Wood material recognition for industrial applications

Haibin Fu, Huaping Liu, Xiaoyan Deng and Fuchun Sun

ABSTRACT
Material recognition is an essential problem in the industrial automation community. In this study, we develop a machine learning method to identify the wood material. We extract various feature descriptors for the sound signal and perform a comprehensive comparison with the common classifier. The experimental validations achieve the best feature combination. Besides, we developed a practical stylus to collect sound signal from the wooden samples and present promising preliminary results.

1. Introduction
With the improvement of people's quality of life and the pursuit of natural life, people's love for wood is increasing. In daily life, people use wood and wood composites in large quantities. And the price and quality of different woods vary widely. Therefore, when we purchase wooden furniture, the wood material recognition of furniture is particularly essential. Most research on the multimodal perception of wood material properties has investigated the knowledge of wood material properties of two modalities such as vision—touch, vision—audition, audition—touch, and vision—action (Fujisaki, Tokita, & Kariya, 2015; Kanaya, Kariya, & Fujisaki, 2016). Due to the different densities of different types of wood, their sound by tapping or sliding is also different. Therefore, the use of sound on the recognition of wood has a certain degree of feasibility (Liu, Sun, & Zhang, 2019).

At present, there are many studies on sound recognition. Feature extraction plays an important role in sound recognition. According to the survey, the extraction methods are mainly the zero crossing rate (Ghosal & Chakraborty, 2009), the time and frequency domain analysis (Pai, Deng, & Sundaresan, 2015), the Linear Prediction cepstral coefficients (LPCC) (Yusnita, Paulraj, & Yaacob, 2011), wavelet transform (Feizifar, Haghifam, & Soleymani, 2012), the Mel Frequency Cepstral Coefficients (Ahmad, Thosar, & Nirmal, 2015) and others. The choice of classification algorithms also has a crucial role for sound recognition, such as artificial neural networks (ANN) (Gao, 2012), extreme learning machine (ELM) (Huang, Zhu, & Siew, 2004), Gaussian Mixture Model (GMM) (Mohammadi & Saeidi, 2008), etc. Based on the fast learning speed and computational efficiency, ELM is more flexible and computationally attractive than traditional learning methods (Wei, Liu, Yan, & Sun, 2016).

Section 2 introduces the framework of this paper. Section 3 presents the sound features HMFFC, LPHMFCC and WLPHMFCC extracted from the sound signal, and introduces the evaluation method of the relative importance of the cepstrum component. Section 4 shows the introduction to Extreme Learning Machine. Section 5 describes the data collection, data analysis and the result of the study. The conclusion is shown in Section 6.

2. Framework
At present, there are various kinds of wooden furniture in the market. The price of furniture varies a lot because of the wood used. Therefore, when we were buying furniture, the judgment of furniture wood material is crucial. However, in our life, there are very few people who understand the recognition of wood material. Due to the different densities of different types of wood, their sound by tapping or sliding is also different. We present a method of using sound to recognize wood material. In this paper, we collect two kinds of sound produced by wood, including sliding sound and tapping sound. We use the sound features and ELM classifier for wood material recognition. The sound features include High-order MFCC features, LPHMFCC features and WLPHMFCC features. The
sound features are evaluated by increasing or decreasing component method to improve the utilization of sound characteristics. Then, we try to find the feature combination which has the best accuracy rate. The overall working process is shown in Figure 1.

3. Feature extraction

3.1. Pre-processing

3.1.1. Pre-emphasis

To boost the higher frequencies of the sound signal, we pre-emphasize the sound signal data through a high pass filter (Ahmad et al., 2015). The high pass filter is given by

\[ H(z) = 1 - \mu z^{-1} \]  

and the \( \mu \) is varied between 0.9 and 1.

3.1.2. Framing

We divide the pre-emphasized sound signal into frames to avoid losing the information of the sound signal. The length of time covered by one frame is about 20–30 ms with an overlap of 50% or more.

3.1.3. Windowing

We choose the Hamming window as the windowing function. Hamming window function (Astuti, Sediono, & Aibinu, 2012) is as follows

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

where \( \alpha = 0.46 \), and \( N \) is the number of samples in each frame.

