

Article

Design and Implementation of Machine Tool Life Inspection System Based on Sound Sensing

Tsung-Hsien Liu ¹, Jun-Zhe Chi ², Bo-Lin Wu ², Yee-Shao Chen ², Chung-Hsun Huang ² and Yuan-Sun Chu ^{2,*} ¹ Communications Engineering Department, National Chung Cheng University, Chiayi 62102, Taiwan² Electrical Engineering Department, National Chung Cheng University, Chiayi 62102, Taiwan

* Correspondence: chu@ee.ccu.edu.tw

Abstract: The main causes of damage to industrial machinery are aging, corrosion, and the wear of parts, which affect the accuracy of machinery and product precision. Identifying problems early and predicting the life cycle of a machine for early maintenance can avoid costly plant failures. Compared with other sensing and monitoring instruments, sound sensors are inexpensive, portable, and have less computational data. This paper proposed a machine tool life cycle model with noise reduction. The life cycle model uses Mel-Frequency Cepstral Coefficients (MFCC) to extract audio features. A Deep Neural Network (DNN) is used to understand the relationship between audio features and life cycle, and then determine the audio signal corresponding to the aging degree. The noise reduction model simulates the actual environment by adding noise and extracts features by Power Normalized Cepstral Coefficients (PNCC), and designs Mask as the DNN's learning target to eliminate the effect of noise. The effect of the denoising model is improved by 6.8% under Short-Time Objective Intelligibility (STOI). There is a 3.9% improvement under Perceptual Evaluation of Speech Quality (PESQ). The life cycle model accuracy before denoising is 76%. After adding the noise reduction system, the accuracy of the life cycle model is increased to 80%.

Keywords: deep learning; DNN; ORM; speech enhancement; machine tools; life period



Citation: Liu, T.-H.; Chi, J.-Z.; Wu, B.-L.; Chen, Y.-S.; Huang, C.-H.; Chu, Y.-S. Design and Implementation of Machine Tool Life Inspection System Based on Sound Sensing. *Sensors* **2023**, *23*, 284. <https://doi.org/10.3390/s23010284>

Academic Editor: Miguel Delgado-Prieto

Received: 20 November 2022

Revised: 15 December 2022

Accepted: 23 December 2022

Published: 27 December 2022



Copyright: © 2022 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

With the development of science and technology, a lot of manpower is replaced by automated equipment, and machine tools can process products more accurately. However, if the equipment is damaged, such as the milling cutter of the machine tool is worn or defective, and the maintenance personnel cannot replace the milling cutter in time, it will lead to the decline and loss of product quality. In the past, the more common maintenance method was reactive maintenance. When the machine is in an abnormal state, a maintenance engineer is dispatched to check the machine state and conduct a fault diagnosis [1]. In order to meet the industry's demand for maintenance technology, the focus of maintenance has gradually developed into technologies such as condition monitoring, predictive maintenance, and early fault diagnosis [1]. Identifying problems early and predicting the life cycle of a machine for early preventive maintenance or updating equipment can avoid costly plant failures or downtime. Milling cutters are an important part of product processing and are often damaged items on machine tools. This paper uses a milling cutter as the object of the experiment. For the milling cutter of the machine tool, image processing or optical instrument detection is generally used, and the vibration method, radiation sensor, or ultrasonic detection are also used. These methods are useful and even accurate. Compared with other sensing and monitoring instruments, sound sensors are inexpensive, portable, and have less computational data. Therefore, this paper proposes to monitor the wear of milling cutters in machine tools by sound, predict the life cycle of milling cutters, make it preventative protection, and replace and repair equipment before problems occur, thereby reducing product loss and maintaining its

accuracy. When recording the milling audio, other noises not related to the milling audio, such as coolers, fans, etc., are also recorded at the same time. In order to avoid the interference of these noises, we developed a noise reduction model applied to milling audio, which can effectively reduce the impact of noise on the life cycle model and obtain more accurate prediction results. Combining these two models, we build a complete life cycle estimation system that can be applied to the operating environment of the machine tool.

The milling process has many details to consider. They may affect life cycle estimation. Eser et al. [2] used milling parameters as input to ANN and used ANN to estimate the surface roughness. Kara et al. [3] tested the effect of hard turning on surface roughness and tool wearing. Regarding the research on life cycle estimation, Li et al. [4] used deep learning and feature extraction to estimate the remaining life of the machine. Chen et al. [5] proposed to use MFCC to capture the features of the punch sound, and then use DNN training to estimate the life cycle of the punch. DNNs are a common practice in speech recognition [6–8] and are also commonly used in life cycle prediction models for machines and tools. Boli et al. [9] analyzed the feasibility of applying mask noise reduction in machine tools. Masking is a common denoising method that is often used as a training target for neural networks. Saleem et al. [10] tested the effect of different masks applied to deep neural networks. Wang et al. [11] proposed that an Ideal Ratio Mask (IRM) can be more objective and more suitable as a training target for DNN. However, the IRM does not consider the correlation between noise and audio, so some studies proposed improvements in this part. Xia et al. [12] proposed to use an Optimal Ratio Mask (ORM) as a training objective for supervised learning to achieve better noise reduction. In this paper, based on masks and DNN, using noise addition and feature extraction, a noise reduction model suitable for machine tools is proposed to improve the estimation of the tool life cycle.

The architecture for the remainder of this article is as follows. In Section 2, we illustrate the establishment of the life cycle model. In Section 3, we illustrate the architecture of the noise reduction model. Experimental results and analysis will be presented in Section 4. Section 5 provides conclusions.

2. Life Cycle Estimation Methodology

As shown in Figure 1, the life cycle prediction system is divided into several parts. First, we record the milling audio. Second, we extract the features of the audio, for this part we use MFCC. In addition, we calculate the correlation coefficient between features and milling times and compare the correlations to find suitable features as the input data of DNN. Finally, after the DNN training is complete, we test the accuracy of the model.

We recorded milling audio in the environment shown in Figure 2 and used the equipment shown in Figure 3. This paper uses Mini 5-Axis-CNC (Figure 3a) as the recording object. As the shaft rotates, it allows precision machining. An aluminum block (Figure 3b) is the machined material. We cut 30 layers on each block. The milling cutter (Figure 3c) is made of tungsten steel, and the cutting shape is spiral. Its side length is 12 mm, its diameter is 4 mm, and the number of teeth is 3. Our milling object is an aluminum block, which is too soft. Milling parameters must be carefully adjusted to suit the material.

This article started testing different milling methods and milling parameters in December 2019 and completed the setup in June 2020. Cutting information is shown in Table 1. After milling, the wear of the milling cutter is difficult to observe with the naked eye. Milling cutter wear must be seen with a microscope. However, the status of the tool can be quickly confirmed by sound. The iPhone is a portable device with easy access and recording capabilities. Using a professional acoustic emission sensor, the recording effect will be better. We use an iPhone 11 as a recorder with a sampling frequency of 44.1 KHz. A voice recorder is located next to the machine tool and it records clear milling audio. The position of the recorder is fixed each time to avoid deviations between recorded audios.

