RESEARCH-ARTICLE

Deep Learning-Based Binaural Speech Signal Enhancement Method

CHENG WEI

YANG ZHILING

Published: 14 July 2023

Citation in BibTeX format

SPML 2023: 2023 6th International Conference on Signal Processing and Machine Learning
July 14 - 16, 2023
Tianjin, China

SPML '23: Proceedings of the 2023 6th International Conference on Signal Processing and Machine Learning (July 2023)
https://doi.org/10.1145/3614008.3614012
ISBN: 9798400707575

Deep Learning-Based Binaural Speech Signal Enhancement Method

Cheng Wei*
Xiamen Ocean Vocational College, 361102 Xiamen, China
chengwei@xmoc.edu.cn

Yang Zhiling
Xiamen Ocean Vocational College, 361102 Xiamen, China
yangzhiling@xmoc.edu.cn

ABSTRACT

The binaural speech signal enhancement decomposition structure is usually monolayered, with limited signal enhancement effect and low signal regularity. In such a context, this paper presents an analysis of the design and validation of a deep learning-based binaural speech signal enhancement method. With the method, the speech and noise features can be extracted according to the actual enhancement requirements and criteria, and signal time-frequency multi-stage decomposition and signal waveform reconstruction can be realized in the multi-stage form to break the limitation of signal enhancement effect and build a deep learning signal enhancement model. The IMCRA spectral subtraction combined processing has been adopted to realize signal enhancement. The test results show that after enhancing the selected binaural speech signals, the distortion of the speech signal enhanced could be significantly improved, with the noise side the signals eliminated, interference reduced, and signals enhanced regularly, which indicates that the enhancement method of such signal could achieve better effect in maintaining the stability of signals during the processing procedure and improving the signal frequency, and therefore is of great application value.

CCS CONCEPTS

- **Design and analysis of algorithms**;

KEYWORDS

Enhancement Method, Speech Recognition, Signal Acquisition

ACM Reference Format:
Cheng Wei and Yang Zhiling. 2023. Deep Learning-Based Binaural Speech Signal Enhancement Method. In *2023 6th International Conference on Signal Processing and Machine Learning (SPML) (SPML 2023), July 14–16, 2023, Tianjin, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3614008.3614012

1 INTRODUCTION

Nowadays, binaural speech technology has been widely used in various fields in society, achieving relatively good results. However, as demands increase and standards change, the transmission of binaural speech signals is subject to certain restrictions under specific environments, resulting in weak signals and poor speech quality, which makes it difficult to meet expectations. To solve this problem, signal enhancement methods are designed. Many traditional binaural speech signal enhancement methods, being one-way, is weak in capturing signals and often poorly targeted though being able to realize the expected processing tasks. The signal enhancement effect achieved with such methods is often insignificant [2]. On top of that, one-way and single-layered signal enhancement structures feature low efficiency, with the binaural speech formed often containing noises that would affect normal use. In this light, this paper presents a study of the design and validation of deep learning-based binaural speech signal enhancement methods. By combining the technology with the binaural speech signal enhancement method, the actual enhancement range could be further expanded, improving the strength of signals within complex radio ranges and forming a signal enhancement structure that is more complete, specific, and flexible. Besides, such an approach is more targeted with higher convertibility and could lay a firm foundation for subsequent innovation and application of binaural speech signal enhancement technology [3].

2 DESIGN OF A DEEP LEARNING-BASED BINAURAL SPEECH SIGNAL ENHANCEMENT METHOD

2.1 Extraction of Speech and Noise Features

To enhancement of binaural speech signals, the first thing that must be done is positioning the speech features and existing noises, and then constructing a stochastic enhancement process based on deep learning technologies [4]. Firstly, the acquisition of the speech signal was performed, after which signal amplitude was measured after preprocessing, as shown in Equation 1:

$$H = (1 - k)^2 \times \varphi + dk \tag{1}$$

In Equation 1: H represents the signal amplitude, k represents continuous syllables, φ represents the signal enhancement range, and φ represents controllable enhancement unit value. The signal amplitude can be calculated based on the above setting. According to the changes in the amplitude, the white noise in speech can be measured, the time domain can be divided, the periodicity of signal waveforms can be summarized and analyzed and the signals can be divided within different time ranges [5].

