B and IV-C), we use oracle masks that can be easily determined from the corrupted spectrograms.

To further evaluate the robustness of the proposed technique under other noisy scenarios, both Gaussian (Section IV-D) and non-Gaussian (Section IV-E), we follow the mask estimation method introduced in [24], whereby we estimate the mask for unreliable entries based on the negative energy criterion. Specifically, we assume the first frame of each utterance to be the region of silence and initialize the noise power spectra to the power spectrum of this frame. The noise spectra are further estimated recursively for subsequent frames. Let $\mathbf{M}(b,d)$ and $\hat{\mathbf{N}}(b,d)$ represent the power spectra of the observed sound signal and the estimated noise respectively, wherein $b$ and $d$ refer to the index of the frame and the frequency band. The noise spectra $\hat{\mathbf{N}}(b,d)$ can be obtained recursively as:

$$\hat{\mathbf{N}}(b,d) = \begin{cases} (1-\lambda)\hat{\mathbf{N}}(b-1,d) + \lambda \mathbf{M}(b,d), \\ \text{if } \mathbf{M}(b,d) < \beta \hat{\mathbf{N}}(b-1,d); \\ \hat{\mathbf{N}}(b-1,d), \text{ otherwise.} \end{cases} \tag{23}$$

where $\lambda$ and $\beta$ are set to 0.95 and 2 respectively following the hyperparameters setting advised in [24] without any further tuning. Then, the unreliable spectral components are identified based on the negative energy criterion in which the component is classified as unreliable if $|\mathbf{M}(b,d)| \leq |\hat{\mathbf{N}}(b,d)|$.

### B. Missing-Feature Reconstruction for Environmental Sound

Table I provides the results for the feature reconstruction on the RWCP environmental sound dataset with oracle mask. It is obvious that all the matrix completion techniques are competent at restoring the missing features and achieve high SNRs across

[3]Assume the singular values are in a descending order, that is, $\lambda_i > \lambda_{i+1}$. The rank $r$ is determined as the first index that satisfying $\sum_1^r \lambda_i / \sum_1^n \lambda_i > \phi$, where $n$ is the total number of singular values. In this work, we set $\phi$ to a value of 0.95.

TABLE I
SPECTROGRAM RECONSTRUCTION RESULTS UNDER DIFFERENT MASK RATIOS. THE RESULTS ARE OBTAINED FROM FIVE INDEPENDENT SIMULATIONS

| Task | Mask Ratio (%) | Reconstruction SNR (dB) | | | |
|---|---|---|---|---|---|
| | | This Work | TNNR-ADMM | SVT | R1MP |
| Environmental Sound Classification | 0 | - | - | - | - |
| | 10 | **89.96 ± 1.44** | 81.88±0.69 | 77.84±0.25 | 50.75±0.34 |
| | 20 | **81.63 ± 0.26** | 70.28±0.64 | 74.68±0.06 | 43.80±0.24 |
| | 30 | **74.42±1.03** | 63.57±0.30 | 71.53±0.07 | 38.66±0.29 |
| | 40 | 64.21±1.41 | 56.46±0.52 | **68.07±0.07** | 33.55±0.23 |
| Speech Recognition | 0 | - | - | - | - |
| | 10 | **77.60 ± 0.13** | 77.03±0.07 | 67.08±0.02 | 44.69±0.21 |
| | 20 | **70.89 ± 0.06** | 67.25±0.13 | 64.68±0.02 | 39.99±0.36 |
| | 30 | **64.67 ± 0.07** | 61.32±0.09 | 62.08±0.04 | 36.78±0.21 |
| | 40 | 58.05±0.06 | 55.32±0.15 | **59.13 ± 0.02** | 33.05±0.32 |