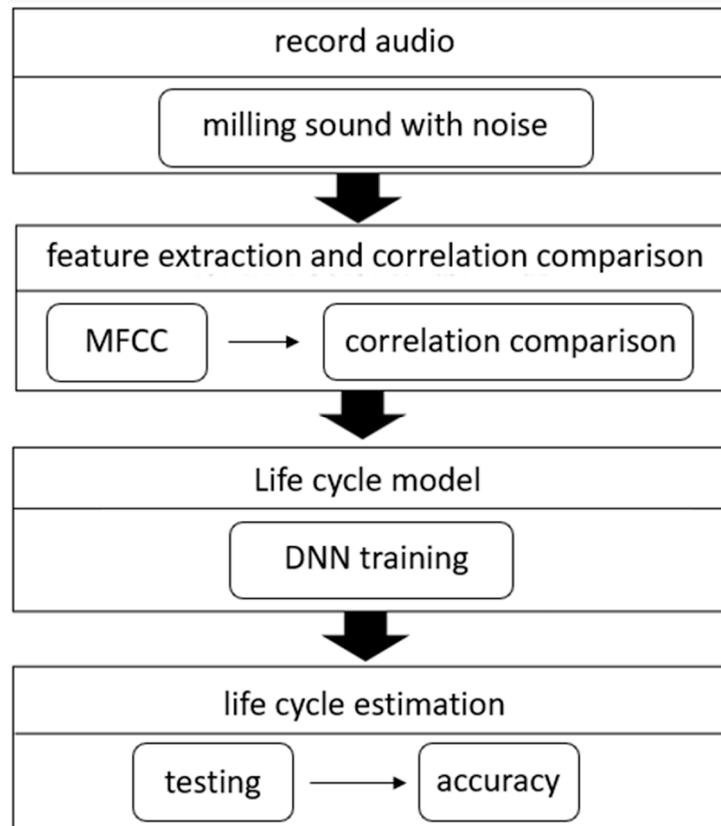


Figure 1. Flow chart of life cycle prediction system.



Figure 2. Audio recording environment.

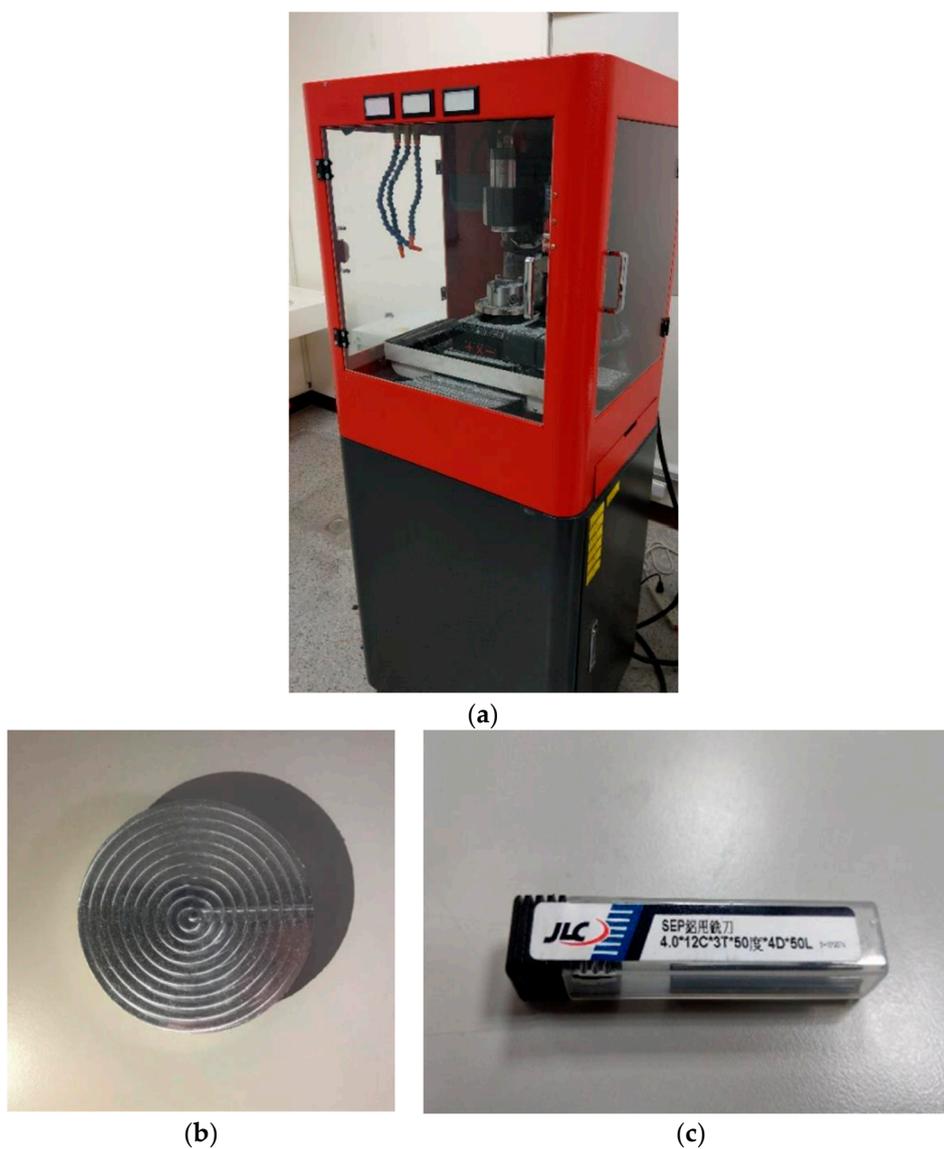


Figure 3. Machine tools and equipment. (a) Mini 5-Axis-CNC small five-axis machine tool; (b) aluminum for cutting; (c) milling cutter for cutting.

Table 1. Cutting information.

Information	Description
Cutting method	Milling
Processing material	aluminum block
Cutting tool	4 mm milling cutter for aluminum
Processing conditions	Dry
Feed rate	500 mm/min
Cutting depth	0.1 mm
Motor revolution	20,000 times/min
Cutting shape	Circle
Tool moving speed	30 mm/s

From September 2020 to May 2021, we recorded milling audio once a week for 5 h. According to the cutting time of the tool, one audio file is recorded every 10 min, and a total of 25 audio files are recorded at a time. We recorded a total of 900 audio files of 5 s each to record the changes in the sound of the tool as the cutting time increased.

There are many algorithms for feature extraction, and MFCC is a more common method [13–16]. It is a speech feature algorithm developed based on human hearing, which can imitate the features obtained by the human ear in different frequency bands.

Figure 4 is a flow chart of the MFCC. In the pre-emphasis stage, the signal will go through a high-pass filter to compensate for the loss of high-frequency signals during the vocalization process. The signal is then split into frames and multiplied by the Hamming window. The Hamming window can solve the spectrum leakage and reduce the discontinuity between the sound frames.

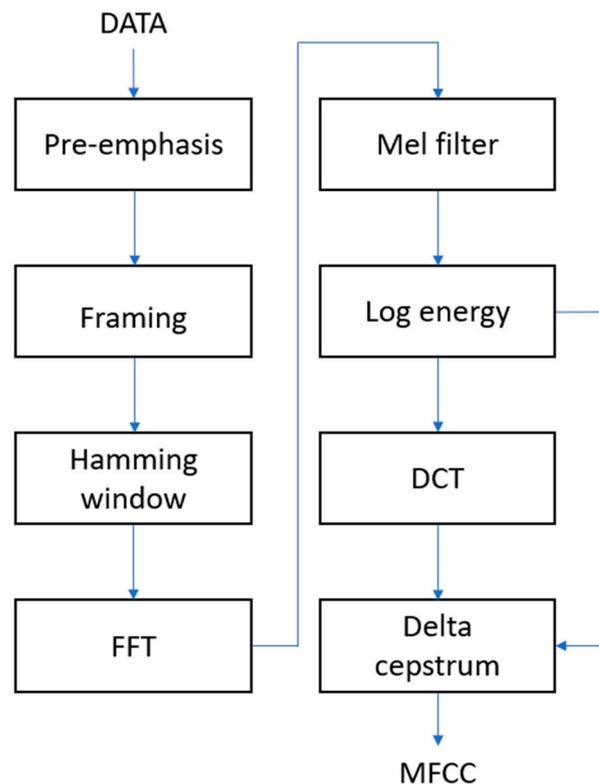


Figure 4. Flow chart of MFCC.

