

## Article

# Multi-Channel Audio Completion Algorithm Based on Tensor Nuclear Norm

Lin Zhu <sup>1,2</sup>, Lidong Yang <sup>1,2,\*</sup> , Yong Guo <sup>3</sup>, Dawei Niu <sup>1</sup> and Dandan Zhang <sup>1</sup>

<sup>1</sup> School of Digital and Intelligence Industry, Inner Mongolia University of Science and Technology, 7 Ardin Street, Baotou 014010, China; zhu59@stu.imust.edu.cn (L.Z.); ndwduo@163.com (D.N.); dandan@imust.edu.cn (D.Z.)

<sup>2</sup> Inner Mongolia Key Laboratory of Pattern Recognition and Intelligent Image Processing, 7 Ardin Street, Baotou 014010, China

<sup>3</sup> School of Science, Inner Mongolia University of Science and Technology, 7 Ardin Street, Baotou 014010, China; guo\_yong@imust.edu.cn

\* Correspondence: yld\_nkd@imust.edu.cn

**Abstract:** Multi-channel audio signals provide a better auditory sensation to the audience. However, missing data may occur in the collection, transmission, compression, or other processes of audio signals, resulting in audio quality degradation and affecting the auditory experience. As a result, the completeness of the audio signal has become a popular research topic in the field of signal processing. In this paper, the tensor nuclear norm is introduced into the audio signal completion algorithm, and the multi-channel audio signals with missing data are restored by using the completion algorithm based on the tensor nuclear norm. First of all, the multi-channel audio signals are preprocessed and are then transformed from the time domain to the frequency domain. Afterwards, the multi-channel audio with missing data is modeled to construct a third-order multi-channel audio tensor. In the next part, the tensor completion algorithm is used to complete the third-order tensor. The optimal solution of the convex optimization model of the tensor completion is obtained by using the convex relaxation technique and, ultimately, the data recovery of the multi-channel audio with data loss is accomplished. The experimental results of the tensor completion algorithm and the traditional matrix completion algorithm are compared using both objective and subjective indicators. The final result shows that the high-order tensor completion algorithm has a better completion ability and can restore the audio signal better.

**Keywords:** multi-channel audio signal; audio recovery; tensor nuclear norm; tensor completion; signal processing



**Citation:** Zhu, L.; Yang, L.; Guo, Y.; Niu, D.; Zhang, D. Multi-Channel Audio Completion Algorithm Based on Tensor Nuclear Norm. *Electronics* **2024**, *13*, 1745. <https://doi.org/10.3390/electronics13091745>

Academic Editor: Esteban Tlelo-Cuautle

Received: 11 March 2024

Revised: 29 April 2024

Accepted: 30 April 2024

Published: 1 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Audio signal is ubiquitous in our daily life. With the development of science and technology, audio has also developed from monophonic audio to dual-channel audio and multi-channel audio, such as 2.1-channel, 5.1-channel, 7.1-channel, and so on [1]. The pursuit of multi-channel audio technology is to restore the various sound effects that humans hear in nature [2]. Therefore, multi-channel audio is closer to the real sound heard by the human ear and provides a better immersive experience for the audience. At present, multi-channel audio is widely used in the fields of movies, TV programs, music production, and game development. However, the process of collecting and transmitting multi-channel audio signal is accompanied by abnormal phenomena such as missing data and audio damage. These phenomena will deteriorate the quality of the final audio, affect the auditory sensation, and even reduce the intelligibility of the audio content. If the damaged audio is applied to other tasks, such as audio recognition and classification tasks, it will affect the final accuracy. Therefore, the restoration of damaged multi-channel audio has become one of the current research hotspots [3].

The core problem related to recovering audio signal is how to establish a link between lost data and known data [4]. At present, some traditional audio signal restoration algorithms have some problems, including the fact that the algorithms are complex and the effect of audio signal recovery is not satisfactory. For instance, the audio restoration algorithm based on sparse decomposition [5] needs more iterations, in order to approach the optimal result. The audio restoration algorithm based on the regression model [6] needs to adjust the model order and other parameters; the problem with this algorithm is hearing distortion. The traditional matrix completion algorithm [7] may incur the problem of information loss and serious performance degradation. At present, these algorithms are not especially used for the restoration of multi-channel audio signals. Therefore, these algorithms do not take into account the spatial position of multi-channel audio and the strong correlation between channels in relation to the impact of the audio completion effect.

As we know, the multi-channel audio signal can be thought of as a multi-dimension model that contains channel, time, and spectrum [8]. However, the traditional matrix model cannot directly process high-order data such as the multi-channel audio signal. It needs to transform the high-order data into a matrix using dimensionality reduction operations. This step will result in the loss of some structural information and alter the effect of signal recovery. At the same time, the representation of multi-dimensional data by the matrix is inefficient [9]. As an extension of the matrix in the high-dimensional space, the tensor has been utilized for multi-dimensional array processing [10]. Therefore, the tensor model can be used to represent high-order data and to directly analyze and process high-order data [11]. This feature can reflect the inherent relationship of multi-factor signals well, so the tensor model has been widely used in image processing [12], computer vision [13], and other fields. Therefore, in order to make full use of the correlation between the various factors of the audio signal [14], researchers have developed tensor completion algorithms.

The tensor completion methods can be roughly grouped into tensor factorization-based methods and tensor completion-based methods [15]. Tensor decomposition can succinctly represent the underlying structure of a tensor; therefore, various tensor factorizations are applied to tensor completion. CANDECOMP/PARAFAC (CP) is one of the well-known tensor factorizations. Currently, one of the most commonly used methods is the CP weighted optimization (CP-WOPT) algorithm. This method is often used for audio completion. CP is a special case of Tucker decomposition and so Tucker is also used for completion. However, compared with Tucker, the tensor train (TT) decomposition has a better ability to represent tensors and can avoid the curse of dimensionality, so it is better for completion. Sedighin et al. [9] used a complete algorithm based on TT decomposition to reconstruct the signal in the multiway delay space. In addition, the tensor nuclear norm is the sum of the singular values of the frontal slices of the tensor after Fourier transformation and it is the tightest convex relaxation of the  $L_1$  norm of the tensor. Studies have shown that methods based on nuclear norm are superior to methods based on tensor factorization. Therefore, the tensor nuclear norm is used by the researchers for signal completion. For example, Ran et al. [15] adopted high-accuracy Low Rank Tensor Completion (HaLRTC) to complete the traffic data.

In this paper, tensor completion based on the tensor nuclear norm is used to recover the data of the multi-channel audio signal. First of all, the multi-channel audio signal is preprocessed to complete the operations of the framework and adding window. Then, the multi-channel audio signal is transformed from the time domain to the frequency domain. The next step is to construct a third-order multi-channel audio tensor. Finally, the multi-channel audio signal with missing data is recovered by using the completion algorithm based on the tensor nuclear norm. By finding the connection between the missing data and the known data, the lost data can be recovered, as much as possible, from some observation data, so that the quality and auditory effect of the multi-channel audio signal are improved.