3.2. High-order MFCC feature (HMFCC)

The MFCC features are widely used in the field of sound signal recognition. It combines the generation mechanism of the sound signal and the auditory perception of the human ear, which can reflect the characteristics of sound signal well, and improves the recognition rate of the algorithm (Dixit, Vidwans, & Sharma, 2016). However, the traditional MFCC features only reflect the static characteristics of the sound signal, the dynamic characteristics of the sound signal can be used to express the difference with the static characteristics. Experiments show that the combination of dynamic and static characteristics can effectively improve the recognition performance of the system (Li, Dai, & Fang, 2001). In this paper, we will extract the first difference and second difference of the sound signal static characteristic parameter to express the dynamic characteristics of the sound signal. And the combination of the sound signal static characteristic parameter and the sound signal dynamic characteristic parameter is called high-order MFCC features (HMFCC).

The extraction process of high-order MFCC features is as follows:

Figure 2 represents the extraction process for HMFCC features. The success of HMFCC is due to the use of Mel spaced filter banks processing of the Fourier Transform which provides robustness to the system.
3.2.1. Static characteristic parameter extraction

MFCC is the cepstrum coefficients extracted in the Mel scale frequency domain. The Mel scale describes the nonlinearity of the human ear frequency. The relationship between the Mel frequency and the frequency is expressed by the following equation (3) (Huang, Xiao, & Zhou, 2015):

\[
f_{\text{mel}} = 2595 \times \log \left(1 + \frac{f}{700}\right)
\]  

where \(f\) is the frequency, in Hz; and \(f_{\text{mel}}\) is the Mel scale frequency.

The detailed steps are as follows:

1) Fast Fourier Transform (FFT)

To obtain the spectrum of the sound signal, the FFT is conducted on the short time frame sound signal \(x(n)\), and the equation (Cao, Zhao, & Wang, 2017) is

\[
X_k = \sum_{n=0}^{N-1} x(n)e^{-j2\pi nk/N}, \quad 0 \leq k \leq N-1
\]  

where \(N\) is the number of samples in each frame.

2) The calculation of power spectrum

The discrete power spectrum of sound signal is got by squaring the spectrum of the sound signal, and the following equation is

\[
P_m = |X_k|^2
\]  

where \(P_m\) is the discrete power spectrum.

3) Mel triangular filter bank

The power spectrum \(P_m\) is filtered through the Mel triangular filter bank. And the Mel filter bank includes \(M\) filters, where \(M\) value is usually between 24 and 40. The Mel filter bank \(H_m(k)\) is calculated as follows

\[
H_m(k) = \begin{cases} 
0, & k < f(m-1) \\
\frac{2(k-f(m-1))}{(f(m+1)-f(m-1))}, & f(m-1) \leq k \leq f(m) \\
\frac{2(f(m+1)-k)}{(f(m+1)-f(m-1))}, & f(m) \leq k \leq f(m+1) \\
0, & k > f(m+1)
\end{cases}
\]  

where \(f(m)\) is the centre frequency, \(m = 1, 2, \ldots, 24\), \(k = 1, 2, \ldots, N/2-1\), and \(\sum_{m=0}^{M-1} H_m(k) = 1\).

Figure 3 represents the Mel triangular filter bank.

4) The calculation of the logarithmic spectrum output by the Mel filter bank

\[
Q_m = \ln \left(\sum_{k=0}^{N-1} P_m H_m(k)\right), \quad 0 \leq m \leq M
\]  

where \(Q_m\) is the logarithmic spectrum, \(M\) is the number of filter banks, and \(M = 24\).

5) The discrete cosine transform (DCT) of the logarithmic spectrum \(Q_m\) yields the MFCC coefficients, and the equation is as follows:

\[
C_n = \sum_{m=0}^{N-1} Q_m \cos \left(\frac{\pi n(m - 0.5)}{M}\right), \quad n = 1, 2, \ldots, L
\]  

where \(C_n\) denotes the MFCC coefficients, \(L\) is the order of MFCC coefficients, and \(L\) is usually 12–16. Here, we set \(L\) to be 13.