Equation (1) is the formula of the Hamming window. N represents the number of data points in the sound frame, and n is the number of data points. Different smoothing effects can be achieved by adjusting the value of α . After testing, the value of α is set to 0.46.

$$W(n, \alpha) = (1 - \alpha) * \cos\left(\frac{2\pi n}{N - 1}\right), 0 \leq n \leq (N - 1) \quad (1)$$

After pre-processing, the signal is converted from the time domain to the frequency domain through Fast Fourier Transform (FFT). The Mel filter will convert frequencies to Mel-frequencies. Mel-frequency represents the sensitivity of the human ear to frequency, and the relationship with frequency is as follows:

$$m = 2595 \log_{10}(1 + 0.0014f) \quad (2)$$

$$f = 700 \left(10^{0.00038m} - 1\right) \quad (3)$$

Mel filters are implemented in an approximate way to achieve filtering. In the low-frequency band, the filters are densely spaced and have a narrow bandwidth. As the frequency increases, so do the spacing and bandwidth. This simulates the phenomenon that the human ear is more sensitive at low frequencies than at high frequencies.

After the processing of the Mel filter, a set of energy data is obtained and converted into logarithmic energy. MFCC can be obtained after the energy is subjected to discrete cosine transform (DCT). However, current MFCCs only have the static characteristics of speech. In the delta cepstrum stage, we will add dynamic characteristics to the MFCC to enhance the recognition effect.

Gammatone Frequency Cepstral Coefficients (GFCC) and MFCC are very similar, they differ in the filters used. MFCC uses a Mel filter; GFCC uses a Gammatone filter. Gammatone filters are linear filters implemented by impulse responses and are widely used in auditory systems. Its mathematical expression is as follows:

$$g(t) = at^{n-1}e^{-2\pi bt} \cos(2\pi f_c t + \varnothing) \quad (4)$$

where a represents the output gain, n represents the order of the filter, and b represents the length of the impulse response, which determines the bandwidth of the wave filter. f_c represents the filter center frequency, and \varnothing represents the phase angle. Since the human ear is insensitive to phase, the phase is generally set to 0.

PNCC is a feature extraction method for noise processing [17,18]. It carries out data processing before extraction and uses the features of different changing speeds of speech and noise to reduce the interference of noise so that PNCC can describe the features of the target speech more accurately.

After comparing several feature extraction methods, MFCC can better represent the change of tool milling noise corresponding to the number of milling times, so MFCC is finally selected for the life cycle model. Before bringing the features into the DNN training, we use the correlation coefficient to judge the change of the features with the milling times of the tool, find the features with high correlation, and then bring these features into the DNN [19].

Equation (5) is a formula for the correlation coefficient. N represents the number of data points. If the MFCC feature (Y) becomes larger as the number of millings (X) increases, the correlation coefficient (r) will be positive, and vice versa. In order to facilitate identification, this paper will take the absolute value of the correlation coefficient, and then find out the features that will change with the increase of milling times.

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}} \quad (5)$$

According to the milling layers, we divided the life of the milling cutters into three cycles. The first cycle milled the fewest layers and the third cycle milled the most. We divide the audio into three parts, 1~300 audio files are in the first cycle, 301~600 audio files are in the second cycle, and 601~900 audio files are in the third cycle. After the audio is marked with a label, it will be divided into a training set and a test set. Figure 5 is the architecture diagram of DNN. According to the correlation coefficient, we select some of the MFCC features as the input of the DNN. The DNN has three hidden layers, each with 64 neurons. The output layer has three neurons, corresponding to the three life cycle stages. Audio files will be categorized into one of the cycles.

In addition, training parameters and the activation function affect the training time and training effect [20]. After testing, the batch size is 10 and the epoch is 300. The ReLU activation function works best. We chose it as the activation function for the life cycle model. After the model was trained, we would calculate the accuracy of the life cycle model. As shown in Figure 6, three colored dots represent audios with different life cycles. The range of values for each cycle is 1. From the first cycle to the third cycle, the range of value is 0 to 3. We calculated accuracy using the number of audios classified in the correct cycle and the total number of audios. Compared with other architectures, this DNN architecture has relatively good accuracy.

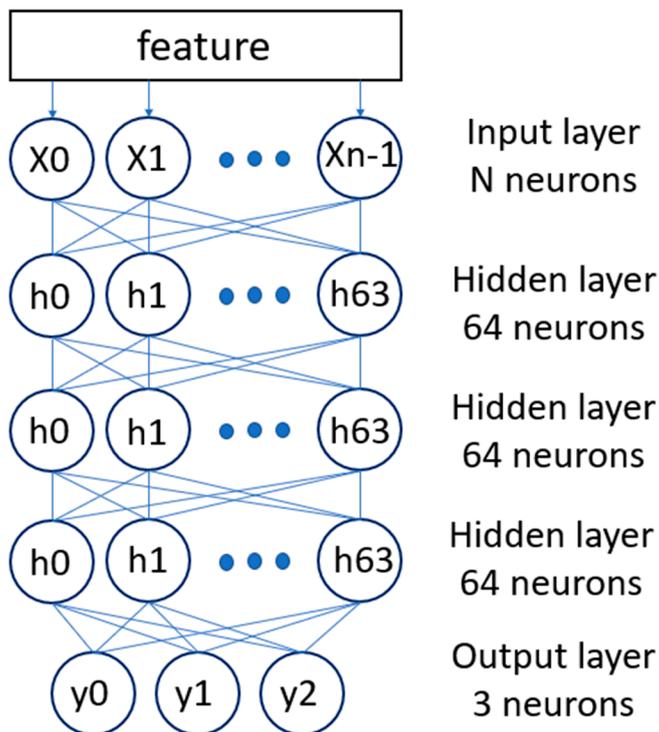


Figure 5. DNN Architecture of Life cycle Model.

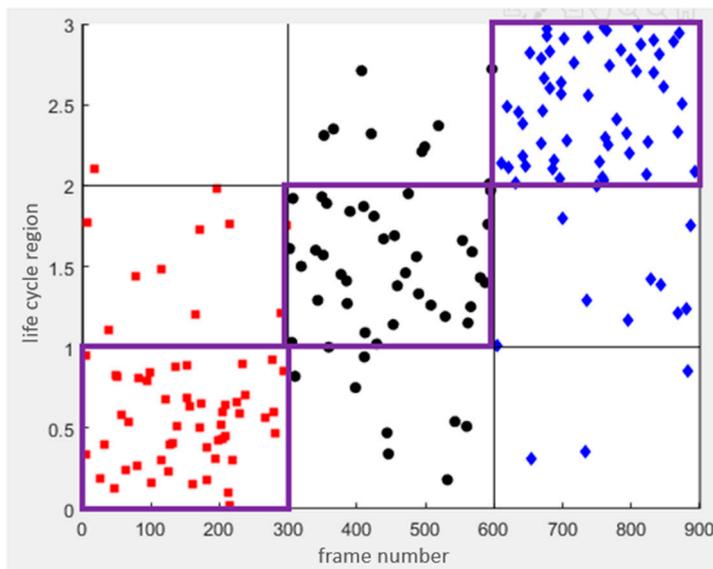


Figure 6. Testing Result of Life Cycle Model.