The rest of this paper is organized as follows: Section 2 introduces the notations and methods. The settings of the experiment are described in Section 3. The experimental results are discussed in Section 4. In Section 5, conclusions are discussed.

## 2. Materials and Methods

### 2.1. Notation and Definitions

A tensor is a high-order generalization of a vector and matrix [16], which represents an element of N-way multifactor space. The order of a tensor is the number of dimensions, also known as modes [17]. An M-order tensor is denoted as  $\mathcal{H} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$ , where  $I_m$  ( $1 \leq m \leq M$ ) is the size of the dimension  $m$ , and the elements in the tensor can be represented by  $h_{i_1 \cdots i_M}$ . For brevity, the main notations in this paper are summarized in Table 1.

**Table 1.** The main notations in this paper are summarized in the table.

Notation	Implication	Notation	Implication
$h$	A scalar	$\mathbf{h}$	A vector
$\mathbf{H}$	A matrix	$\mathcal{H}$	A tensor
$\mathcal{H}(i, :, :)$	The $i$ -th horizontal slice of $\mathcal{H}$	$\mathcal{H}(:, j, k)$	The column fiber of $\mathcal{H}$
$\mathcal{H}(:, j, :)$	The $j$ -th lateral slice of $\mathcal{H}$	$\mathcal{H}(i, :, k)$	The row fiber of $\mathcal{H}$
$\mathcal{H}(:, :, k)$	The $k$ -th frontal slice of $\mathcal{H}$	$\mathcal{H}(i, j, :)$	The tube fiber of $\mathcal{H}$
$\ \mathcal{H}\ _F$	$\ \mathcal{H}\ _F = \sqrt{\sum_{i_1=1}^{I_1} \cdots \sum_{i_M=1}^{I_M} h_{i_1 \cdots i_M}^2}$	$\ \mathbf{H}\ $	$\ \mathbf{H}\  = \max_i \sigma_i(\mathbf{H})$
$\ \mathcal{H}\ _1$	$\ \mathcal{H}\ _1 = \sum_{i_1=1}^{I_1} \cdots \sum_{i_M=1}^{I_M}  h_{i_1 \cdots i_M} $	$\ \mathbf{H}\ _*$	$\ \mathbf{H}\ _* = \sum_i \sigma_i(\mathbf{H})$

The inner product between two tensors is similar to the inner product between two matrices. The inner product is a scalar, which is the sum of the product of the elements at the corresponding positions. Given two tensors  $\mathcal{H}, \mathcal{B} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$  that have the same size, the inner product of tensors is defined as  $\langle \mathcal{H}, \mathcal{B} \rangle = \sum_{i_1=1}^{I_1} \cdots \sum_{i_M=1}^{I_M} h_{i_1 \cdots i_M} b_{i_1 \cdots i_M}$ , accordingly. The mode- $d$  matricization (the mode- $d$  unfolding) of a tensor is the rearrangement of the fibers of the tensor into a new matrix  $H_{(d)}$ , according to the direction of dimension  $d$ ; its mathematical expression is  $\mathbf{H}_{(d)} \in \mathbb{R}^{I_d \times \prod_{n \neq d} I_n}$ .

**Definition 1.** (T-product [18]): Given two tensors  $\mathcal{H}_1 \in \mathbb{R}^{I_1 \times I_2 \times I_3}$  and  $\mathcal{H}_2 \in \mathbb{R}^{I_2 \times I_4 \times I_3}$ , the t-product of them is defined as

$$\mathcal{H} = \mathcal{H}_1 * \mathcal{H}_2 = \text{fold}(\text{bcirc}(\mathcal{H}_1) \cdot \text{unfold}(\mathcal{H}_2)) \quad (1)$$

where  $\text{bcirc}(\cdot)$  represents the block circulant matrix,  $\text{unfold}(\cdot)$  represents the matricization of the tensor, and  $\text{fold}(\cdot)$  is the inverse operation. The result of t-product is also a tensor  $\mathcal{H} \in \mathbb{R}^{I_1 \times I_4 \times I_3}$ , whose  $(i, j)_{th}$  tube fiber is  $\mathcal{H}(i, j, :) = \sum_{l=1}^{I_2} \mathcal{H}_1(i, l, :) \bullet \mathcal{H}_2(l, j, :)$ , where  $\bullet$  indicates cycle convolution [19].

**Definition 2.** (T-SVD [20]): Given a tensor  $\mathcal{H} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ , it can be factored as

$$\mathcal{H} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top \quad (2)$$

where  $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$  is an f-diagonal tensor, whose frontal slices are all diagonal matrices.  $\mathcal{U} \in \mathbb{R}^{I_1 \times I_1 \times I_3}$  and  $\mathcal{V} \in \mathbb{R}^{I_2 \times I_2 \times I_3}$  are two orthogonal tensors.  $\mathcal{V}^\top \in \mathbb{R}^{I_2 \times I_2 \times I_3}$  is the conjugate transpose of  $\mathcal{V}$ .

**Definition 3.** (Tensor Tubal Rank [20,21]): Given a tensor  $\mathcal{H} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$  with the t-SVD decomposition  $\mathcal{H} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$ , its tensor tubal rank that is denoted as  $\text{rank}_t(\mathcal{H})$  is defined as the number of non-zero singular tubes of  $\mathcal{S}$ . The expression is

$$\text{rank}_t(\mathcal{H}) = \sum_i 1(\mathcal{S}(i, i, :) \neq 0) \quad (3)$$

Since the tensor tubal rank is determined by the first frontal slice,  $\mathcal{S}(i, i, :)$  in the expression of the tensor tubal rank can be written as  $\mathcal{S}(i, i, 1)$ .

**Definition 4.** (Tensor Nuclear Norm [20]): Given a tensor  $\mathcal{H} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ , whose  $t$ -SVD decomposition is  $\mathcal{H} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$ , the mathematical formula of tensor nuclear norm is defined as

$$\|\mathcal{H}\|_* = \langle \mathcal{S}, \mathcal{I} \rangle = \sum_{i=1}^r \mathcal{S}(i, i, 1) \quad (4)$$

where  $r = \text{rank}_t(\mathcal{H})$  and  $\mathcal{I}$  is an identity tensor.

## 2.2. Tensor Completion Algorithms

Tensor completion involves building an optimization model using a small amount of known signal data and then recovering the original signal with a high probability, by solving the model. An incomplete tensor with missing elements is given and, in order to obtain the complete tensor, we need to obtain the global information by using the rank function of the tensor. The rank function is an effective means to obtain the global information of the data. However, the rank function is a non-convex function and the result is not guaranteed to be optimal. As a consequence, traditional matrix completion involves using the nuclear norm of the matrix to approximate the rank function and transforming it into a problem of convex optimization [22], to obtain a global optimal solution. However, the matrix model is two-dimensional, so it is necessary to reduce the dimension when processing multi-dimensional data such as multi-channel audio signals. This operation will result in the loss of structural information. Audio signal recovery cannot be performed well and the recovery effect is not satisfactory [23]. The tensor completion algorithm is a new technique for recovering lost data, which is developed from matrix completion algorithms. It can be considered as a high-order extension of matrix completion. The tensor completion method can directly process multi-dimensional signals with high-order structures and can restore data. The convex relaxation technology is used by the tensor completion algorithm to transform the rank minimization problem, which is an NP-hard problem, into a convex optimization problem of nuclear norm [24]. The robust tensor recovery problem is formulated as a convex problem, which can then be solved to obtain the optimal solution. Afterwards, the data recovery for a high-order signal with data loss is performed.