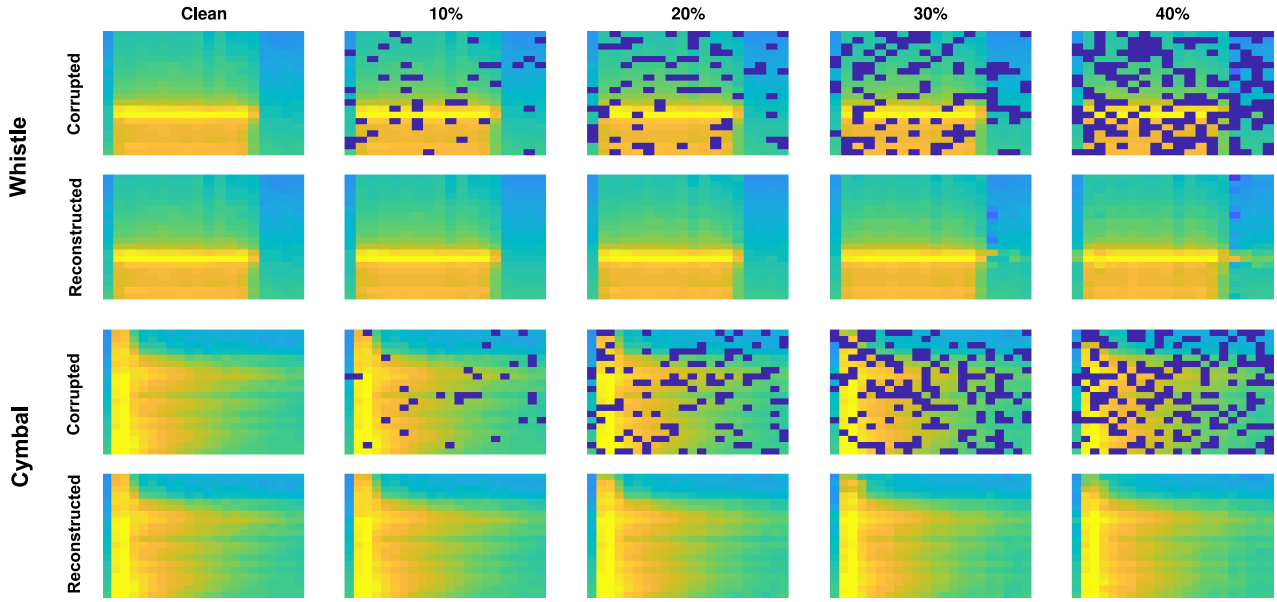


Fig. 2. Illustration of the missing feature reconstruction on the RWCP dataset. The dark blue squares refer to location where the orignal features are masked. Our proposed matrix compeletion technique can restore the missing features with high quality across different noise levels.

different corruption levels. Among the four matrix completion techniques studied in this work, our proposed technique consistently outperforms the TNNR-ADMM, SVT and R1MP methods, except for the challenging scenario with a 40% mask ratio. Herein, the mask ratio is defined as the percentage of the spectral components that are corrupted by noise and considered as unreliable. That is, $1 - \|\Omega\|/(|n1| * |n2|)$, where $\|\Omega\|$ denotes the number of non-zero elements in $\Omega$. To allow a better qualitative evaluation of the proposed matrix completion technique, we provide two examples from the 'whistle' and 'cymbal' classes as shown in Fig. 2. The reconstructed spectrograms exhibit high similarities to the original clean spectrograms across different corruption levels, suggesting the proposed technique can effectively compensate for missing features.

As the classification results given in Table II, the DNN-based classifier is highly sensitive to the masking noise, and the classification accuracy drops substantially from 99.40% to 35.70% when only 10% of the features are corrupted. In contrast, with the proposed method, the DNN-based classifier can still maintain a high classification accuracy after feature reconstruction and achieves a mean accuracy of 91.80% even under a corruption level of 40%. This promising result again highlights the effectiveness of the proposed matrix completion technique, which significantly improves the robustness of sound classification systems to missing features.

In addition to the superior feature reconstruction capability, our proposed matrix completion technique is highly efficient compared to the TNNR-ADMM and SVT methods. This can be explained by the fact that the proposed method does not require the time-consuming hyperparameter search. As shown in Fig. 4, under the mask ratio of 10%, our proposed method achieves a speedup of 2.0 and 2.4 times compared to the TNNR-ADMM and SVT methods respectively. It is worth noting that our required computation time reduces with a growing level of corruption, this is due to the fact that less non-zero values are involved in the process of computation according to the

TABLE II
CLASSIFICATION RESULTS OF NEURAL NETWORK CLASSIFIERS UNDER DIFFERENT MASK RATIOS. THE RESULTS ARE OBTAINED FROM FIVE INDEPENDENT SIMULATIONS

| Task | Mask Ratio (%) | Classification Accuracy (%) | | | | |
|---|---|---|---|---|---|---|
| | | Corrupted | Reconstructed | | | |
| | | | This Work | TNNR-ADMM | SVT | R1MP |
| Environmental Sound Classification | 0 | 99.40±0.55 | - | - | - | - |
| | 10 | 35.70±2.64 | **98.80 ± 0.27** | 98.10±0.55 | 97.60±0.42 | 93.00±2.03 |
| | 20 | 33.50±2.62 | **98.10 ± 0.65** | 96.00±0.71 | 96.80±0.27 | 88.00±2.55 |
| | 30 | 29.20±4.44 | **96.20 ± 0.27** | 94.10±0.82 | 95.60±0.42 | 82.20±4.70 |
| | 40 | 26.70±2.66 | 91.80±0.67 | 91.40±0.42 | **92.60 ± 0.41** | 74.90±6.67 |
| Speech Recognition | 0 | 98.77±0.34 | - | - | - | - |
| | 10 | 23.72±0.13 | **98.66 ± 0.13** | 98.19±0.14 | 97.89±0.38 | 87.37±2.46 |
| | 20 | 13.58±0.22 | **98.27 ± 0.22** | 97.57±0.18 | 97.45±0.36 | 79.56±3.19 |
| | 30 | 11.42±0.19 | **97.54 ± 0.19** | 96.08±0.37 | 96.83±0.47 | 75.89±2.58 |
| | 40 | 10.67±0.19 | **95.79 ± 0.19** | 92.30±0.70 | 95.37±0.34 | 68.27±4.34 |