Table 2 is the correlation between MFCC features and tool loss. First, we judge the 14 features separately and then arrange the features of each sound file according to the order of the milling times of the tool. Finally, we can calculate the correlation between each feature and the number of times the tool has been milled. However, the correlation of some features is very low. In order to avoid features that are not related to the milling times of the tool from affecting the accuracy, the 14 features in this paper are reduced in order from low to high correlation, as shown in Table 3. Finally, it is calculated that the accuracy rate is the highest when 6 features are brought into the tool life cycle model, which can reach 76%. In addition, regarding the ratio of the training set and test set, as shown in Table 4, we tried

90% training and 10% testing, 80% training and 20% testing, and 70% training and 30% testing, and found that 80% training data and 20% testing data has the most accurate data.

Table 2. Correlation between MFCC features and the number of milling layers.

Features MFCC	Original File
m1	0.5061
m2	0.3674
m3	0.5065
m4	−0.6690
m5	0.6595
m6	−0.7395
m7	0.8560
m8	0.6102
m9	0.7605
m10	−0.1400
m11	−0.7142
m12	−0.5461
m13	0.3731
m14	0.7723

Table 3. Tool life cycle model accuracy.

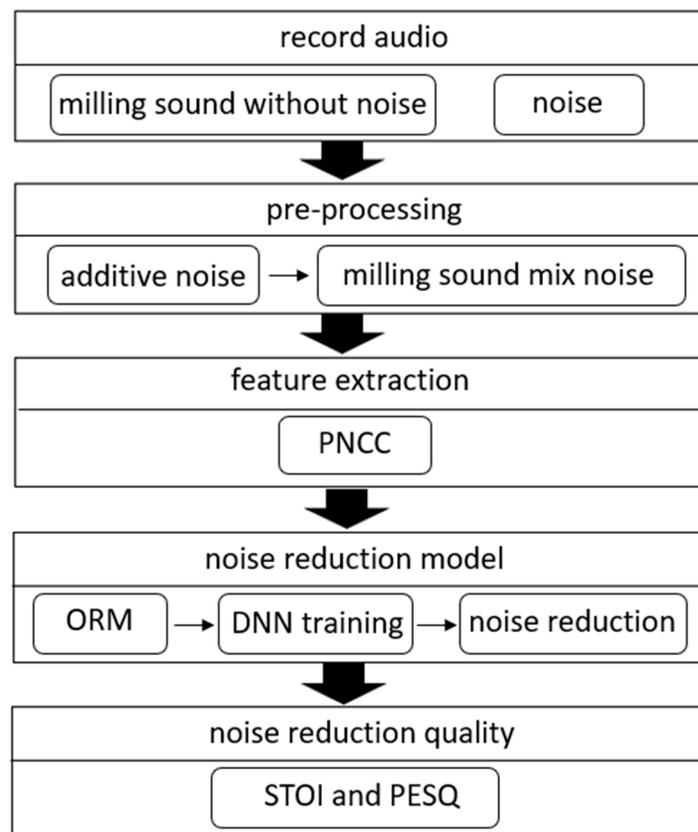
Before Noise Reduction	
Number of Features	Accuracy
1	52%
2	56%
3	58%
4	66%
5	72%
6	76%
7	75%
8	72%
9	68%
10	68%
11	58%
12	55%
13	55%
14	57%

Table 4. Model accuracy under different training data.

Test Data Percentage	Accuracy
Test data (90%) Training materials (10%)	75%
Test data (80%) Training materials (20%)	76%
Test data (70%) Training materials (30%)	73%

3. Noise Reduction Methodology

The audio of milling is similar to the audio of noise, and it is most noticeable in the low-frequency part, and the noise cannot be separated using ordinary low-frequency filters [21–23]. Moreover, the sound of processing sometimes varies greatly, which affects life cycle estimation. In this paper, mask-based supervised speech separation is used to formulate these problems as a supervised learning problem [24,25]. Figure 7 is a flowchart of the noise reduction model. This paper will first record the milling audio and noise, and then use them to generate the Optimal Ratio Mask (ORM) as the learning target of DNN [26,27]. In addition, this paper extracts the feature, PNCC, from the test signal as the input data of DNN. After DNN training is completed, the degree of noise suppression can be determined by features. Finally, we will use the PESQ and STOI to evaluate the effect of noise reduction.

**Figure 7.** Flow chart of noise reduction model.

We used the milling sound recorded without a fan on as the clean audio. Furthermore, we recorded the sound of the cooling fan when the machine tool was not milling as noise and mixed it with the milling sound at 10 db to get a noise-like signal. Finally, we create masks using clean audio and noisy audio. Mask is a commonly used method for noise

reduction, which is usually trained by deep learning. First, add the clean signal to the known noise, after the calculation of the mask, and then enter the deep learning training, after obtaining the mask model of the relationship between the target audio and noise, the speech separation or signal enhancement can be achieved.

A signal (x) with additive noise (n) can be mathematically represented as:

$$y(t) = x(t) + n(t) \quad (6)$$

After Fourier transforming Equation (6), we get Equation (7):

$$Y(t, f) = X(t, F) + N(t, f) \quad (7)$$

Each time-frequency unit is called a TF unit. In the frequency domain, the value of each TF unit falls in the complex domain. From the perspective of amplitude, it is necessary to restore $|X|$, and the Mask matrix should be as follows:

$$M = \frac{|X|}{|Y|} = \frac{|X|}{|X| + |N|} \quad (8)$$

IRM calculates the energy ratio of each TF-unit signal and noise and adjusts the energy of the TF-unit. IRM reflects the degree of noise suppression of each TF unit, which can further improve the quality of the separated audio. The equation of IRM is as follows:

$$IRM(t, f) = \left(\frac{|S(t, f)|^2}{|S(t, f)|^2 + |N(t, f)|^2} \right)^\beta \quad (9)$$

The value range of $IRM(t, f)$ is between 0 and 1, and β is usually 0.5. $|S(t, f)|^2$ and $|N(t, f)|^2$ are the signal and noise energy values in the TF unit.