In this paper, the tensor nuclear norm is introduced into the multi-channel audio signal recovery task; the tensor nuclear norm is an extension of the matrix nuclear norm. The nuclear norm of a tensor can be considered as a convex combination of the nuclear norms of the matrices obtained after the tensor is matricized at each order. The robust tensor completion algorithm based on the tensor nuclear norm minimization (TC-TNN) [25] is used to complete the multi-channel audio signal with missing data. The results of the experiment are compared to the results of the robust tensor completion algorithm based on the sum of the matrix nuclear norm minimization (TC-SNN) [26], CANDECOMP/PARAFAC weighted optimization algorithm (CP-WOPT) [27], and the traditional matrix completion algorithm [28]. The following primarily introduces the tensor completion algorithms.

### 2.2.1. Robust Tensor Completion Based on the Tensor Nuclear Norm Minimization

Classical algorithms are affected by large amounts of noise and, therefore, cannot work properly. In order to better solve their sensitivity to noise and to improve their robustness, the TC-TNN algorithm is used for restoration, whose aim is to recover the low-rank tensor damaged by sparse errors. In this algorithm, the noise only needs to be assumed to be sparse, regardless of the strength of the noise. It is able to recover intrinsically low-rank parts from large and sparse noise-contaminated observations. Therefore, its robustness is stronger than that of classical algorithms.

The completion model of the TC-TNN algorithm [25] is defined as follows:

$$\min_{\mathcal{L}, \mathcal{S}} \|\mathcal{L}\|_* + \lambda \|\mathcal{S}\|_1, \text{ s.t. } \mathcal{T} = \mathcal{L} + \mathcal{S} \quad (5)$$

where  $\lambda = 1/\sqrt{\max(I_1, \dots, I_{M-1})I_M}$  and  $\mathcal{T}$  is a tensor which can be decomposed as  $\mathcal{L}$  and  $\mathcal{S}$ .  $\mathcal{L}$  is a low-rank part and  $\mathcal{S}$  is a sparse part; both components are of arbitrary size. The setup of  $\mathcal{S}$  improves the robustness of the algorithm. In the audio completion experiment, the low-rank part of the audio tensor is the original audio and the sparse part is the part that causes damage.

The Lagrange function of (8) is written as follows:

$$L(\mathcal{L}, \mathcal{S}, \mathcal{Z}, \mu) = \|\mathcal{L}\|_* + \lambda \|\mathcal{S}\|_1 + \langle \mathcal{Z}, \mathcal{L} + \mathcal{S} - \mathcal{T} \rangle + \frac{\mu}{2} \|\mathcal{L} + \mathcal{S} - \mathcal{T}\|_F^2 \quad (6)$$

where  $\mathcal{Z}$  is an auxiliary variable. Using the Alternating Direction Method of Multiplier (ADMM) [29] to solve the problem of the TC-TNN algorithm, the specific steps are as follows:

- (1) Update the original audio part  $\mathcal{L}$ . The optimal solution of  $\mathcal{L}$  is

$$\mathcal{L}^{k+1} = \underset{\mathcal{L}}{\operatorname{argmin}} \|\mathcal{L}\|_* + \frac{\mu^k}{2} \left\| \mathcal{L} + \mathcal{S}^k - \mathcal{T} + \frac{\mathcal{Z}^k}{\mu^k} \right\|_F^2 \quad (7)$$

- (2) Update the sparse part  $\mathcal{S}$ . The optimal solution of  $\mathcal{S}$  is

$$\mathcal{S}^{k+1} = \underset{\mathcal{S}}{\operatorname{argmin}} \lambda \|\mathcal{S}\|_1 + \frac{\mu^k}{2} \left\| \mathcal{L}^{k+1} + \mathcal{S} - \mathcal{T} + \frac{\mathcal{Z}^k}{\mu^k} \right\|_F^2 \quad (8)$$

- (3) Update  $\mathcal{Z}$ . The update of the dual variable is given by

$$\mathcal{Z}^{k+1} = \mathcal{Z}^k + \mu^k (\mathcal{L}^{k+1} + \mathcal{S}^{k+1} - \mathcal{T}) \quad (9)$$

Finally, the errors of the obtained values are calculated according to the following formula, and then the errors are compared with the allowable error. If the errors are less than the allowable error, the completed result is returned.

$$\begin{cases} \|\mathcal{L}^{k+1} - \mathcal{L}^k\|_\infty \leq tol \\ \|\mathcal{S}^{k+1} - \mathcal{S}^k\|_\infty \leq tol \\ \|\mathcal{L}^{k+1} + \mathcal{S}^{k+1} - \mathcal{T}\|_\infty \leq tol \end{cases} \quad (10)$$

### 2.2.2. Robust Tensor Completion Based on the Sum of the Matrix Nuclear Norm Minimization

The completion model of the TC-SNN algorithm [26] is defined as follows:

$$\min_{\mathcal{L}, \mathcal{S}} \sum_{m=1}^M \lambda_m \|\mathcal{L}_{(m)}\|_* + \|\mathcal{S}\|_1, \text{ s.t. } \mathcal{T} = \mathcal{L} + \mathcal{S} \quad (11)$$

The robustness of this algorithm also lies in the fact that  $\mathcal{S}$  only needs to be assumed to be sparse, regardless of the strength. It has the ability to recover the intrinsically low-rank part from large and sparse noise-contaminated observations. The low-rank part of the audio tensor is the original audio and the sparse part is the part that causes damage.

The following equivalent problem is obtained by introducing auxiliary variable  $\mathcal{Z}_m$  for further optimization

$$\min_{\mathcal{L}, \mathcal{S}} \sum_{m=1}^M \frac{1}{2} \|\mathcal{L}_m - \mathcal{Z}_m\|^2 + \mu \|\mathcal{S}\|_1 \quad (12)$$

where  $\mathcal{Z}_m = \mathcal{T} + \mu \Lambda_m - \mathcal{S}$  and  $\Lambda_m$ ,  $m = 1, 2, \dots, M$ , is a Lagrangian operator related to  $M$  constraints. The global optimal solution to the problem is the best rank- $k$  approximation of  $\mathcal{Z}$ .