3.2.2. Dynamic characteristic parameter extraction

1) the first difference of MFCC

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{mel_filter_bank.png}
\caption{The Mel triangular filter bank.}
\end{figure}
The equation of the first difference of MFCC is as follows (Cao et al., 2017):

\[
D_1(t) = \begin{cases} 
C_{(t+1)} - C_{(t)}, & t < k \\
\frac{1}{L} \sum_{k=1}^{L} i \cdot (C_{(t+1)} - C_{(t-1)}), & k \leq t < L - k \\
C_{(t)} - C_{(t-1)}, & k \geq L - k
\end{cases}
\]

(9)

where \(C_{(t)}\) is the \(t\)-th MFCC coefficient for each frame; \(D_{1}(t)\) is the \(t\)-th first difference of MFCC coefficient for each frame. The \(k\) represents the time difference and is set to 2.

(2) the second difference of MFCC

The equation of the second difference of MFCC is as follows (Cao et al., 2017):

\[
D_2(t) = \begin{cases} 
D_{(t+1)} - D_{(t)}, & t < k \\
\frac{1}{L} \sum_{k=1}^{L} i \cdot (D_{(t+1)} - D_{(t-1)}), & k \leq t < L - k \\
D_{(t)} - D_{(t-1)}, & k \geq L - k
\end{cases}
\]

(10)

where \(D_{(t)}\) is the \(t\)-th first difference of MFCC coefficient for each frame; \(D_{2}(t)\) is the \(t\)-th second difference of MFCC coefficient for each frame. The \(k\) represents the time difference and is set to 2.

In summary, the HMFCC is a 39-dimensional feature vector, which is constructed by the MFCC feature, the first difference of MFCC and the second difference of MFCC.

### 3.3. LPHMFCC feature

Based on the calculation of MFCC, the Linear Prediction Coefficients (LPC) are incorporated into the calculation of MFCC coefficients (Zhang, Wang & Xia, 2011). The LPC power spectrum of the sound signal is passed through the triangular filter bank, and then transformed it into the time domain by DCT transform, and the new characteristic parameter LPMFCC is obtained. The LPCMFCC coefficients are used to calculate the first difference and the second difference to get the LPHMFCC coefficients.

The extraction processes of LPHMFCC features is shown in Figure 4, and the extraction algorithm is as follows:

1. Pre-processing: the sound signal is processed by pre-emphasis, framing and windowing.
2. Calculate the LPCC of each frame of the sound signal, and the length of the LPCC is equal to the length of one frame of the sound signal.
3. According to the LPCC to obtain the frequency response of the sound model, and then calculate the power spectrum of the sound signal.
4. Filter processing of the sound signal is performed by the Mel filter bank, and then the logarithmic power spectrum is calculated.
5. The LPMFCC is obtained by the DCT of the logarithmic power spectrum of the sound signal.
6. The LPCMFCC coefficients are used to calculate the first difference and the second difference to get the LPHMFCC coefficients.

### 3.4. WLPHMFCC feature

The traditional Mel frequency cepstrum coefficient assumes that the sound signal is short and stable, obtained with a fixed window Fourier transform. From the principle of uncertainty, this assumption makes the details of the spectrum of the sound fuzzy and loses certain information. The use of discrete wavelet Fourier transform to replace the fast Fourier transform in the extraction process of the traditional MFCC coefficients, can be better to solve this problem.

The experiment uses discrete binary wavelet, its expression (Zhang, Zhang, & Liang, 2008) is:

\[
W_{f}(\phi, \psi) = \int_{R} f(\tau) \phi_{\psi, \phi}(\tau) d\tau
\]

(11)

\(f(\tau)\) is the limited energy signal, and \(\phi_{\psi, \phi}(\tau)\) is a binary wavelet whose expression is as follows:

\[
\phi_{\psi, \phi}(\tau) = 2^{-\phi/2} \varphi(2^{-\phi} \tau - \psi), (\phi, \psi \in Z)
\]

(12)

In this paper, the extraction process LPHMFCC coefficients are improved by discrete wavelet Fourier transform, and we will obtain the new characteristic parameters.
Figure 5. The extraction processes of WLPHMFCC features.