ORM is an improvement of IRM, which calculates the mean square error (MSE) of the target audio $S(t, f)$ and the masked audio $\hat{S}(t, f)$. Equation (10) represents the transformation between $S(t, f)$ and $\hat{S}(t, f)$, and $\gamma(t, f)$ represents ORM. Equation (11) is the calculation of MSE.

$$\hat{S}(t, f) = \gamma(t, f)[S(t, f) + N(t, f)] \quad (10)$$

$$L(t, f) = \frac{1}{T} \sum |\hat{S}(t, f) - S(t, f)|^2 \quad (11)$$

Substituting Equation (10) into Equation (11) can get the following equation:

$$L(t, f) = \frac{1}{T} \left(\sum (\gamma(t, f) - 1)^2 S(t, f)^2 + \gamma(t, f)^2 N(t, f)^2 + \sum 2\gamma(t, f)(\gamma(t, f) - 1) \mathcal{R}(S(t, f)N^*(t, f)) \right) \quad (12)$$

In Equation (12), the superscript * denotes the conjugate operator, and $\mathcal{R}(\cdot)$ returns the real part of the complex number. Partially differentiate $L(t, f)$ with respect to $\gamma(t, f)$ and let $\partial L(t, f) / \partial \gamma(t, f) = 0$ to find the minimum MSE, the result is as in (13).

$$\gamma(t, f) = \frac{|S(t, f)|^2 + \mathcal{R}(S(t, f)N^*(t, f))}{|S(t, f)|^2 + |N(t, f)|^2 + 2\mathcal{R}(S(t, f)N^*(t, f))} \quad (13)$$

In Equation (14), because the range of $\gamma(t, f)$ value is $(-\infty, +\infty)$, the hyperbolic tangent function (\tanh) formula is added to converge the ORM value, and K can be adjusted with the value of c to face different noise sources and target audio.

$$ORM(t, f) = K \frac{1 - e^{-c\gamma(t, f)}}{1 + e^{-c\gamma(t, f)}} \quad (14)$$

ORM is the result of calculating the minimum MSE, and the complex part is also added to the calculation, so compared with IRM, the details are clearer, and the relationship between audio and noise can be subtler.

After feature extraction, the noisy signal will be input into the DNN as the basis for noise reduction. The feature extraction method will affect the effect of the noise reduction system. Figure 8 is the flow chart of PNCC. In the pre-processing stage, the signal will go through FFT and Gammatone Filter Bank to convert it into features similar to the human auditory model.

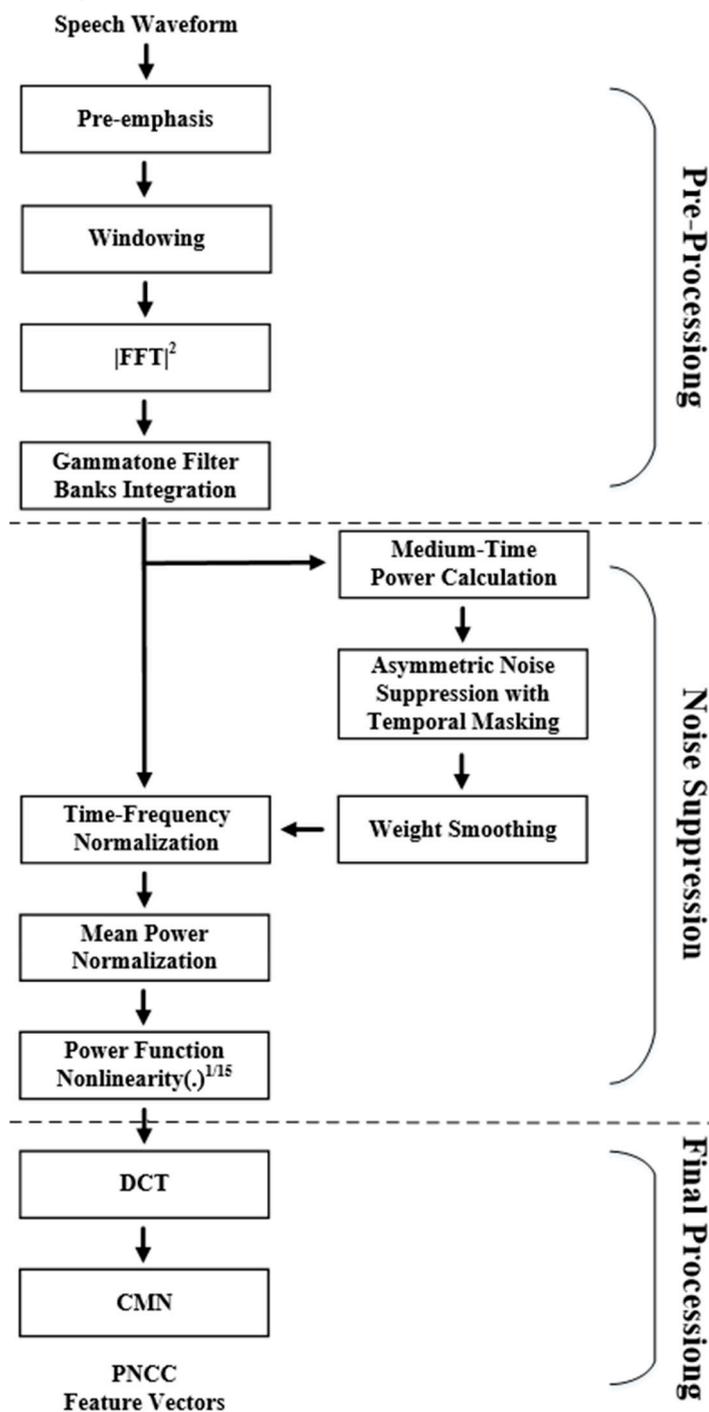


Figure 8. Flow chart of PNCC.

In the noise suppression stage, PNCC undergoes more complex processing to reduce the impact of noise. In the part of the medium-time power calculation, because the noise

energy changes slower than the speech, a larger window can be used to obtain better results. Then, the background noise is suppressed by Asymmetric Noise Suppression, where temporal masking can reduce the influence of reverberation. Spectral weight smoothing enables speech enhancement and noise compensation, while mean power normalization reduces the effects of amplitude scaling. In addition, the energy is processed by an exponential nonlinear function, making it more in line with human auditory perception.

After the noise suppression stage, DCT is used to convert it into acoustic parameters, and then Cepstral Mean Normalization (CMN) is performed on the features to make the distribution between the data more similar, and the PNCC operation is completed. Compared with other feature extraction methods, PNCC handles noise more completely, and it has better effects on audio with noise.

After the signal is denoised, it is necessary to score the denoising quality. However, there is no corresponding database and scoring method for the sound of the machine tool. Therefore, this paper uses the scoring mechanisms of speech, STOI and PESQ, as the scoring method for machine tool noise reduction.

PESQ is a common method for evaluating speech quality. It compares the difference between the original input signal and the processed signal and uses the MOS scoring method to quantify the difference between the ideal speech model and the actual output signal. As shown in Figure 9, the PESQ scoring method corresponds to the MOS score of 4.5 to -0.5 , representing the best and worst scores, respectively.

User Satisfaction	MOS
Very Satisfied	4.4
Satisfied	4.0
Some User Dissatisfied	3.6
Many User Dissatisfied	3.1
Nearly ALL User Dissatisfied	2.6
Not Recommended	1.0

Figure 9. PESQ Quality Chart.

STOI is a measure of speech intelligibility. STOI is scored by comparing the clean speech and the speech to be evaluated, and the value range is 0–1. The higher the value, the better the voice quality.

Tables 5 and 6 are the comparison of different feature extraction methods and the noise reduction model trained by the mask. It can be found that the combination of PNCC and

ORM has the best effect under the PESQ and STOI scores. Finally, we choose PNCC and ORM as the input of DNN.

Table 5. Feature Extraction Method and Mask Comparison 1.

PESQ		MFCC	
MASK	IRM	ORM	
Noise decibel	10 dB	10 dB	
before	3.76	3.76	
after	3.82	3.84	
promote	0.06	0.08	
lift rate	1.6%	2.1%	
PESQ		PNCC	
MASK	IRM	ORM	
Noise decibel	10 dB	10 dB	
before	3.76	3.76	
after	3.87	3.91	
promote	0.11	0.15	
lift rate	2.9%	3.9%	
PESQ		GFCC	
MASK	IRM	ORM	
Noise decibel	10 dB	10 dB	
before	3.76	3.76	
after	3.87	3.86	
promote	0.11	0.1	
lift rate	2.9%	2.6%	

Table 6. Feature Extraction Method and Mask Comparison 2.