### 2.2.3. CP Weighted Optimization Algorithm

Given a tensor  $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ , whose rank is  $R$ , and a weighted tensor  $\mathcal{W} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ , whose elements is

$$w_{i_1 i_2 i_3} = \begin{cases} 1, & t_{i_1 i_2 i_3} \text{ is not lost} \\ 0, & t_{i_1 i_2 i_3} \text{ is lost} \end{cases} \quad (13)$$

CP-WOPT minimizes the weighted function, by finding the matrices  $\mathbf{A} \in \mathbb{R}^{I_1 \times R}$ ,  $\mathbf{B} \in \mathbb{R}^{I_2 \times R}$ , and  $\mathbf{C} \in \mathbb{R}^{I_3 \times R}$ . The weighted function [27] is as follows:

$$W(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \sum_{i_3=1}^{I_3} w_{i_1 i_2 i_3}^2 \left[ t_{i_1 i_2 i_3}^2 - 2t_{i_1 i_2 i_3} \sum_{r=1}^R a_{i_1 r} b_{i_2 r} c_{i_3 r} + \left( \sum_{r=1}^R a_{i_1 r} b_{i_2 r} c_{i_3 r} \right)^2 \right] \quad (14)$$

### 2.3. Modified Discrete Cosine Transform

Modified discrete cosine transform (MDCT) is a transform related to Fourier transform, based on the fourth type of discrete cosine transform (DCT). The DCT is a transform in the real number domain. The DCT has orthogonal transformation properties, and the basis vectors of its transformation matrix are very similar to the eigenvectors of the Toeplitz matrix, which reflects the correlation characteristics of the audio signal. Therefore, the DCT is considered a quasi-optimal transformation when performing orthogonal transformations on audio signals. However, the DCT will produce boundary artifacts when the signal is framed. As a consequence, the MDCT is proposed; this transformation performs windowing and overlapping after framing. The MDCT can effectively eliminate boundary artifacts because of the overlapping nature and its energy compressibility, which is similar to that of the DCT. Therefore, the MDCT is widely used in audio processing. Assume that the MDCT of a sequence  $z_m$  of length  $M$  is  $Z_k$ , and its forward and inverse transformation expressions are

$$Z_k = \sum_{m=0}^{M-1} z_m \cos \left[ \frac{\pi}{2M} \left( 2m + 1 + \frac{M}{2} \right) (2k + 1) \right], k \in \left[ 0, \frac{M}{2} - 1 \right] \quad (15)$$

$$z_m = \sum_{k=0}^{M/2-1} Z_k \cos \left[ \frac{\pi}{2M} \left( 2m + 1 + \frac{M}{2} \right) (2k + 1) \right], m \in [0, M - 1] \quad (16)$$

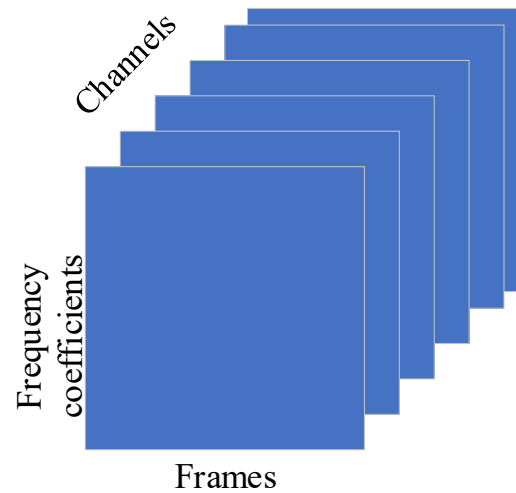
## 3. Experimental Setup

### 3.1. Audio Signal Modeling

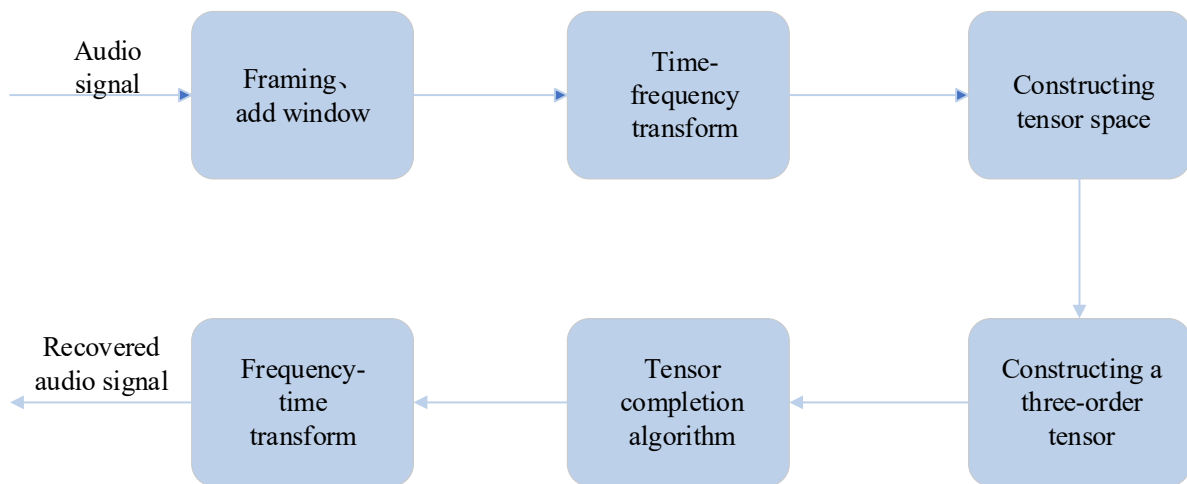
Although vectors and matrices are easier to process, in order to retain more structural correlations, this paper constructs the multi-channel audio signal into a third-order tensor. The multi-channel audio can have more than two axes of variation, such as channel, frame, and feature [30]. First of all, the multi-channel audio signal is divided into frames and then added the window. The frame length is generally set to 10–30 ms. There will be a partial overlap between the frames, to avoid the discontinuity of the audio signal in the time domain. The product of the frame length and sampling frequency is the number of samples in each frame. After that, the MDCT is performed on the frame samples of the processed audio signal and the audio signal is transformed from the time domain to the frequency domain, to obtain the frequency domain coefficients of each frame sample. The MDCT



transformation has anti-symmetric characteristics and, as a result of this, the number of frequency domain coefficients is equal to half the number of samples in the time domain. The frequency domain coefficients are selected as the characteristic parameters and as one of the orders of the audio tensor. Next, the multi-channel audio signal can be constructed into a third-order tensor, which is represented by  $\mathcal{H} \in \mathbb{R}^{I_p \times I_f \times I_c}$ , where  $I_p$  represents the coefficients after frequency domain transformation,  $I_f$  represents the frame samples, and  $I_c$  represents the number of channels of the audio signal. Its structure is shown in Figure 1. Then, the audio tensor is completed through the structural relationship within the audio signal in the frequency domain. The process of audio signal recovery is shown in Figure 2.