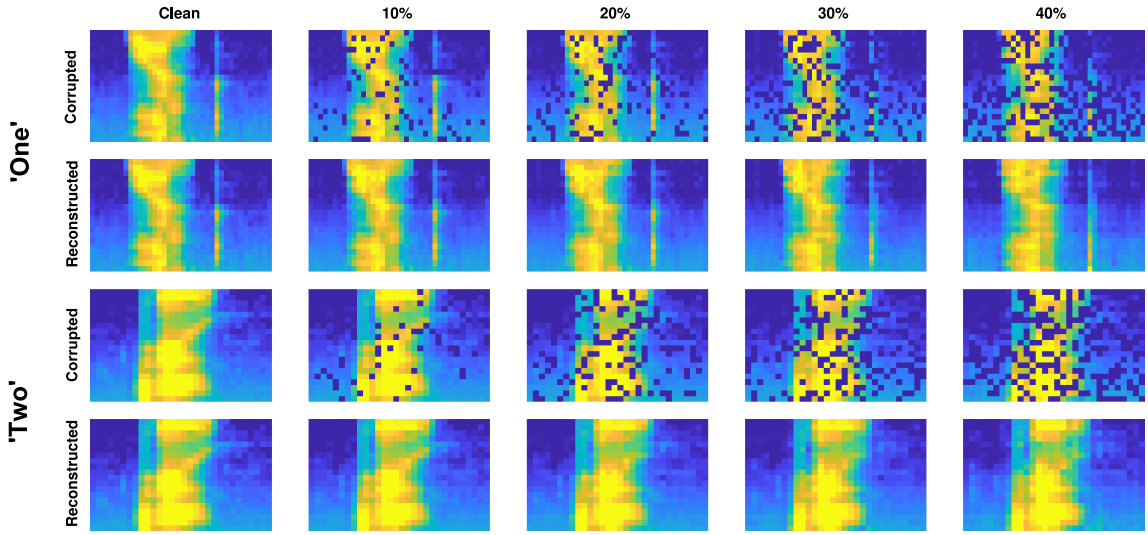


Fig. 3. Illustration of the missing feature reconstruction on the TIDIGITS dataset.The dark blue squares refer to location where the orignal features are masked. Our proposed matrix compeletion technique can restore the missing features with high quality across different noise levels.

solutions of Equations (20) and (22). In contrast, the computation time remains relatively stable for the TNNR-ADMM and SVT methods across different corruption levels. In the SVT, since the singular values exceeding the thresholding parameter $\tau$ and the corresponding singular vectors are required to compute per iteration, it is computationally demanding for handling large-scale dataset. We note that the R1MP converges at a speed that almost two orders of magnitude faster than the other methods in terms of the CPU time. The reasons are twofold. The most time-consuming step of the R1MP method is to compute the top singular vector pair of a sparse matrix, with only $\mathcal{O}(|\Omega|)$ operations at each iteration. Furthermore, an economic weight updating rule is exploited to avoid to store all rank-one matrices in the current OMP basis set. Nonetheless, because of these two reasons, its feature reconstruction performance is significantly inferior to other methods.

### C. Missing-Feature Reconstruction for Speech

As shown in Fig. 3, the proposed matrix completion technique allows a high-fidelity reconstruction of the corrupted spectrogram features of speech signals. This observation can be explained by the high SNRs given in Table I, where the SNRs remain higher than 58.05 dB under different corruption levels. Moreover, our proposed technique achieves superior reconstruction results and consistently outperforms other matrix completion techniques, except for the severely degraded scenario with a mask ratio of 40% where the SVT method achieves a slightly better result.

Similar to the observation in the environmental sound classification task, we notice that the DNN-based speech recognizer is also highly sensitive to the masking noise as the classification accuracies drop rapidly with an increasing amount of feature corruption. As the results are given in Table II, the proposed matrix completion technique effectively alleviates the effect of missing features and the classification accuracy degradation remains less than 3% for the masking ratio up to 40%. Notably, the classification accuracy can still maintain at 95.79% under the challenging scenario where 40% of the spectrogram features are masked.

In terms of the computational time, our proposed method increases substantially to 0.316 seconds (at a masking ratio