STOI		MFCC	
MASK	IRM	ORM	
Noise decibel	10 dB	10 dB	
before	0.7963	0.7963	
after	0.8351	0.8496	
promote	0.0388	0.0536	
lift rate	4.8%	2.1%	
STOI		PNCC	
MASK	IRM	ORM	
Noise decibel	10 dB	10 dB	
before	0.7963	0.7963	
after	0.8362	0.8506	
promote	0.0399	0.0543	
lift rate	5%	6.8%	
STOI		GFCC	
MASK	IRM	ORM	
Noise decibel	10 dB	10 dB	
before	0.7963	0.7963	
after	0.8362	0.8492	
promote	0.0399	0.0529	
lift rate	5%	6.6%	

Figure 10 is the architecture of DNN. Before extracting the signal, the signal will extract the features through the Gammatone Filter and temporarily store them. After

feature extraction, IRM and ORM are used as training targets to perform Label actions. If the training target is IRM, assign a value to the Label according to the ratio of the signal to noise, and the range of the Label is between 0 and 1. If the training target is ORM, the value of Label is between $-K$ and K . Label the T-F Unit captured by each feature, so that the supervised learning effect can be more accurate according to the value of the Label during the subsequent DNN training. After the DNN obtains the value of the Label and the audio features, it will use the features to make a noise reduction mask suitable for the domain, which is equivalent to the inverse conversion of the Gammatone Filter, and the noise reduction is completed.

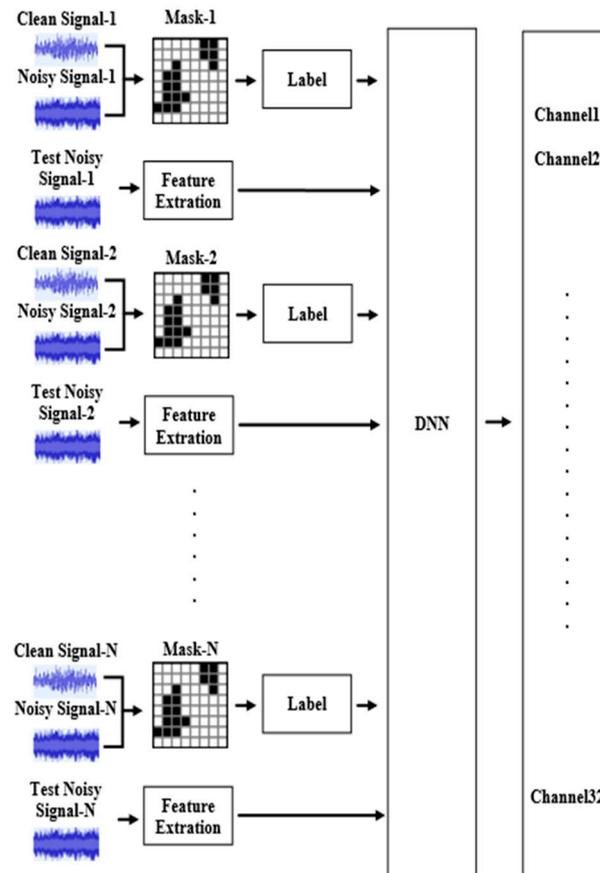


Figure 10. DNN training Mask architecture diagram.

Figure 11 shows the DNN training process of 1 channel. DNN will be trained according to the features of the label and the noise-added signal, and obtain a noise reduction mask suitable for audio. Each DNN corresponds to a frequency band, and 5 sound frames are learned at a time. Combining the DNNs of multiple channels together can achieve a mask suitable for complete audio and complete noise reduction.

Table 7 shows the noise reduction effect of different hidden layers and Gammatone Filter channels. It can be found that 4-layer hidden layers and 64-channel Gammatone Filter will have better training results. Therefore, the noise reduction model finally selects the 64-channel Gammatone Filter to extract features and establishes 4 hidden layers with a total of 1024 neurons. We also tested different parameters for the noise reduction model. We use ReLU as an activation function. The learning rate is 0.001, the epoch is 350 and the batch size is 32.

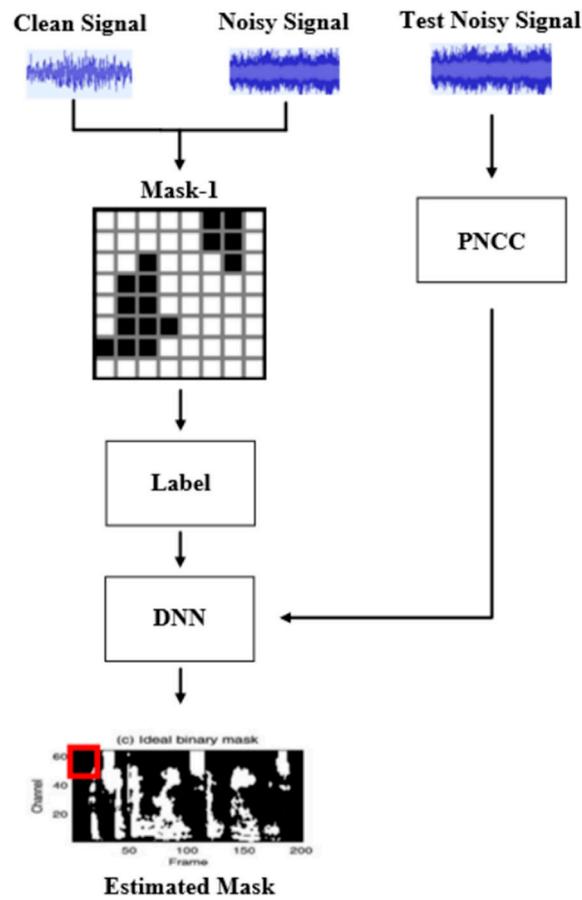


Figure 11. DNN trained per channel.

Table 7. Scoring of DNN Layer and Channel Selection.

STOI		3 Layers	
Aisle	32	64	128
Promote	0.0187	0.0193	0.0193
PESQ		3 layers	
Aisle	32	64	128
Promote	0.07	0.12	0.11
STOI		4 layers	
Aisle	32	64	128
Promote	0.0533	0.0543	0.0543
PESQ		4 layers	
Aisle	32	64	128
Promote	0.09	0.15	0.12
STOI		5 layers	
Aisle	32	64	128
Promote	0.0549	0.0519	0.234
PESQ		5 layers	
Aisle	32	64	128
Promote	0.07	0.09	0.09

Table 8 shows the noise reduction models trained with different data ratios. It can be found that the training results of 70% of the training set and 30% of the test set will have a better noise reduction effect.

Table 8. Test data ratio selection.

STOI		PNCC_DNN ORM	
Ratio	Training (80%) Testing (20%)	Training (70%) Testing (30%)	
SNR (dB)	10	10	
Before	0.7963	0.7963	
After	0.8327	0.8506	
Promote	0.0364	0.0543	
Lift rate	4.5%	6.8%	
PESQ		PNCC_DNN ORM	
Ratio	Training (80%) Testing (20%)	Training (70%) Testing (30%)	
SNR (dB)	10	10	
Before	3.76	3.76	
After	3.89	3.91	
Promote	0.13	0.15	
Lift rate	3.4%	4%	

4. Experimental Results and Discussion

It can be seen from Table 9 that the noise reduction system in this paper can effectively reduce the noise whether the noise is larger or the target signal is larger. It is obvious from the scores of STOI and PESQ that the noise reduction system in this paper can reduce the noise when the SNR is -4 dB to 4 dB.