- wavelet LPHMFCC coefficients (recorded as WLPHMFCC).

The flowchart in Figure 5 illustrates the feature extraction processes for WLPHMFCC.

The WLPHMFCC coefficients are obtained by the LPHMFCC coefficients. The improvements are described below (Zhang, Yang & Sun 2011):

1. In the process of feature extraction of WLPHMFCC, we replace FFT with discrete wavelet Fourier transform. N layers wavelet coefficients are obtained by wavelet transform of N-1 layers. And then transform each layer wavelet coefficients to get its spectrum through FFT. For FFT, the number of points for each converted frame must be an integer multiple of 2, so the wavelet coefficients of each layer are automatically zeroed to obtain an integer power of 2. The resolution of the spectrum is improved.

2. After obtaining the spectrum of each layer wavelet coefficients, it is necessary to synthesize the spectrum to obtain the complete spectrum of each frame signal according to its frequency level. The spectrum of the first layer wavelet coefficients is placed at the highest bit of the spectrum, and the other layers are discharged in turn.

3.5. Evaluation method of relative importance of cepstrum component

To find the best combination of HMFCC, LPHMFCC, and WLPHMFCC, we use the increase and decrease component method to evaluate the relative importance of the cepstrum components (Zhou, Li, & Li, 2013). The average contribution rate of each cepstrum component is calculated by the following equation:

$$R(\nu) = \frac{1}{\eta} \left[ \sum_{\theta > \nu} (p(\nu, \theta) - p(\nu + 1, \theta)) + \sum_{\theta < \nu} (p(\theta, \nu) - p(\theta, \nu - 1)) \right]$$

where $R(\nu)$ represents the average contribution of the $\nu$-th order cepstrum component, $\eta$ is cepstrum order number, $p(\nu, \theta)$ indicates the cepstrum coefficient feature recognition rate when order from $\nu$ to $\theta$.

If the average contribution rate is positive, it indicates that the feature tends to increase the recognition rate. The negative contribution rate suggests that the feature tends to reduce the recognition rate. We retain the characteristic parameters of the positive contribution rate and discard the characteristic parameters with negative contribution rate.

4. Classifier development

Most of the feed-forward neural networks use the gradient descent method, which causes slow training speed, weak generalization ability and other shortcomings. The ELM algorithm randomly generates the connection weights between the input layer and the hidden layer and the threshold of the hidden layer neurons, and there is no need to adjust during the training process. Only need to set the number of hidden layer neurons, you can get a unique optimal solution. This algorithm has the advantages of fast learning speed and good generalization ability. In this study, the model of basic ELM is shown in Figure 6.

Suppose there are Q arbitrary samples $(x_j, t_j)$, where $x_j = [x_{j1}, x_{j2}, \ldots, x_{jn}]^T \in \mathbb{R}^n$, $t_j = [t_{j1}, t_{j2}, \ldots, t_{jm}]^T \in \mathbb{R}^m$. Here, $x_j$ represents the feature matrix of HMFCC, LPHMFCC and WLPHMFCC extracted from sound signal. The parameters $n$ and $m$ are the dimensions of input and target vector respectively. In this paper, $n$ represents the
dimension of the sound feature matrix, and \( m \) represents the number of wood materials in this paper. For a hidden layer of a single hidden layer of the neural network (Huang et al., 2004) can be expressed as

\[
\sum_{i=1}^{L} \beta_i g(w_i \cdot x_j + b_i) = o_j, \quad j = 1, 2, \cdots, Q \tag{14}
\]

where \( \beta_i \) is the weight vector between the \( i \)-th hidden node and the output nodes, \( g(x) \) is the activation function, \( w_i \cdot x_j \) represents the inner product of \( w_i \) and \( x_j \), \( b_i \) is the bias of the \( i \)-th hidden node, \( o_j \) is the output vector of the \( j \)-th training sample. The number of hidden nodes in this paper is \( l \).