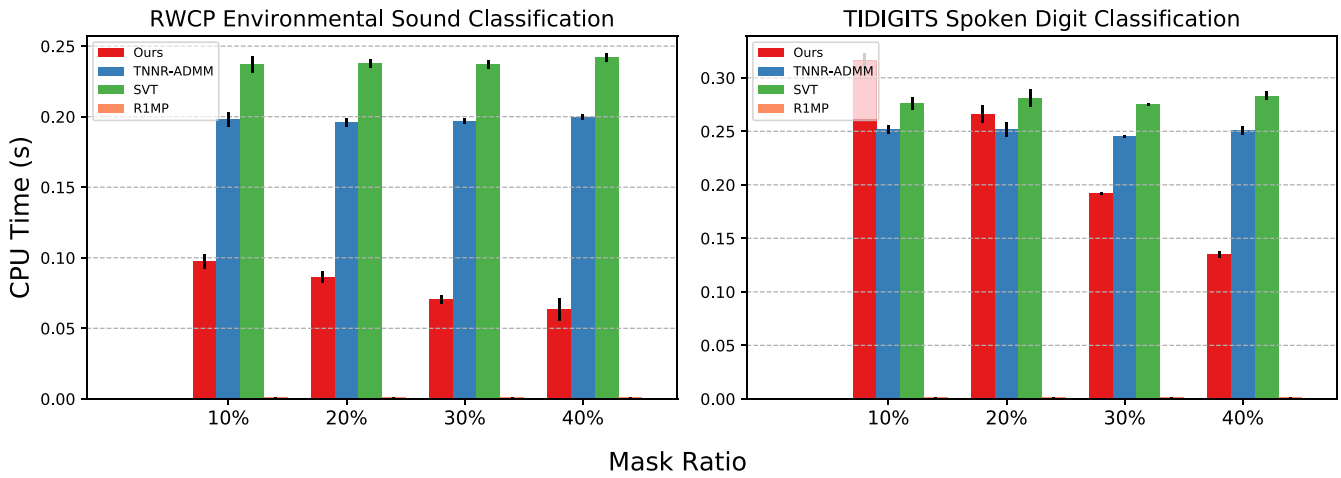


Fig. 4. Compare the required CPU time for missing feature reconstruction tasks under different corruption levels. For clarity, the results of different matrix completion techniques are color-coded. Note that the R1MP method requires around two order of magnitude less CPU times than the rest of the methods. Although our method is guaranteed to converge, it requires more iterations to obtain the optimal $q$ on the TIDIGITS dataset.

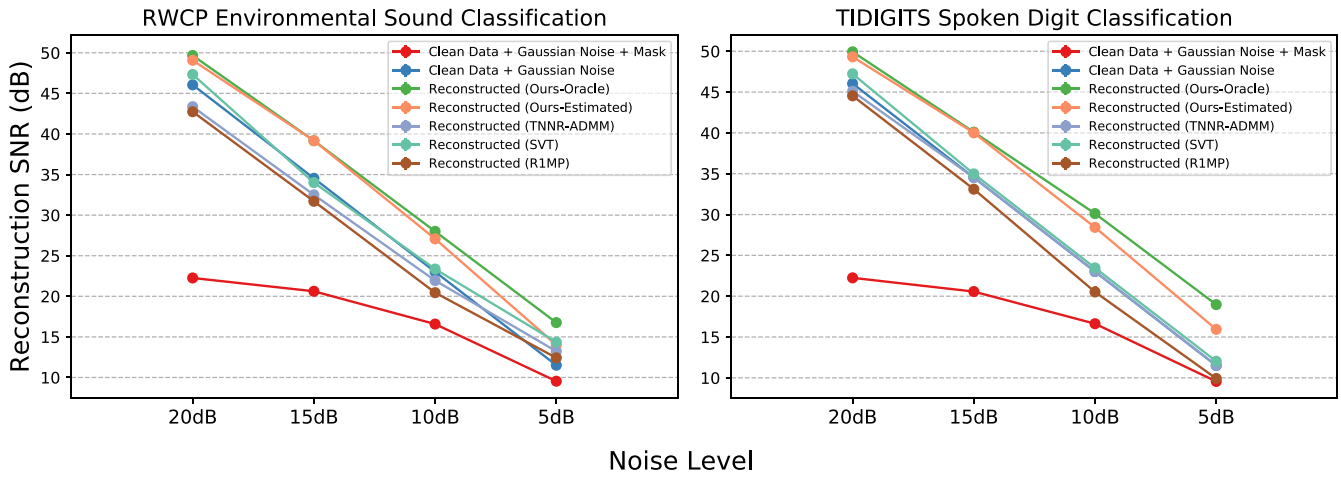


Fig. 5. Compare the feature reconstruction performance of different matrix completion techniques under varying Gaussian noise levels.

of 10%) on the TIDIGITS dataset, which is 3.2x longer than that required for the RWCP dataset. A thorough investigation reveals that the average number of frames increases from 20 (RWCP) to 38 (TIDIGITS), which can partially account for the computational time increment. On the other hand, although the proposed method is parameter tuning-free, $q$ is determined automatically by the number of iterations when our method converges. For speech signals in the TIDIGITS dataset, finding $q$ from the extracted spectrogram features is harder compared to that in the RWCP dataset, thereby requires a longer time to converge. In contrast, the TNNR-ADMM and SVT methods are less sensitive to the higher temporal dimensionality of the speech signal, and the required CPU runtime only increases slightly from the earlier timing analysis on the RWCP dataset. While our proposed method becomes competitive when more features are corrupted, and requires only about 50% of the CPU time to these two methods when 40% features are masked.

### D. Missing-Feature Reconstruction Under Gaussian Noise

Besides the impulse noise, the sound signals are also exposed to other less severe quasi-stationary noise in real acoustic environments and communication channels. To investigate the applicability and effectiveness of the proposed matrix completion technique to these scenarios, we add Gaussian noise at different signal-to-noise ratios (SNR), ranging from 5 dB to 20 dB, to the corrupted spectrogram that has a fixed mask ratio of 10%. As the feature reconstruction results are shown in Fig. 5, the SNRs deteriorate substantially when adding Gaussian and impulse noises to the spectrogram feature. The TNNR-ADMM and R1MP methods can effectively restore the masked features, while the reconstruction result is slightly poorer than the original clean feature with Gaussian noise. The feature reconstructed by the SVT method achieves an average SNR that is higher than the original clean feature with Gaussian noise, it suggests that the SVT method can not only restore the masked feature but also compensate for the Gaussian noise. Our method consistently