Table 9. Noise reduction effect of different noises.

STOI		PNCC_DNN ORM	
Noise source		Cooling fan	
SNR (dB)	-4	0	4
Before	0.1948	0.3271	0.4873
After	0.5094	0.5895	0.6810
Promote	0.3146	0.2642	0.1937
Lift rate	162%	80%	40%
PESQ		PNCC_DNN ORM	
Noise source		Cooling fan	
SNR (dB)	-4	0	4
Before	1.312	1.495	1.189
After	2.233	2.233	2.861
Promote	0.920	0.73	1.042
Lift rate	70%	49%	57%

The SNRs of the audios in Table 10 are all -4 db, using PNCC to extract features. It can be found that compared with IRM, ORM has better results as the training target of the noise reduction model. The noise added to the noise, whether it is mechanical sound or human voice, has been significantly improved after the noise reduction system, so the noise reduction system in this article has a good noise reduction effect in different environments.

Table 10. Different ratio of noise reduction effect.

Machine Tool Cutting Sound (Noise Decibel −4 dB) PNCC				
Noise Source	Cooler		Factory Noise	
Mask	IRM	ORM	IRM	ORM
Before	1.512	1.512	1.233	1.233
After	2.499	2.805	2.466	2.820
Promote	0.987	1.293	1.233	1.587
Lift rate	65%	86%	100%	129%
Noise source	Human voice		White noise	
Mask	IRM	ORM	IRM	ORM
Before	1.145	1.145	1.421	1.421
After	2.558	2.835	2.456	2.807
Promote	1.413	1.690	1.035	1.386
Lift rate	123%	148%	73%	98%

As shown in Figure 12, before using milling audio for life cycle estimation, the audio will be passed through a noise reduction system to eliminate the influence of noise, and then feature extraction and correlation coefficient analysis will be performed. Table 11 is the comparison of the correlation coefficients before and after noise reduction. Comparing the correlation coefficients between the 14 features and the milling times of the tool, it can be found that m1, m2, m4, m5, m6, m7, m8, m9, and m10 have all improved. This can know that when reducing noise, the change of the sound can be found from the features, and a better effect can be obtained.

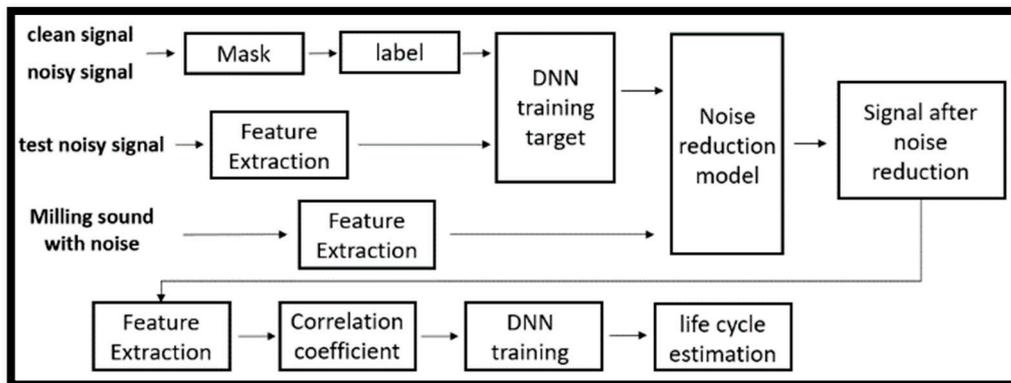


Figure 12. Flow chart of life cycle estimation system after noise reduction.

Table 11. Comparison of correlation coefficients between MFCC and milling times after noise reduction.

Features MFCC	Original File	After Noise Reduction	Lift Rate
m1	0.5061	0.5633	11%
m2	0.3674	0.8316	126%
m3	0.5065	0.3541	−15%
m4	−0.6690	−0.7334	9%
m5	0.6595	0.6913	5%
m6	−0.7395	−0.7592	3%
m7	0.8560	0.8613	0.5%
m8	0.6102	0.6888	12%
m9	0.7605	0.7721	1.5%
m10	−0.1400	−0.1717	3%
m11	−0.7142	−0.2020	−71%
m12	−0.5461	−0.1635	−70%
m13	0.3731	0.3650	−2%
m14	0.7723	0.4494	−41%

As shown in Table 12, the accuracy can be increased to 0.8 when 6 features are selected after noise reduction. It can be seen that the selection of the number of features and the existence of noise will indeed affect the judgment of the tool milling sound in the life cycle model. Therefore, this paper finally chooses to use 6 features as the number of features for judging the life cycle of the tool and uses the noise reduction system to increase the accuracy.

Table 12. The accuracy of the tool life cycle model after noise reduction.

Before Noise Reduction		After Noise Reduction		Lift Rate
Number of Features	Accuracy	Number of Features	Accuracy	
1	52%	1	61%	17%
2	56%	2	61%	9%
3	58%	3	64%	10%
4	66%	4	68%	3%
5	72%	5	77%	7%
6	76%	6	80%	5%
7	75%	7	79%	5%
8	72%	8	77%	7%
9	68%	9	77%	13%
10	68%	10	72%	6%
11	58%	11	68%	17%
12	55%	12	68%	24%
13	55%	13	64%	16%
14	57%	14	66%	16%

5. Conclusions

This paper successfully establishes a machine tool life cycle model based on deep learning and uses the noise reduction model to effectively remove the fan sound contained in the milling sound of the machine tool. Three methods are compared in the life cycle model to capture the sound features of the machine tool. Among them, MFCC has the best performance, and in the noise reduction model, PNCC has the best performance. Under the STOI score, PNCC works well with IRM and ORM. The lift rate reaches 5% and 6.8%, which are higher scores than other methods. PNCC also performs best under the PESQ score. In terms of noise reduction, Mask compares IRM and ORM. Finally, ORM is selected as the target of DNN learning. On the DNN architecture, four hidden layers are selected as the architecture of the entire neural network. It performs better on PESQ and STOI scores than other architectures.

We detect the SNR value of the machine tool site at 10 dB and reduce the noise of tool milling according to this SNR value. The scores of STOI and PESQ have been improved respectively, indicating that this noise reduction model can make noise reduction in line with the ambient sound of the scene. Finally, the noise reduction model is integrated with the tool life cycle model, and noise reduction is carried out according to the actual SNR value of the machine tool site of 10 dB. In this way, the accuracy of the tool life cycle model is increased from 76% to 80%.

There are many different noises in the operation of machine tools. Increasing the types of noise can improve the noise reduction model. Also, when changing the milling parameters, the milling sound also changes. Current lifecycle models test relatively few processing parameters. Testing more parameters may increase the accuracy of life cycle estimation and bring it closer to industrial applications. Finally, the iPhone is not a dedicated recording device. Recording equipment can use industry-standard acoustic emission sensors. The recording effect will then be better, and the life cycle model would be more accurate.