The goal of single-hidden neural network learning is to minimize the output error and can be expressed as

\[
\sum_{j=1}^{Q} ||o_j - t_j|| = 0 \tag{15}
\]

It means that there are \( \beta_i, w_i \) and \( b_i \) to make

\[
\sum_{j=1}^{L} \beta_i g(w_i \cdot x_j + b_i) = t_j, \quad j = 1, 2, \cdots, Q \tag{16}
\]

It can be expressed as a matrix

\[
H\beta = T \tag{17}
\]

where \( H \) is the output of the hidden layer node, \( \beta \) is the output weight, and \( T \) is the desired output (Huang, Zhu, & Siew, 2006; Yang & Wu, 2016).

\[
H = \begin{bmatrix}
g(w_1 \cdot x_1 + b_1) & \cdots & g(w_l \cdot x_1 + b_l) \\
\vdots & \ddots & \vdots \\
g(w_1 \cdot x_Q + b_1) & \cdots & g(w_l \cdot x_Q + b_l)
\end{bmatrix}_{Q \times l} \tag{18}
\]

\[
\beta = \begin{bmatrix}
\beta_1^T \\
\vdots \\
\beta_l^T
\end{bmatrix}_{l \times m}, \quad T = \begin{bmatrix}
t_1^T \\
\vdots \\
t_Q^T
\end{bmatrix}_{Q \times m} \tag{19}
\]

Traditionally in order to train an SLFN, we need to find \( \hat{w}_i, \hat{b}_i, \beta_i (i = 1, 2, \cdots, l) \) such that

\[
||H(\hat{w}_i, \hat{b}_i) \hat{\beta}_i - T|| = \min_{w,b,\beta} ||H(w_i, b_i) \beta_i - T|| \tag{20}
\]

Figure 7. The 9 different wood (include Toon, Paulownia, Cunninghamia, Firmiana, Tzumu, Pyinkado, Bubinga, Red oak and Red cherry).

Figure 8. The environment and methods of collecting data. First, the overall structure of the collection system. Secondly, collecting impact sound signal. Thereafter, collecting sliding sound signal.
**Figure 9.** The example of the sliding sound signal and tapping sound signal of four kinds of wood (Paulownia, Cunninghamia, Toon and Tzumu).

**Figure 10.** The average contribution of each order cepstrum component for wood classification of HMFFC, LPHMFCC and WLPHMFFC on the sliding sound.
which is equivalent to minimizing the cost function
\[ E = \sum_{j=1}^{Q} \left( \sum_{i=1}^{L} \beta_i g (w_i \cdot x_j + b_i) - t_j \right)^2 \]  
(21)

In the ELM algorithm, once the input weight \( w_i \) and the hidden layer \( b_i \) are randomly determined, the output matrix \( H \) of the hidden layer is uniquely determined. The training single hidden layer neural network can be transformed into solving a linear system \( H\beta = T \). And the output weight \( \beta \) can be determined \( \hat{\beta} = H^+ T \)  
(22)

where \( H^+ \) is the Moore-Penrose generalized inverse of the matrix (Cao & Lin, 2015; Cao, Lin, & Huang, 2012; Huang, Zhou, & Ding, 2012).

### 5. Experiments and analysis

#### 5.1. Data collection

In this experiment, we collect the sliding sound signal and the impact sound signal of 9 different wood surfaces, in which each sliding sound signal duration of 4 s and impact sound signal duration of 1 s. Figure 7 shows the

---

**Table 1.** The accuracy rate of the original features and processed features on the sliding sound.

| Features    | Original features | Processed features |
|-------------|-------------------|--------------------|
|             | Average(%) | Max(%) | Average(%) | Max(%) |
| HMFCC       | 28.89       | 31.11 | 49.50       | 52.22 |
| LPHMFCC     | 28.79       | 32.22 | 35.25       | 37.78 |
| WLPHMFCC    | 47.37       | 50.00 | 51.82       | 53.33 |
9 different wood which includes Toon, Paulownia, Cunninghamia, Firmiana, Tzumu, Pyinkado, Bubinga, Red oak and Red cherry. In the experiment, we use the MC108 microphone to collect the sound signal. The sensitivity of the microphone sensor is $-47\text{dB} \pm 4\text{dB}$. In this study, the sampling frequency of the recorded sound data is 44100 Hz, and the sampling accuracy is 16 bit. To improve the robustness of the experiment, when the sound signal is collected, we use the different scan force and scan velocities to slide on the wood surface, and the position of the tap on the wood surface is not fixed. In Strese, Schuwerk, & lepure, 2016, the range of scan force is 0–3N, and the range of scan velocity is 0–400 mm/s. Figure 8 shows the environment and method for collecting data. In this study, we use the Python language programme to collect sound data in the Ubuntu 14.04 system environment.