TABLE III
COMPARISON OF DIFFERENT MATRIX COMPLETION TECHNIQUES FOR MISSING-FEATURE RECONSTRUCTION ON THE NOISE CORRUPTED SAMPLES FROM THE TIDIGITS DATASET. THE AVERAGE CLASSIFICATION RESULTS (%) OVER THREE INDEPENDENT SIMULATIONS ARE REPORTED

| Factory | 0 dB | 5 dB | 10 dB | 15 dB | 20 dB | Average |
|---|---|---|---|---|---|---|
| TNNR-ADMM | 35.85 | 75.10 | 90.23 | 95.29 | 97.57 | 78.81 |
| SVT | 34.62 | 67.15 | 89.79 | 92.88 | 96.77 | 76.24 |
| R1MP | 26.80 | 46.63 | 57.62 | 73.27 | 78.15 | 56.49 |
| This Work | **36.96** | **76.22** | **91.15** | **96.04** | **98.47** | **79.77** |
| Car | 0 dB | 5 dB | 10 dB | 15 dB | 20 dB | Average |
| TNNR-ADMM | 34.74 | 62.01 | 91.48 | 96.59 | 97.21 | 76.41 |
| SVT | 33.00 | 58.29 | 89.87 | 96.05 | 96.43 | 74.73 |
| R1MP | 23.63 | 37.98 | 57.57 | 77.64 | 81.29 | 55.62 |
| This Work | **34.82** | **62.02** | **92.58** | **97.74** | **97.68** | **76.97** |
| Babble | 0 dB | 5 dB | 10 dB | 15 dB | 20 dB | Average |
| TNNR-ADMM | 37.42 | 65.08 | 85.90 | 95.58 | 97.59 | 76.31 |
| SVT | 35.70 | 60.18 | 83.96 | 93.76 | 96.94 | 74.11 |
| R1MP | 22.84 | 37.32 | 59.90 | 71.92 | 79.82 | 54.36 |
| This Work | **37.51** | **65.50** | **86.68** | **96.25** | **98.05** | **76.80** |
| Gaussian | 0 dB | 5 dB | 10 dB | 15 dB | 20 dB | Average |
| TNNR-ADMM | 41.00 | 69.73 | 93.61 | 96.86 | 98.09 | 79.86 |
| SVT | 39.45 | 64.78 | 91.46 | 95.14 | 96.85 | 77.54 |
| R1MP | 34.21 | 46.98 | 59.50 | 71.89 | 78.94 | 58.30 |
| This Work | **41.11** | **70.12** | **93.88** | **96.91** | **98.13** | **80.03** |

TABLE IV
COMPARISON OF DIFFERENT MATRIX COMPLETION TECHNIQUES FOR MISSING-FEATURE RECONSTRUCTION ON THE NOISE CORRUPTED SAMPLES FROM THE RWCP DATASET. THE AVERAGE CLASSIFICATION RESULTS (%) OVER THREE INDEPENDENT SIMULATIONS ARE REPORTED

| Factory | 0 dB | 5 dB | 10 dB | 15 dB | 20 dB | Average |
|---|---|---|---|---|---|---|
| TNNR-ADMM | 48.00 | 64.00 | 91.33 | 96.17 | 97.67 | 79.43 |
| SVT | 46.50 | 62.83 | 89.67 | 95.00 | 96.00 | 78.00 |
| R1MP | 39.00 | 52.83 | 83.00 | 85.17 | 86.33 | 69.27 |
| This Work | **53.00** | **66.33** | **92.17** | **97.83** | **99.67** | **81.80** |
| Car | 0 dB | 5 dB | 10 dB | 15 dB | 20 dB | Average |
| TNNR-ADMM | 50.50 | 73.83 | 92.17 | 96.33 | **99.50** | 82.47 |
| SVT | 49.83 | 71.67 | 91.17 | 95.83 | 97.67 | 81.23 |
| R1MP | 38.33 | 61.83 | 72.00 | 89.67 | 93.83 | 71.13 |
| This Work | **54.18** | **75.33** | **94.17** | **97.67** | 99.33 | **84.13** |
| Babble | 0 dB | 5 dB | 10 dB | 15 dB | 20 dB | Average |
| TNNR-ADMM | 62.17 | 85.83 | 94.50 | 96.00 | 99.00 | 84.63 |
| SVT | 59.83 | 83.83 | 94.67 | 95.50 | 98.50 | 84.67 |
| R1MP | 51.33 | 67.33 | 81.33 | 77.17 | 91.00 | 69.29 |
| This Work | **65.00** | **88.00** | **97.17** | **98.83** | **99.67** | **87.25** |
| Gaussian | 0 dB | 5 dB | 10 dB | 15 dB | 20 dB | Average |
| TNNR-ADMM | 73.00 | 78.50 | 96.83 | 97.67 | 98.67 | 88.93 |
| SVT | 76.83 | 81.67 | 97.00 | 98.00 | 98.83 | 90.47 |
| R1MP | 62.33 | 62.83 | 81.83 | 84.17 | 92.67 | 76.77 |
| This Work | **77.00** | **83.33** | **98.00** | **98.50** | **99.50** | **91.27** |

outperforms all these matrix completion methods by more than 2 dB for both datasets. We also note that the estimated mask works well under low noise scenarios, where SNR $\geq$ 15 dB. While the quality of the estimated mask degrades under more adverse scenarios, adversely affecting the feature reconstruction results. Nevertheless, the reconstruction results are still better than or comparable to other baseline methods that are using oracle mask. This result demonstrates a promising prospect of applying the proposed matrix completion technique to improve system performance under adverse acoustic conditions. It is also worth noting that the classification accuracy could be further improved by injecting the prior-knowledge of the type and data statistic of the noise during the DNN-based classifier training [46].