Generally speaking, featuring certain perceptual features, the waveform of the speech signal is continuous, and would change with time [7]. Noise is one of the features of all acoustic signals [8], and would increase or disappear under the direct impact of external environments and specific factors. The sound source, formed randomly, is relatively difficult to control.

[*] Corresponding author.

SPML 2023, July 14–16, 2023, Tianjin, China

ACM ISBN 979-8-4007-0757-5/23/07…$15.00
https://doi.org/10.1145/3614008.3614012

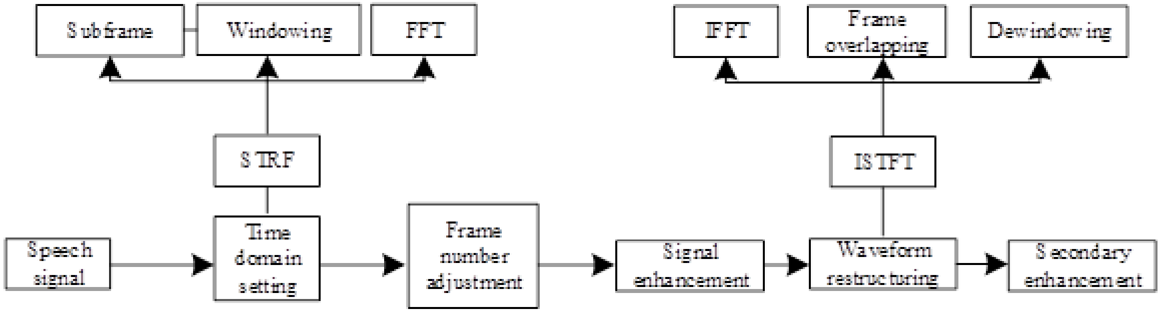

Figure 1: Diagram of signal time-frequency decomposition and signal waveform reconstruction

Table 1: Deep learning signal enhancement model index parameter table

Deep learning signal enhancement model indicator	Directional parameter standard	Measured parameter standard
Sampling frequency/kHz	6.5	8.2
Frame overlapping	192 sampling points	203 sampling points
Conjugate enhancement times	12	18

2.2 Multi-Order Decomposition and Signal Waveform Reconstruction of Signal Time-Frequency

After extracting speech and noise features, deep learning technology was applied to perform signal time-frequency decomposition and signal waveform reconstruction. Featuring certain continuity, the time-frequency domain amplitude spectrum of the binaural speech signal was related to the power spectrum, which should also be processed timely during signal enhancement. The specific nature of the binaural speech signal was analyzed to ensure a strong correlation of the signal in a short time. Personalized speech processing technologies in the deep learning field and the Short Time Fourier Transform (STFT) were used to perform directional transformation against the signal to obtain the distribution state of the binaural speech signal in different frequency backgrounds. STFT was used to divide the signal into different segments and transformed the signal into a windowing short sequence signal, as shown in Figure 1

The processing of signal time-frequency decomposition and signal waveform reconstruction was completed According to FIgure1. The results of the above-mentioned deep learning processing indicate the directional change of the frequency dimension of the binaural speech signal, which could help further maintain the signal stability, minimize frame overlapping, window the processing speed of signal enhancement, and thereby provide a reference future signal enhancement.