**Figure 1.** A three-order audio tensor.



**Figure 2.** Audio recovery flow chart.

### 3.2. Experiment Settings

The multi-channel audio signals used in the completion experiment are the 5.1-channel audio signals that are common in our daily life and that are downloaded from the Internet. The specific content of the audio is popular music. The format of the multi-channel audio used in the audio completion experiment is WAV. In total, 50 segments of audio are used in the experiment, the sampling frequency is 48 kHz, and the sampling bit depth is 16 bit. The duration of each audio segment is 10 s. Taking a piece of audio in the dataset as an example, the dynamic range of the left channel of this piece of audio is 46.82 dB, the dynamic range of the right channel is 50.77 dB, the dynamic range of the center channel is 47.62 dB, the dynamic range of the low frequency enhanced channel is 47.04 dB, the dynamic range

of the left surround channel is 50.92 dB, and the dynamic range of the right surround channel is 45.37 dB. The dynamic range of music is generally 40–60 dB and the audio used in the experiment is within this range. After framing, the length of each frame is set to 20 ms; thus, the number of samples contained in each frame is 960. The overlap between two frames is set to 50%; thus, the frame shift is 10 ms. The time–frequency conversion is completed through the MDCT transformation and the number of frequency domain coefficients obtained is 480. Then, the three-order audio tensor can be constructed.

The missing data of the multi-channel audio signal is due to the method of random loss, where data loss occurs at random locations. The total missing data rate is set to 15%, 30%, 45%, 60%, and 75%, respectively. According to the above experimental settings, a third-order audio tensor with data missing can be constructed. Then, four kinds of audio completion algorithms, the TC-TNN algorithm, the TC-SNN algorithm, the CP-WOPT algorithm, and the robust matrix completion (RMC) algorithm, are used, respectively, to carry out audio recovery experiments on audio signals with missing data. The allowable error of the experiment is  $10^{-8}$  and the maximum number of iterations is 500.

The recovery effect of the audio signal is evaluated using both objective and subjective evaluation indicators. The objective evaluation indicator is the relative standard error (RSE), while the subjective evaluation indicator is Multiple Stimuli with Hidden Reference and Anchor (MUSHRA). The completion experiment is conducted on a DELL 7050 computer with a 3.6 GHz CPU of Intel Core i7 and 16 GB RAM and the simulation software is Matlab (R2019a).

## 4. Results and Discussion

### 4.1. Objective Evaluation

In the audio recovery experiment, for each missing data rate, the experiment was repeated 10 times for each audio, to avoid coincidence. Then, the results of the recovery experiment are objectively evaluated. The objective evaluation indicator is RSE, which is a measure of the difference between the original signal and the recovered signal. The lower the value of RSE, the better the effect of audio signal recovery. RSE is defined as follows:

$$\text{RSE} = \frac{\|\mathcal{H}^* - \overline{\mathcal{H}}\|_F}{\|\overline{\mathcal{H}}\|_F} \quad (17)$$

where  $\mathcal{H}^*$  represents the tensor after completion and  $\overline{\mathcal{H}}$  represents the tensor without data loss.

Table 2 records the RSE of the multi-channel audio signals restored using the four audio completion algorithms, as well as the RSE are the average values of the results of the experiment of 50 pieces of multi-channel audio.

**Table 2.** The RSE of four audio recovery algorithms.

Missing Data Rate	TC-TNN	TC-SNN	CP-WOPT	RMC
15%	0.0115	0.0196	0.0239	0.0401
30%	0.0238	0.0282	0.0441	0.0598
45%	0.0391	0.0476	0.0603	0.0722
60%	0.0661	0.0731	0.0904	0.1048
75%	0.0948	0.1075	0.1093	0.1513

In addition, this experiment also counts the time taken, using different audio completion algorithms, to restore multi-channel audio signal, which is represented by CPU running time (CPU time). The results of this experiment are shown in Table 3.



**Table 3.** CPU time of four audio recovery algorithms (s).

Missing Data Rate	TC-TNN	TC-SNN	CP-WOPT	RMC
15%	86	112	94	67
30%	89	122	102	76
45%	94	135	117	86
60%	105	150	131	101
75%	127	171	154	121

It can be seen from Table 2 that, under all conditions of missing data rate, the value of RSE obtained using the TC-TNN algorithm is the lowest. This phenomenon shows that the audio recovery capability of the TC-TNN algorithm is the best, in all cases. It shows that the operation of constructing the multi-channel audio signal into a tensor can make full use of the inherent relationship of high-order structure, as well as recovering the missing data of the multi-channel audio signal better, so that the recovery ability of this tensor completion algorithm is stronger and the audio recovery quality is higher.

The recovery ability of the TC-SNN algorithm and the CP-WOPT algorithm are medium, in terms of the four completion algorithms, and these algorithms also carry out tensor modeling of multi-channel audio signals. Compared with the results of the RMC algorithm, the RSE of the TC-SNN algorithm and the CP-WOPT algorithm are lower. The TC-SNN algorithm is based on the matrix nuclear norm, while the TC-TNN algorithm is based on the tensor nuclear norm. The tensor nuclear norm can be considered as a higher extension of the matrix nuclear norm. As a result, the tensor nuclear norm has a high-order structure and its intrinsic correlation is stronger, compared to the matrix nuclear norm. For this reason, in each case of missing data rate, the RSE of the TC-SNN algorithm is slightly higher than the TC-TNN algorithm, as well as the recovery ability of the TC-SNN algorithm being slightly weaker than that of the TC-TNN algorithm. In the aspect of data completion, the nuclear norm is superior to the tensor factorization. Therefore, the recovery ability of the CP-WOPT algorithm is slightly weaker than that of the TC-TNN algorithm and the TC-SNN algorithm.

It can be seen from the two tables that the RMC algorithm takes relatively less time to recover the multi-channel audio signal compared to the other three algorithms, but the RSE of this algorithm is the highest among the four methods, owing to the fact that the RMC algorithm does not consider the spatial structure and other correlations of high-order data, and even loses the high-order structural information in the process of restoring the audio signal. As a result, its recovery ability is not as good as the tensor completion algorithm.

The tensor completion algorithm requires a large number of iterative operations to obtain the optimal solution, which leads to a long time for multi-channel audio signal recovery. However, compared to the traditional matrix completion algorithm, the audio recovery quality of the tensor completion algorithm is better, indicating that this type of algorithm increases the complexity of the algorithm, in exchange for a better audio recovery effect.

In addition, for the reason that the TC-TNN algorithm has the best recovery effect in the completion experiment, the spectrograms of the original audio and the audio that is recovered using the TC-TNN and RMC algorithms are shown in Figures 3–5. From top to bottom, they are the left channel, the right channel, the center channel, the low frequency enhanced channel, the left surround channel, and the right surround channel. In the spectrogram, the depth of the color indicates the energy of the frequency, the horizontal stripes represent the formant information, and the vertical stripes represent the pitch information. The denser the stripes, the higher the pitch. It can be seen that the energy of the frequency points of the corresponding channels of the audios is slightly different in Figures 3 and 4. However, the position and number of horizontal and vertical stripes are roughly same. The difference between Figures 3 and 5 is greater and some of the horizontal stripes are blurred. Hence, the spectrogram can also show that the audio recovery effect of the TC-TNN algorithm is better.