The classification results using the DNN-based classifier are provided in Tables III and IV for the TIDIGITS and RWCP datasets, respectively. For a fair comparison with other non-Gaussian noises evaluated in Section IV-E, no impulse noises are added here. By restoring the features from noise, the proposed method can achieve a promising result for SNRs above 10 dB with an accuracy above 90%. While the results degrade rapidly for more adverse acoustic environments, especially for the TIDIGITS dataset. This can be attributed to the fact that no prior information about the audio signals nor the noises have been exploited by the proposed matrix completion methods. We note that the feature reconstruction performance for the Gaussian Noise is superior to other more realistic noises in the following section. This is due to a higher level of variability and hence complexity involved in those more realistic acoustic environments.

### *E. Missing-Feature Reconstruction Under Non-Gaussian Noise*

To further investigate the feature reconstruction performance of the proposed method under real acoustic environments, we conduct experiments by adding noise samples from the NOISEX-92 dataset [48] onto the clean audio samples at different SNRs, ranging from 0 dB to 20 dB. We provide the classification results in Tables III and IV for the TIDIGITS and RWCP datasets, respectively. Across all the noisy scenarios been tested (i.e., Factory, Car, and Babble), the reconstructed features from the TNNR-ADMM, SVT, and our methods have shown promising results under low noise scenarios (15 dB and 20 dB), with a classification accuracy above 90%. The performance degrades smoothly until 5 dB where accuracies of above 60% can still be achieved. Among all the methods been evaluated, our proposed method consistently outperforms the other methods across all the testing scenarios and audio characteristics, i.e., environmental sound and speech. Below 5 dB SNR, classification accuracy drops significantly. This can be explained by the fact that the matrix completion methods do not exploit any prior knowledge of the noisy environments and the audio samples, therefore, it remains challenging to apply these methods to the highly corrupted samples.

## V. CONCLUSION

In this work, we apply a matrix completion technique to tackle the missing-feature reconstruction task. Most of the existing matrix completion techniques require hyperparameter tuning which is costly and difficult in practice. To resolve this problem, a parameter tuning-free matrix completion method based on matrix factorization is proposed. The proposed method can restore the missing features with high fidelity and computation efficiency, as demonstrated with both speech and environmental sound signals. Moreover, the experiments in speech recognition and environmental sound classification tasks also highlight the importance of high-quality audio features and the effectiveness of the proposed matrix completion technique in addressing the feature degradation problems including impulse and quasi-stationary Gaussian noises. To investigate the applicability of our

method to more challenging acoustic environments, our method allows optimizing to $\ell_p$-norm ($p \geq 1$) to handle non-Gaussian noises, e.g., factory, car, and babble noises as used in our experimental evaluation. We demonstrate high robustness to an intermediate level of such noises with the proposed matrix completion method. However, given no prior knowledge about the acoustic environments and audio samples, it remains challenging to apply matrix completion techniques to reconstruct highly corrupted audio features. Therefore, matrix completion should be considered as complementary to other sparse imputation methods [16], [17], [19] when the required prior information is not available.

It is worth noting that the problem tackled by the proposed method is different from that in the conventional audio inpainting framework, where short-period, continuous audio segments are missing. The proposed method is not applicable to recover the data wherein all the components in a particular column are erroneous or missing. As future work, we will explore different strategies to remove this constraint. The possible solutions include characterizing the corrections among corrupted columns by exploiting low-rank Hankel property [49], or conducting imputation on a modeling framework using information across multiple references, solved by the matrix co-clustering factorization (MCCF) in iterative update [50].

## APPENDIX

*Proof of Proposition 1:* At first, by substituting the $\mathbf{R}_q$ into (15), we can see that the problems $\bar{\mathcal{G}}$ in (15) and $\mathcal{G}$ in (10) are the same. Regarding the condition in (12), we have:

$$\langle \mathbf{X}_q, \mathbf{R}_q \rangle = \langle \mathbf{u}_q \mathbf{v}_q^T, \mathbf{R}_q \rangle = \sigma_{\max}(\mathbf{R}_q) \tag{24}$$

based on the definition of $\mathbf{X}_q = \mathbf{u}_q \mathbf{v}_q^T$, where $\sigma_{\max}(\mathbf{R}_q)$ is the maximum singular value of the residual matrix $\mathbf{R}_q$. According to the analysis of rank-one matrix approximation in [28], we get:

$$||\mathbf{R}_q||_p^p \leq ||\mathbf{R}_{q-1}||_p^p - \langle \mathbf{X}_{q-1}, \mathbf{R}_{q-1} \rangle^p \tag{25}$$

$$= \left(1 - \frac{\sigma_{\max}^p(\mathbf{R}_{q-1})}{||\mathbf{R}_{q-1}||_p^p}\right) ||\mathbf{R}_{q-1}||_p^p, \tag{26}$$

because of the convexity of $\ell_p$-norm at $p \geq 1$, where $\langle \cdot \rangle^p$ denotes the $\ell_p$-correlation [35]. From the recurrence relation, we conclude:

$$||\mathbf{R}_q||_p^p \leq ||\mathbf{M}_\Omega||_p^p \prod_{\ell=1}^{q-1} \left(1 - \frac{\sigma_{\max}^p(\mathbf{R}_\ell)}{||\mathbf{R}_\ell||_p^p}\right). \tag{27}$$