2.3 Construction of a Deep Learning Signal Enhancement Model

After signal time-frequency decomposition and signal waveform reconstruction, a binaural speech signal enhancement model was constructed based on deep learning technologies. Since there is always noise in speech signals, the suppression structure was set according to the features of noise to eliminate the noises in the signal, after which a signal mapping program was established in the initial model to establish the enhancement sequence and input the features of binaural speed signals and noise signals into the model, with specific control indexes shown in Table 1

According to Table 1, the index parameters of the deep learning-based signal enhancement model were set to reduce noise, and deep learning technologies were applied to establish a multidimensional classification processing program for the signal. Based on the actual demands and standards of signal enhancement, the structure of the deep learning-based signal enhancement model was constructed, as shown in FIgure 2:

The design and study of the structure of the deep learning signal enhancement model were completed according to Figure 2. The signal features were continuously extracted to establish the mapping between the binaural speech signal and the pure speech feature masking, and the steps of signal enhancement were simplified to improve the enhancement capability of the model.

2.4 Enhancement of Signal through IMCRA Spectral Subtraction Combined Processing

IMCRA spectral subtraction combined processing was performed to enhance the signal after the construction of the deep learning signal enhancement model. Then, based on the established masking mapping relationship and the MCRA spectral subtraction combined correction structure, the enhancement sequence of the binaural speech signal was corrected according to different masking values. After the real-time features were extracted, the spectral subtraction joint correction structure was established, as shown in Figure 3

The design and adjustment of the IMCRA spectral subtraction joint signal enhancement correction structure were completed according to Figure 3. Based on the changes of signal features, IMCRA

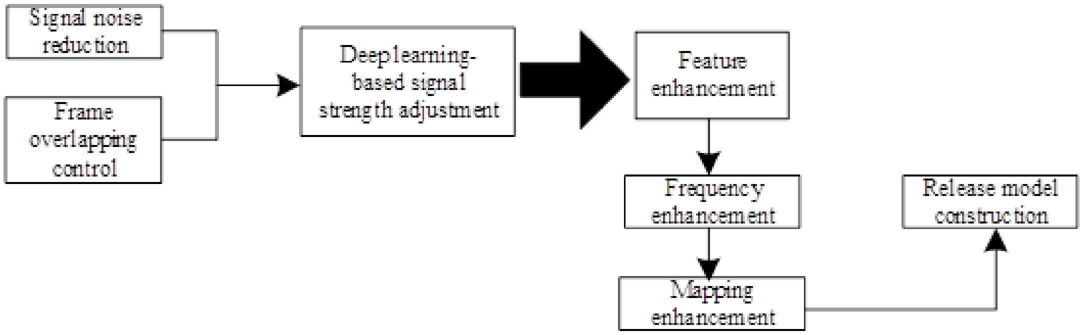
Signal noise reduction
Frame overlapping control
Deep learning-based signal strength adjustment
Feature enhancement
Frequency enhancement
Mapping enhancement
Release model construction

Figure 2: Structural Diagram of the deep learning signal enhancement model

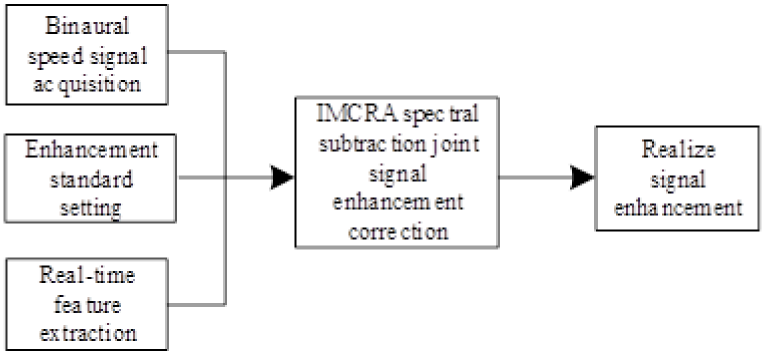
Binaural speed signal acquisition
Enhancement standard setting
Real-time feature extraction
IMCRA spectral subtraction joint signal enhancement correction
Realize signal enhancement

Figure 3: Diagram of IMCRA spectral subtraction joint signal enhancement correction structure

spectral subtraction was performed to realize covered correction and enhancement of signal. After the built-in noise was eliminated or maintained, the probability of signal distortion was minimized.