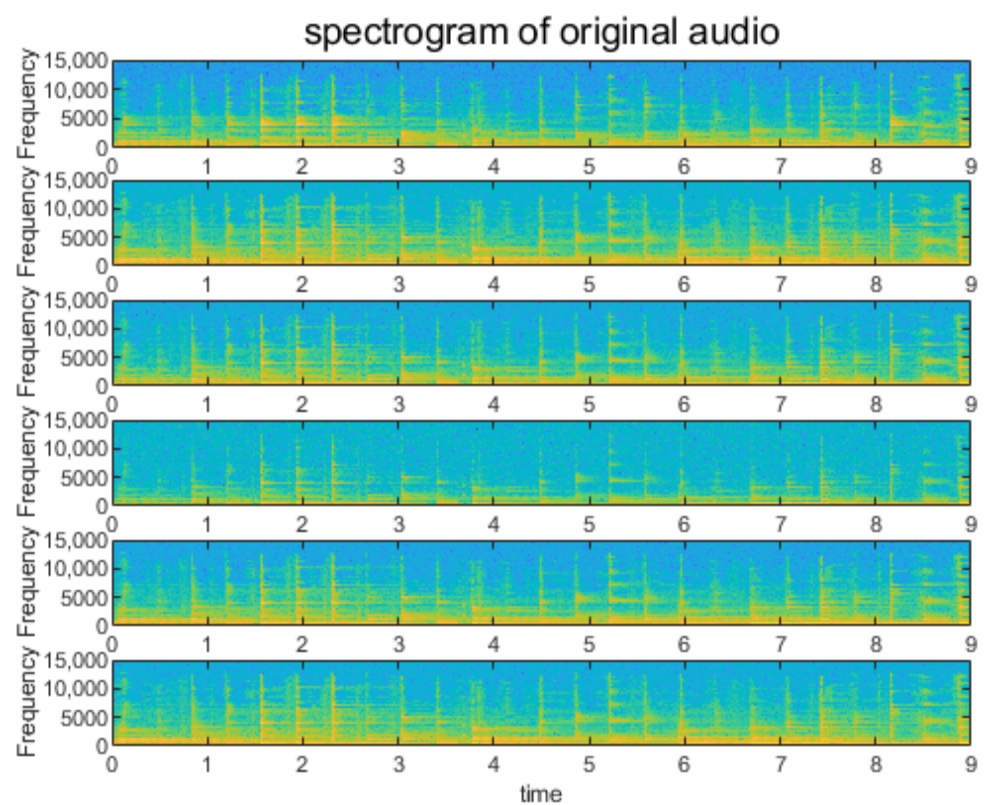


Figure 3. The spectrogram of the original audio.

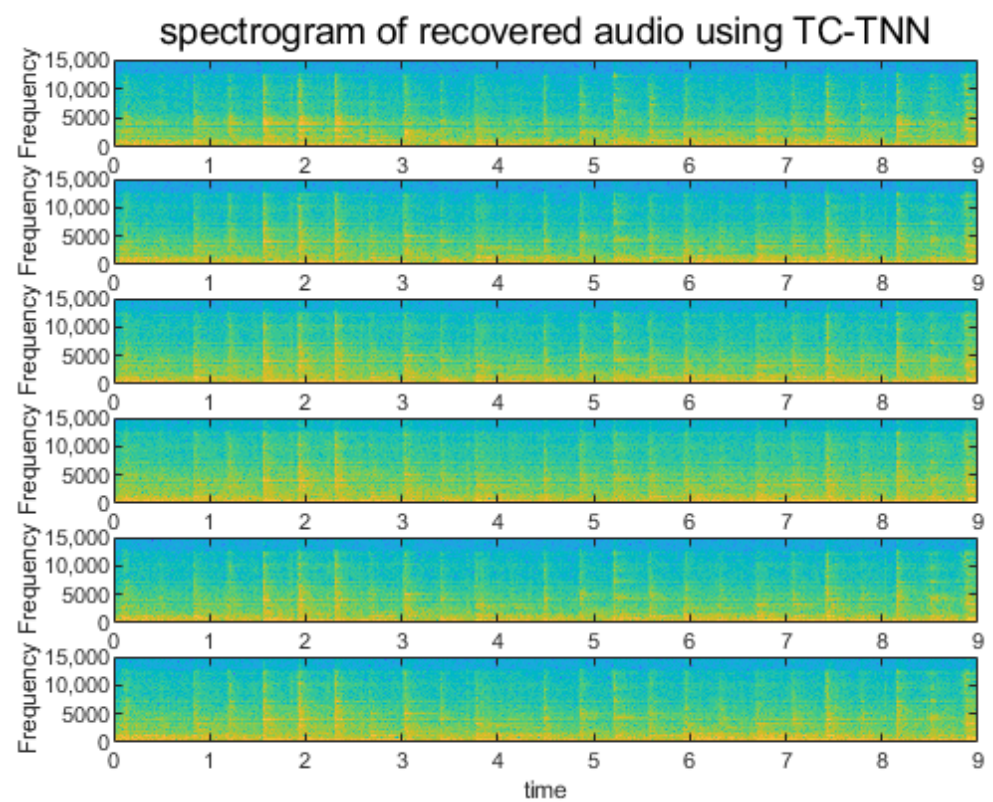
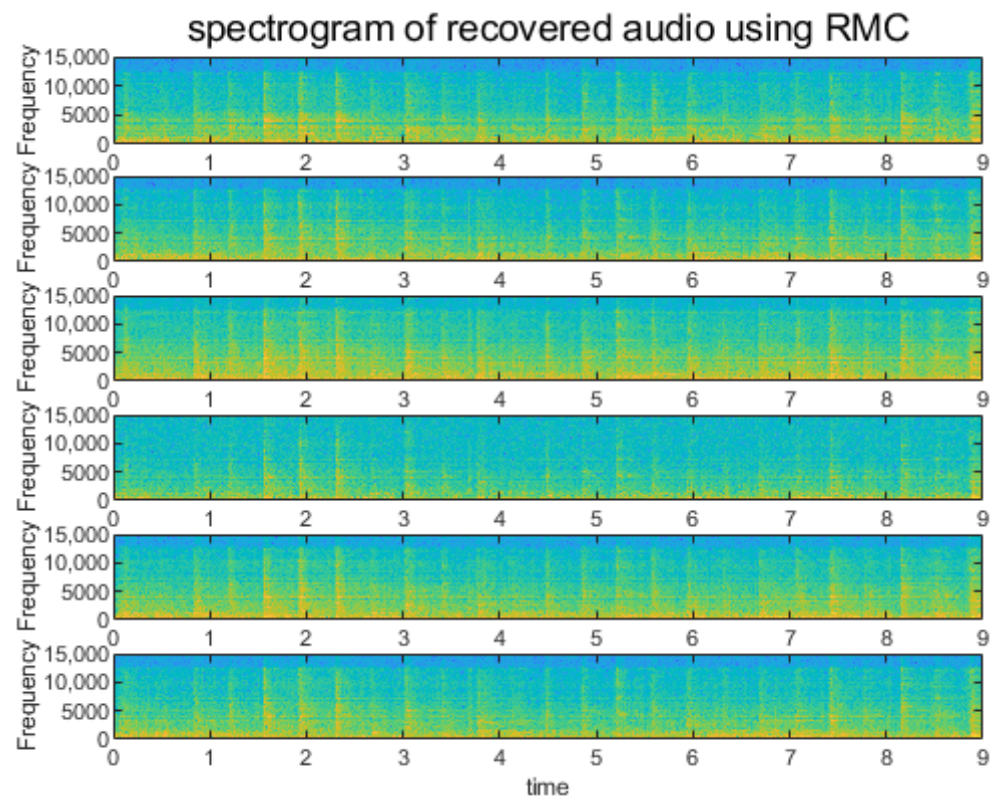


Figure 4. The spectrogram of the audio recovered using the TC-TNN algorithm.