Since $0 < \frac{1}{\text{rank}(\mathbf{R}_\ell)} \leq \frac{\sigma_{\max}(\mathbf{R}_\ell)}{||\mathbf{R}_\ell||_p^p} \leq 1$, $||\mathbf{R}_q||_p^p \leq \alpha^{q-1} ||\mathbf{M}_\Omega||_p^p$ for $0 \leq \alpha < 1$. Besides, for each $\mathbf{\Upsilon}_\iota$, $\iota = 1, \ldots, 2q$, it is easy to observe that the least squares problem in (15) is convex and continuous when the others are fixed. Therefore, the conditions from (11) to (14) are satisfied.

## ACKNOWLEDGMENT

The authors would like to thank Prof. Haizhou Li for his insightful comments and for proofreading the initial version of the manuscript. They would also like to thank the handling editor and anonymous reviewers for their constructive feedbacks and suggestions, which have greatly strengthened the quality of the work. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of Singapore National Research Foundation, the Agency for Science, Technology and Research (A*STAR), and Zhejiang Lab.

## REFERENCES

[1] I. McLoughlin, H. Zhang, Z. Xie, Y. Song, and W. Xiao, "Robust sound event classification using deep neural networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 3, pp. 540–552, Mar. 2015.
[2] D. Yu and L. Deng, *AUTOMATIC SPEECH RECOGNITION: A Deep Learning Approach*. Berlin, Germany: Springer, 2016.
[3] D. Snyder, D. Garcia-Romero, D. Povey, and S. Khudanpur, "Deep neural network embeddings for text-independent speaker verification," in *Proc. INTERSPEECH*, 2017, pp. 999–1003.
[4] Y. Wang *et al.*, "Tacotron: A fully end-to-end text-to-speech synthesis model," 2017, *arXiv:1703.10135*.
[5] H. Chen, C. Leung, L. Xie, B. Ma, and H. Li, "Multitask feature learning for low-resource query-by-example spoken term detection," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 8, pp. 1329–1339, Dec. 2017.
[6] O. Dehzangi, B. Ma, E. S. Chng, and H. Li, "Discriminative feature extraction for speech recognition using continuous output codes," *Pattern Recognit. Lett.*, vol. 33, no. 13, pp. 1703–1709, 2012.
[7] J. Ryu, B. Lee, and D. Kim, "semg signal-based lower limb human motion detection using a top and slope feature extraction algorithm," *IEEE Signal Process. Lett.*, vol. 24, no. 7, pp. 929–932, Jul. 2017.
[8] A. Biem, S. Katagiri, and Biing-Hwang Juang, "Pattern recognition using discriminative feature extraction," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 500–504, Feb. 1997.
[9] J. Yang, R. K. Das, and H. Li, "Significance of subband features for synthetic speech detection," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 2160–2170, Nov. 2020.
[10] J. Wu, E. Yilmaz, M. Zhang, H. Li, and K. C. Tan, "Deep spiking neural networks for large vocabulary automatic speech recognition," *Front. Neurosci.*, vol. 14, 2020, Art. no. 199.
[11] Z. Pan, Y. Chua, J. Wu, M. Zhang, H. Li, and E. Ambikairajah, "An efficient and perceptually motivated auditory neural encoding and decoding algorithm for spiking neural networks," *Front. Neurosci.*, vol. 13, 2019, Art. no. 1420.
[12] Z. Pan, J. Wu, M. Zhang, H. Li, and Y. Chua, "Neural population coding for effective temporal classification," in *Proc. Int. Joint Conf. Neural Netw.*, 2019, pp. 1–8.
[13] Z. Pan, H. Li, J. Wu, and Y. Chua, "An event-based cochlear filter temporal encoding scheme for speech signals," in *Proc. Int. Joint Conf. Neural Netw.*, 2018, pp. 1–8.
[14] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "A regression approach to speech enhancement based on deep neural networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 1, pp. 7–19, Jan. 2015.
[15] Y.-L. Chang, K.-Y. Lee, P.-Y. Wu, H.-y. Lee, and W. Hsu, "Deep long audio inpainting," 2019, *arXiv:1911.06476*.
[16] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 3, pp. 922–932, Mar. 2012.
[17] P. Záviška, P. Rajmic, O. Mokrý, and Z. Průša, "A proper version of synthesis-based sparse audio declipper," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brighton, U.K., May 2019, pp. 591–595.
[18] A. Marafioti, N. Perraudin, N. Holighaus, and P. Majdak, "A context encoder for audio inpainting," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 12, pp. 2362–2372, Dec. 2019.
[19] C. Gaultier, N. Bertin, and R. Gribonval, "Cascade: Channel-aware structured cosparse audio declipper," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Calgary, AB, Canada, Apr. 2018, pp. 571–575.
[20] G. Kutyniok, "Theory and applications of compressed sensing," *GAMM-Mitteilungen*, vol. 36, no. 1, pp. 79–101, 2013.
[21] F. Cherfaoui, V. Emiya, L. Ralaivola, and S. Anthoine, "Recovery and convergence rate of the FrankWolfe algorithm for the M-exact-sparse problem," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7407–7414, Nov. 2019.