3 METHOD TESTING

This paper mainly focuses on the analysis and validation study of the application effect of the deep learning-based binaural speech signal enhancement method. Considering the authenticity and reliability of the final test results, a comparative analysis was conducted against the testing results obtained according to the requirements and standards for binaural speech signal enhancement. Then, the initial test environment was developed by applying deep learning technology.

3.1 Test Preparation

The test environment required for the application of the binaural speech signal enhancement method was established and associated with signal enhancement based on deep learning technologies. Firstly, a segment of the binaural speech signal was intercepted as the main target under test to set a basic speech signal processing environment and perform preprocessing against the segment of speech signal based on deep learning technology and ICA technology. In most cases, the intercepted binaural speech signal segments would all contain certain noises, which would not only affect the sound quality but also weaken the signal and lead to negative effects. In this case, the Gaussian white noise signal was used as another input port of ICA to perform hybrid separation processing. Secondly, the FastICA algorithm was adopted to eliminate some of the Gaussian white noises in the signal in advance, with the directional frequency set as 8kHz and the recognition accuracy of the signal being 14 bit. Thirdly, the abnormal data and information existing in the binaural speech signal were collected and the enhancement matrix of the speech signal was established based on deep learning and data mining structure. Refer to Equation 2 for details:

$$W = \begin{bmatrix} -0.79545a & -0.6241 \\ 0.62014 & 0.7451\,(\beta+\delta)^2 \end{bmatrix} \tag{2}$$

In equation 2: W represents the speech signal enhancement matrix, a represents the whitening deviation, β represents the coefficient of convergence, and δ represents the enhancement threshold. The speech signal enhancement matrix was set based on the above settings, while the processing conditions of signal enhancement were established based on the signal enhancement matrix that has been set.

In this part, the real-time signal was collected as the binaural speech changes to monitor frequency changes first. Processing was then conducted through the signal enhancement matrix to establish the enhancement environment, the range of frequency changes and the restricted enhancement area of the desired signal,

Table 2: Basic Signal Enhancement Control Index and Value Setting Table

Basic signal enhancement control indicators	Basic parameter index	Measured parameter index
Enhanced response vector ratio	1.3	1.6
Minimum mean square error	0.21	0.15
Expected signal vector	16.35	18.44
Optimal enhancement factor	5.4	8.1
Signal shock response value	-6.351	-8.114

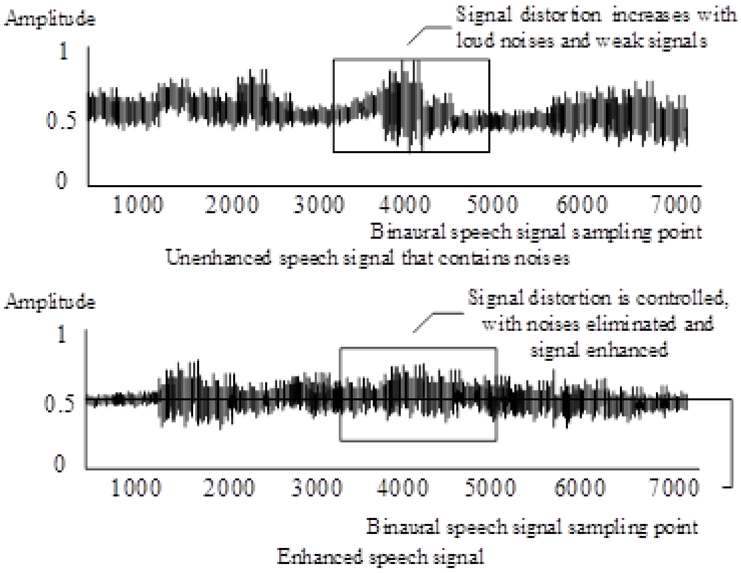

Figure 4: Diagram of the comparison analysis of signal enhancement test results

so as to measure the expected value of signal enhancement and form basic restriction conditions. Then, basic signal enhancement control indicators and values were set based on the aforesaid steps (as shown in Table 2):

Basic signal enhancement control indexes and corresponding values were set according to Table 2. Next, deep learning technology was applied to establish a directional binaural speech signal capturing procedure, correlate with the enhancement structure, and thereby form a complete signal enhancement environment, after which studies of specific measurements were conducted based on deep learning technology.