**Figure 5.** The spectrogram of the audio recovered using the RMC algorithm.

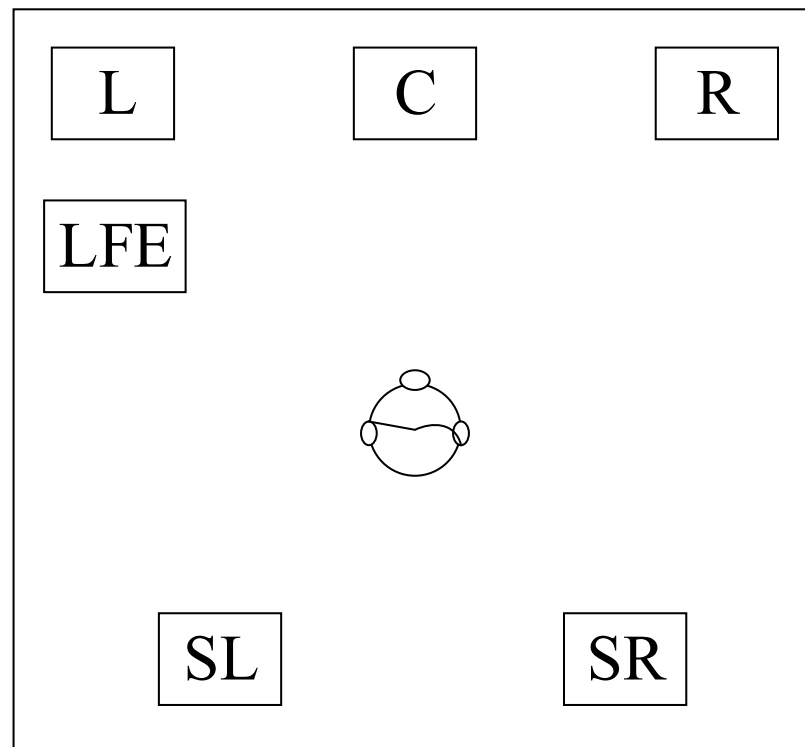
#### 4.2. Subjective Evaluation

The purpose of recovering the multi-channel audio signal with data loss is to improve the quality of the multi-channel audio, improve the intelligibility of the audio content, and to obtain a better auditory sensation. As a consequence, it is necessary to subjectively evaluate the results of the experiment and test the restoration quality of multi-channel audio in subjective hearing. The subjective evaluation indicator is the MUSHRA method. This test method is recommended by the International Telecommunication Union and was first used for the subjective evaluation of streaming media and the relevant coding of communication. The main feature of the MUSHRA method is to mix the lossless audio into the test corpus as a reference, with the total loss audio as an anchor. Through the double-blind listening test, the measured audio, the hidden reference audio, and the anchor audio are subjectively scored. This test method requires experienced listeners that need to be trained to be familiar with the test process and scoring rules before the formal test. The original multi-channel audio without data loss is generally used as a reference signal. During the formal test, the listeners scored the audio signal by comparing the multi-channel audio signal without data loss to the multi-channel audio signal after completion. The scores are integers ranging from 0 to 100 and the corresponding evaluations range from poor to very good.

In this experiment, ten experienced listeners, including five men and five women, were selected to conduct subjective audiometry and score the multi-channel audio. The subjective audiometry is performed in a quiet audio lab and the room reverberation time is 0.5 s. The equipment is a kind of 5.1-channel stereo device, with a dynamic range of 86 dB, and the equipment is placed according to the 5.1-channel schematic diagram, as shown in Figure 6. L is the left channel, R is the right channel, C is the center channel, LFE is the low frequency enhanced channel, SL is the left surround channel, and SR is the right surround channel. The length of each piece of measured audio is 10 s. The interval test is performed between the reference audio and the measured audio and repeated three times to prevent misjudgment. The total time of a complete test is controlled within 15 min to 20 min, which



is also aimed at preventing misjudgment due to auditory fatigue. The average score of the MUSHRA test of 50 pieces of multi-channel audio is used as the subjective evaluation of the audio restoration quality. The results are shown in Table 4.



**Figure 6.** The 5.1-channel schematic diagram.

**Table 4.** Average MUSHRA test scores of multi-channel audio.

Missing Data Rate	TC-TNN	TC-SNN	CP-WOPT	RMC
15%	95	88	85	81
30%	92	85	80	78
45%	88	81	76	73
60%	83	75	71	67
75%	76	70	65	59

It can be seen from Table 4 that, in the case of various missing data rates, the MUSHRA test score of the TC-TNN algorithm is the highest among several completion algorithms, and the MUSHRA test scores of the TC-SNN algorithm and the CP-WOPT algorithm are also in the middle, which is consistent with the objective evaluation. For all these completion algorithms, the MUSHRA test score decreases as the missing data rate increases, which means that there is an inverse relationship between the audio recovery quality and the missing data rate. In particular, when the missing data rate is more than half, the audio recovery quality drops sharply. For the reason that when the data are lost too much, the structural correlation is weakened and it becomes difficult to mine the connection between the lost data and the known data, the multi-channel audio signal cannot be well recovered and the audio recovery quality will eventually decline.

## 5. Conclusions

In the field of audio signal processing, audio restoration tasks have attracted wide attention. In this paper, the tensor completion algorithm is used to restore the multi-channel audio signal with data loss. First of all, the multi-channel audio signal with data loss is

constructed as a third-order tensor, after signal preprocessing and time–frequency transformation. Afterwards, the audio recovery is carried out using the completion algorithm based on the tensor nuclear norm, and the optimal solution of the tensor completion convex optimization model is obtained by using convex relaxation technology. Then, the completed audio tensor is obtained. At last, it will be converted into multi-channel audio. The results of the experiment are compared to the experimental results of the traditional matrix completion algorithm, based on the objective and subjective indicators. It can be seen from the experimental results that the tensor completion algorithm is better able to recover audio signal with data loss and has a higher recovery ability compared to the traditional method. The tensor completion algorithm models the problem of audio data recovery using a mathematical model and optimizes the model to solve the global value, so as to achieve the purpose of data recovery. The tensor completion method provides a new way to recover the lost data of multi-channel audio and effectively improves the quality of the recovered audio. Therefore, the tensor completion algorithm has a good application prospect in the field of audio signal processing.