[22] Y. C. Eldar and A. V. Oppenheim, "Filterbank reconstruction of bandlimited signals from nonuniform and generalized samples," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2864–2875, Oct. 2000.
[23] C. Tzagkarakis and A. Mouchtaris, "Reconstruction of missing features based on a low-rank assumption for robust speaker identification," in *Proc. IISA 2014, 5th Int. Conf. Inf., Intell., Syst. Appl.*, Chania, Greece, 2014, pp. 432–437.
[24] B. Raj and R. M. Stern, "Missing-feature approaches in speech recognition," *IEEE Signal Process. Mag.*, vol. 22, no. 5, pp. 101–116, Sep. 2005.
[25] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, Jun. 2012.
[26] J.-F. Cai, E. J. Candés, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010.
[27] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2117–2130, Sep. 2013.
[28] Z. Wang, M.-J. Lai, Z. Lu, W. Fan, H. Davulcu, and J. Ye, "Rank-one matrix pursuit for matrix completion," in *Proc. 31st Int. Conf. Mach. Learn.*, ser. Res., E. P. Xing and T. Jebara, Eds., vol. 32, no. 2. Bejing, China: PMLR, 22–24 Jun. 2014, pp. 91–99.
[29] A. Narayanan and D. Wang, "Ideal ratio mask estimation using deep neural networks for robust speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 7092–7096.
[30] W. Kim and J. H. L. Hansen, "A novel mask estimation method employing posterior-based representative mean estimate for missing-feature speech recognition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 5, pp. 1434–1443, Jul. 2011.
[31] Q. Liu, C. Yang, Y. Gu, and H. C. So, "Robust sparse recovery via weakly convex optimization in impulsive noise," *Signal Process.*, vol. 152, pp. 84–89, 2018.
[32] E. J. Candés and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.
[33] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–722, 2009.
[34] E. J. Candés and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.
[35] W. Zeng and H. C. So, "Outlier-robust matrix completion via $\ell_p$ -minimization," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1125–1140, Mar. 2018.
[36] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proc. 45th Annu. ACM Symp. Theory Comput. - STOC' 13*, Palo Alto, CA, USA, Jun. 2013, pp. 665–674.
[37] Y. Zhou, D. Wilkinson, R. Schreiber, and R. Pan, "Large-scale parallel collaborative filtering for the netflix prize," in *Proc. Algorithmic Aspects Inf. Manage.*, R. Fleischer and J. Xu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 337–348.
[38] Q. Liu, X.-P. Li, and J. C. Yang, "Optimum co-design for image denoising between type-2 fuzzy identifier and matrix completion denoiser," *IEEE Trans. Fuzzy Syst.*, to be published doi: 10.1109/TFUZZ.2020.3030498.
[39] M. A. Davenport and J. Romberg, "An overview of low-rank matrix recovery from incomplete observations," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 4, pp. 608–622, Jun. 2016.
[40] W. Zeng, H. C. So, and L. Huang, "$\ell_p$-MUSIC: Robust direction-of-arrival estimator for impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4296–4308, Sep. 2013.
[41] M. Razaviyayn, M. Hong, and Z. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2012.
[42] D. P. O'Leary, "Robust regression computation computation using iteratively reweighted least squares," *SIAM J. Matrix Anal. Appl.*, vol. 11, no. 3, pp. 466–480, May 1990.
[43] D. Gamarnik and S. Misra, "A note on alternating minimization algorithm for the matrix completion problem," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1340–1343, Oct. 2016.
[44] R. G. Leonard and G. Doddington, "Tidigits speech corpus," *Linguistic Data Consortium, Philadelphia*, 1993.
[45] T. Nishiura and S. Nakamura, "An evaluation of sound source identification with rwcp sound scene database in real acoustic environments," in *Proc. IEEE Int. Conf. Multimedia Expo*, vol. 2, Lausanne, Switzerland, 2002, pp. 265–268.
[46] J. Wu, Y. Chua, M. Zhang, H. Li, and K. C. Tan, "A spiking neural network framework for robust sound classification," *Front. Neurosci.*, vol. 12, 2018, Art. no. 836.
[47] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[48] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: Ii. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 12, no. 3, pp. 247–251, 1993.
[49] S. Zhang and M. Wang, "Correction of corrupted columns through fast robust Hankel matrix completion," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2580–2594, May 2019.
[50] B. Jiang *et al.*, "SparRec: An effective matrix completion framework of missing data imputation for GWAS," *Sci. Rep.*, vol. 6, no. 1, Dec. 2016, Art. no. 35534.