Author Contributions: Conceptualization, Y.-S.C. (Yuan-Sun Chu); Methodology, Y.-S.C. (Yuan-Sun Chu), C.-H.H. and T.-H.L.; Software, J.-Z.C.; Investigation, J.-Z.C.; Writing—original draft preparation,

B.-L.W. and Y.-S.C. (Yee-Shao Chen); Writing—review and editing, B.-L.W., Y.-S.C. and Y.-S.C. (Yuan-Sun Chu); Supervision, Y.-S.C. (Yuan-Sun Chu); Project administration, Y.-S.C. (Yuan-Sun Chu). All authors have read and agreed to the published version of the manuscript.

Funding: This study is funded by National Science and Technology Planning Project. The number is MOST 110-2221-E-194-044.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tsai, Y.T. Applying a Case-Based Reasoning Method for Fault Diagnosis During Maintenance. *Proc. Inst. Mech. Eng. Part C* **2009**, *223*, 2431–2441. [[CrossRef](#)]
2. Eser, A.; Askar Ayyildiz, E.; Ayyildiz, M.; Kara, F. Artificial Intelligence-Based Surface Roughness Estimation Modelling for Milling of AA6061 Alloy. *Adv. Mater. Sci. Eng.* **2021**, *2021*, 5576600. [[CrossRef](#)]
3. Kara, F.; Karabatak, M.; Ayyildiz, M.; Nas, E. Effect of machinability, microstructure and hardness of deep cryogenic treatment in hard turning of AISI D2 steel with ceramic cutting. *J. Mater. Res. Technol.* **2020**, *9*, 969–983. [[CrossRef](#)]
4. Li, X.; Zhang, W.; Ding, Q. Deep learning-based remaining useful life estimation of bearings using multiscale feature extraction. *Reliab. Eng. Syst.* **2019**, *182*, 208–218. [[CrossRef](#)]
5. Chen, P.-J.; Wu, T.-L.; Lin, T.-J.; Lai, Y.-H.; Wu, T.-L.; Yeh, C. Life Period Estimation of Stamping Process Using Punch Sounds and Deep Neural Network. In Proceedings of the 2019 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA), Xi'an, China, 19–21 June 2019; pp. 625–628.
6. Furnon, N.; Serizel, R.; Essid, S.; Illina, I. DNN-Based Mask Estimation for Distributed Speech Enhancement in Spatially Unconstrained Microphone Arrays. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 2310–2323. [[CrossRef](#)]
7. Martin-Doñas, J.M.; Jensen, J.; Tan, Z.-H.; Gomez, A.M.; Peinado, A.M. Online Multichannel Speech Enhancement Based on Recursive EM and DNN-Based Speech Presence Estimation. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 3080–3093. [[CrossRef](#)]
8. Lee, M.-C.; Yeh, S.-C.; Chang, J.-W.; Chen, Z.-Y. Research on Chinese Speech Emotion Recognition Based on Deep Neural Network and Acoustic Features. *Sensors* **2022**, *22*, 4744. [[CrossRef](#)]
9. Boli, W.; Poyi, C.; Junzhe, C.; BoJun, X.; Chu, Y.-S.; Hou, T.-C. Noise Reduction System Applied to Machine Tools. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Penghu, Taiwan, 15–17 September 2021; pp. 1–2.
10. Saleem, N.; Khattak, M.I.; Al-Hasan, M.; Qazi, A.B. On Learning Spectral Masking for Single Channel Speech Enhancement Using Feedforward and Recurrent Neural Networks. *IEEE Access.* **2020**, *8*, 160581–160595. [[CrossRef](#)]
11. Wang, Y.; Narayanan, A.; Wang, D. On Training Targets for Supervised Speech Separation. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 1849–1858. [[CrossRef](#)]
12. Xia, S.; Li, H.; Zhang, X. Using optimal ratio mask as training target for supervised speech separation. In Proceedings of the Asia-Pacific Signal and Information Processing Association 9th Annual Summit and Conference, Kuala Lumpur, Malaysia, 12–15 December 2017. pp. 163–166.
13. Lee, J.; Choi, H.; Park, D.; Chung, Y.; Kim, H.-Y.; Yoon, S. Fault Detection and Diagnosis of Railway Point Machines by Sound Analysis. *Sensors* **2016**, *16*, 549. [[CrossRef](#)]
14. Das, A.; Guha, S.; Singh, P.K.; Ahmadian, A.; Senu, N.; Sarkar, R. A Hybrid Meta-Heuristic Feature Selection Method for Identification of Indian Spoken Languages From Audio Signals. *IEEE Access.* **2020**, *8*, 181432–181449. [[CrossRef](#)]
15. Cabral, F.S.; Fukai, H.; Tamura, S. Feature Extraction Methods Proposed for Speech Recognition Are Effective on Road Condition Monitoring Using Smartphone Inertial Sensors. *Sensors* **2019**, *19*, 3481. [[CrossRef](#)]
16. Al-Ali, A.K.H.; Dean, D.; Senadji, B.; Chandran, V.; Naik, G.R. Enhanced Forensic Speaker Verification Using a Combination of DWT and MFCC Feature Warping in the Presence of Noise and Reverberation Conditions. *IEEE Access.* **2017**, *5*, 15400–15413. [[CrossRef](#)]
17. Kim, C.; Stern, R.M. Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2016**, *24*, 1315–1329. [[CrossRef](#)]
18. Tamazin, M.; Gouda, A.; Khedr, M. Enhanced Automatic Speech Recognition System Based on Enhancing Power-Normalized Cepstral Coefficients. *Appl. Sci.* **2019**, *9*, 2166. [[CrossRef](#)]
19. Nassif, A.B.; Shahin, I.; Attili, I.; Azzeh, M.; Shaalan, K. Speech Recognition Using Deep Neural Networks: A Systematic Review. *IEEE Access.* **2019**, *7*, 19143–19165. [[CrossRef](#)]
20. Erkan, Ö.; Işık, B.; Çiçek, A.; Kara, F. Prediction of Damage Factor in end Milling of Glass Fibre Reinforced Plastic Composites Using Artificial Neural Network. *Appl. Compos. Mater.* **2012**, *20*, 517–536. [[CrossRef](#)]
21. Serizel, R.; Moonen, M.; Wouters, J.; Jensen, S.H. Integrated Active Noise Control and Noise Reduction in Hearing Aids. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2010**, *18*, 1137–1146. [[CrossRef](#)]
22. Schasse, A.; Martin, R. Estimation of Subband Speech Correlations for Noise Reduction via MVDR Processing. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 1355–1365. [[CrossRef](#)]

23. Souden, M.; Araki, S.; Kinoshita, K.; Nakatani, T.; Sawada, H. A Multichannel MMSE-Based Framework for Speech Source Separation and Noise Reduction. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2013**, *21*, 1913–1928. [[CrossRef](#)]
24. Lee, G.W.; Kim, H.K. Multi-Task Learning U-Net for Single-Channel Speech Enhancement and Mask-Based Voice Activity Detection. *Appl. Sci.* **2020**, *10*, 3230. [[CrossRef](#)]
25. Bao, F.; Abdulla, W.H. A Novel Training Target of DNN Used for Casa-Based Speech Enhancement. In Proceedings of the 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC), Tokyo, Japan, 17–20 September 2018. pp. 346–350.
26. Wang, D.; Chen, J. Supervised Speech Separation Based on Deep Learning: An Overview. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 1702–1726. [[CrossRef](#)] [[PubMed](#)]
27. Jiang, W.; Wen, F.; Liu, P. Robust Beamforming for Speech Recognition Using DNN-Based Time-Frequency Masks Estimation. *IEEE Access.* **2018**, *6*, 52385–52392. [[CrossRef](#)]

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.