3.2 Test Procedure and Result Analysis

In the aforesaid test environment, the binaural speech signal enhancement was analyzed and tested based on deep learning technology. Firstly, the signal capture procedure was used to calibrate the internal weak areas to facilitate subsequent enhancement, with the speech adjusted as monaural and undergoing 16-bit quantization. With the help of the speech processing hormone in deep learning, a cyclic signal enhancement sequence was constructed through adaptive beams, and the signal vectors were placed in a 200 × 200 signal enhancement matrix by the column order to realize overlapping noise reduction. Then, based on the aforesaid steps, the amplitude of the signal was measured, multiple sampling points in the binaural speech signal were selected for testing, and the measured signal state was compared with the initial state (as shown in Figure 4):

According to Figure 4, the analysis of the test results was completed. After enhancing the selected binaural speech signal, the speech signal distortion was significantly improved, with the noise inside the signal eliminated, the interference reduced, and the signal enhanced regularly. This indicates that this signal enhancement method could achieve better effects while maintaining the stability of the signal and improving signal frequency, and is of great practical value.

4 CONCLUSION

According to the analysis of the design and study of the deep learning-based binaural speech signal enhancement method, compared with the traditional speech signal enhancement structure, the signal enhancement form constructed in this paper by using the deep learning technology is more flexible and resilient, being more targeted with higher integrity. In complex background environments, a method with an improved embedded speech recognition engine is adopted to improve the recognition rate of binaural speech structure in reverberant environments, while ensuring that the speech signal can be transmitted smoothly, which could significantly improve the speech recognition effect and strengthen the signal.

ACKNOWLEDGMENTS

This paper is supported by Research Projects for high-level talents in Xiamen Ocean Vocational College: Nonlinear Control Method Research of Air-Breathing Hypersonic Vehicles(KYG202101).

REFERENCES

[1] Zhang Tao,Wang Zeyu,Hu Mengxue,Zhao Xin1,et al.2022.A Speech Enhancement Method Based on a Parabolic Center-Microphone Preprocessing and Transfer Learning.Journal of Tianjin University(Science and Technology) (Oct.2022),1053-1060.

[2] Tian Binpeng,Dong Wenfang,Zhang Kun,et al.2022.Deep speech enhancement based on time-frequency mask for propeller interference of rotor aircraft.Telecommunication Engineering,(July 2022),947-952.

[3] Bai Haojun,Zhang Tianqi,Liu Jianxing,Ye Shaopeng.(2022).Speech enhancement combining accurate ratio masking and deep neural network.ACTA ACUSTICA,(May 2022),394-404.

[4] Zhang Pengcheng,Guo Haiyan,Yang Zhen,Yang Yang.(2022).A multi−channel speech enhancement method based on graph post-filtering .Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition)(Apr.2022),66-71.

[5] Zhang Ruiqi,Li Chisheng,Chen Ying.(2021).Improved short-wave signal enhancement method based on generative adversarial network.Modern Electronics Technique(Sep.2021),56-60.

[6] Fang Zhongqi,Wang Shanbing.(2021).Research on Distributed Multichannel Speech Enhancement Algorithm with Colored Noise.Technology Innovation and Application(Apr. 2021),71-73.

[7] Jiang Yixin,Zhang Hongbing.(2021).Application of Speech Signal Enhancement Technology in Speech Recognition.Electronic Technology & Software Engineering(Mar. 2021),70-71.