**Author Contributions:** Conceptualization, L.Z. and L.Y.; methodology, L.Z.; software, L.Y. and Y.G.; validation, L.Z., L.Y. and Y.G.; formal analysis, L.Y. and Y.G.; investigation, D.N. and D.Z.; resources, D.N.; data curation, D.Z.; writing—original draft preparation, L.Z.; writing—review and editing, L.Y. and Y.G.; visualization, L.Z.; supervision, L.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Natural Science Foundation of China (NO. 62161040), the Science and Technology Project of Inner Mongolia Autonomous Region (NO. 2021GG0023), the Program for Young Talents of Science and Technology in Universities of Inner Mongolia Autonomous Region (NO. NJYT22056), the Natural Science Foundation of Inner Mongolia Autonomous Region (NO. 2021MS06030), the Fundamental Research Funds for Inner Mongolia University of Science and Technology (NO. 2023RCTD029), and the Science and Technology Project of Inner Mongolia Autonomous Region (NO. 2023YFSW0006).

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors, on request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Palenchar, J. Dolby Atmos, DTS:X, Multiroom Come to More Soundbars. *TWICE* **2017**, *32*, 1–2.
2. Deng, G. Change from simple to Complex 20 Years of History of Audio-visual technology (Multi-Channel Technology). *Home Theater Tech.* **2017**, *237*, 10–12.
3. Adler, A.; Emiya, V.; Jafari, M.G.; Elad, M.; Gribonval, R.; Plumbley, M.D. Audio Inpainting. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 922–932. [[CrossRef](#)]
4. Yang, L.; Wang, J.; Xie, X.; Zhao, Y.; Kuang, J. Low Rank Tensor Completion for Recovering Missing Data in Multi-channel Audio Signal. *J. Electron. Inf. Technol.* **2016**, *38*, 394–399.
5. Adler, A.; Emiya, V.; Jafari, M.G.; Elad, M.; Gribonval, R.; Plumbley, M.D. A Constrained Matching Pursuit Approach to Audio Declipping. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; pp. 329–332.
6. Godsill, S.H.; Rayner, P.J.W. *Digital Audio Restoration: A Statistical Model Based Approach*; Springer-Verlag: Berlin/Heidelberg, Germany, 1998.
7. Shi, J.; Zheng, X.; Wei, Z.; Yang, W. Survey on algorithms of low-rank matrix recovery. *Appl. Res. Comput.* **2013**, *30*, 1601–1605.
8. Wang, J.; Xie, X.; Kuang, J. Microphone array speech enhancement based on tensor filtering methods. *China Commun.* **2018**, *15*, 141–152. [[CrossRef](#)]
9. Sedighin, F.; Cichocki, A.; Yokota, T.; Shi, Q. Matrix and Tensor Completion in Multiway Delay Embedded Space Using Tensor Train, with Application to Signal Reconstruction. *IEEE Signal Process. Lett.* **2020**, *27*, 810–814. [[CrossRef](#)]
10. Zheng, H.; Shi, Z.; Zhou, C.; de Almeida, A.L.F. Coarray Tensor Completion for DOA Estimation. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *59*, 5472–5486. [[CrossRef](#)]
11. Yang, J.; Yang, B.; Tang, Z. Research about Link Clustering Algorithm Based on Tensor Analysis. *Appl. Res. Comput.* **2011**, *28*, 833–837.
12. Vlasic, D.; Brand, M.; Pfister, H.; Popovic, J. Face Transfer with Multilinear Models. *ACM Trans. Graph.* **2005**, *24*, 426–433. [[CrossRef](#)]

13. Vasilescu, M.A.O.; Terzopoulos, D. Multilinear subspace analysis of image ensembles. In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 18–20 June 2003; p. II-93. [\[CrossRef\]](#)
14. Tan, H.; Wu, Y.; Feng, G.; Wang, W.; Ran, B. A New Traffic Prediction Method Based on Dynamic Tensor Completion. *Procedia-Soc. Behav. Sci.* **2013**, *96*, 2431–2442. [\[CrossRef\]](#)
15. Ran, B.; Tan, H.; Wu, Y.; Jin, P.J. Tensor based missing traffic data completion with spatial–temporal correlation. *Phys. A Stat. Mech. Its Appl.* **2016**, *446*, 54–63. [\[CrossRef\]](#)
16. Cichocki, A.; Zdunek, R.; Phan, A.H. *Nonnegative Matrix and Tensor Factorizations*; John Wiley & Sons: Boston, MA, USA; Chichester, UK, 2009; pp. 21–31.
17. Wang, J.; Liu, M.; Xie, X.; Kuang, J. Compression of Head-Related Transfer Function Based on Tucker and Tensor Train Decomposition. *IEEE Access* **2019**, *7*, 39639–39651. [\[CrossRef\]](#)
18. Kilmer, M.E.; Martin, C.D. Factorization strategies for third-order tensors. *Linear Algebra Appl.* **2011**, *435*, 641–658. [\[CrossRef\]](#)
19. Kilmer, M.E.; Braman, K.; Hao, N.; Hoover, R.C. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM J. Matrix Anal. Appl.* **2013**, *34*, 148–172. [\[CrossRef\]](#)
20. Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; Yan, S. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 925–938. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Zhang, Z.; Ely, G.; Aeron, S.; Hao, N.; Kilmer, M. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3842–3849.
22. Li, W.; Xu, A.; Dai, H.; Wang, F. Signal denoising method based on matrix rank minimization and statistical modification. *J. Vib. Shock.* **2015**, *34*, 38–44.
23. Huang, H. Multidimensional Signal Recovery using Tensor Network Decompositions. Master’s Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2021.
24. Yang, L. Study on Multifactor Audio Signal Modeling and Applications Based on Tensor Analysis. Ph.D. Thesis, Beijing Institute of Technology, Beijing, China, 2016.
25. Wang, A.; Song, X.; Wu, X.; Lai, Z.; Jin, Z. Robust Low-tubal-rank Tensor Completion. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3432–3436.
26. Goldfarb, D.; Qin, Z. Robust Low-Rank Tensor Recovery: Models and Algorithms. *SIAM J. Matrix Anal. Appl.* **2014**, *35*, 225–253. [\[CrossRef\]](#)
27. Ding, W.; Sun, Z.; Wu, X.; Yang, Z.; Solé-Casals, J.; Caiafa, C.F. Tensor completion algorithms for estimating missing values in multi-channel audio signals. *Comput. Electr. Eng.* **2022**, *97*, 107561. [\[CrossRef\]](#)
28. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust Principal Component Analysis? *J. ACM (JACM)* **2011**, *58*, 1–37. [\[CrossRef\]](#)
29. Gandy, S.; Recht, B.; Yamada, I. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Probl.* **2011**, *27*, 25010–25028. [\[CrossRef\]](#)
30. Yang, L.; Liu, M.; Wang, J.; Xie, X.; Kuang, J. Tensor completion for recovering multichannel audio signal with missing data. *China Commun.* **2019**, *16*, 186–195.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.