For each kind of wood, we collected 100 samples for each of the sliding and tapping sound, 50 of which are used as the training set and 50 as the testing set. We use different wood blocks of the same wood to collect training set and testing set. In this experiment, the pre-emphasis coefficient is 0.97, the length of one frame is 256, and the length of a frame overlap is 128.

5.2. Experiment result

In this experiment, we collected 1800 sound data samples including training set and testing set. Figure 9 shows the normalized amplitude example of the sliding sound signal and the tapping sound signal of four kinds of wood of Paulownia, Cunninghamia, Toon and Tzumu. From Figure 9, we can find that the tapping sound signal of

![Figure 12](image-url)

**Figure 12.** The average contribution of each order cepstrum component for wood classification of HMFC, LPHMFC and WLPHMFC on the tapping sound.
Paulownia and Cunninghamia is similar, but the sliding sound signal of Paulownia and Cunninghamia is different. However, the sliding sound signal of Toon and Tzumu is similar, and the tapping sound signal of Toon and Tzumu is different.

We extract HMFCC, LPHMFCC, WLPHMFCC features. From each sound sample, HMFCC, LPHMFCC, and WLPHMFCC are all 39-dimension features. In the next experiment, we find a combination of the three features with the best classification effect.

### 5.2.1. Experiments on sliding sound

We extract HMFCC, LPHMFCC and WLPHMFCC features from each sound sample, and then use the ELM algorithm classifier to classify the wood based on the obtained features. We set the number of hidden layer neurons from 1 to 3001, where each interval 100 take a value. In the

| Features | Original features | Processed features |
|----------|------------------|-------------------|
| HMFCC    | Average(%)       | Max(%)            |
|          | 52.73            | 57.78             |
| LPHMFCC  | 63.67            | 66.67             |
| WLPHMFCC | 42.32            | 45.56             |

| Features | Average(%)       | Max(%)            |
|----------|------------------|-------------------|
| HMFCC    | 57.07            | 61.11             |
| LPHMFCC  | 64.04            | 68.89             |
| WLPHMFCC | 47.27            | 50.00             |

### Table 2. The accuracy rate of the original features and processed features on the tapping sound.

| Feature origin | Feature combinations | Dimension |
|----------------|----------------------|-----------|
| Slide          | HMFCC + WLPHMFCC     | 56        |
| Tap            | HMFCC + LPHMFCC      | 60        |
|                | HMFCC + WLPHMFCC     | 54        |
| Slide + Tap    | WLPHMFCC + LPHMFCC   | 52        |
following, this paper will describe the classification of the obtained features and the features processed by the increase and decrease component method.

Figure 10 shows the average contribution of each order cepstrum component for wood classification of HMFCC, LPHMFCC and WLPHMFCC on the sliding sound.

According to the increasing and decreasing component method, the average contribution of each cepstrum component for wood classification of HMFCC, LPHMFCC and WLPHMFCC are calculated. Finally, we select 25 orders HMFCC cepstrum components with the highest contribution, 23 orders LPHMFCC cepstrum components and 31 orders WLPHMFCC cepstrum components. Figure 11 reflects the tendency of the accuracy rate of the features before and after processing on the sliding sound.

Table 1 shows the maximum values and the average values of the accuracy rate of the original features and processed features on the sliding sound. It can be seen from the above analysis that the accuracy of the classification has been improved after the feature has been processed by the increase and decrease component method on the sliding sound.

5.2.2. Experiments on tapping sound
As with the method of sliding sound processing, we extract HMFCC, LPHMFCC and WLPHMFCC features from the tapping sound, and then use the ELM algorithm classifier to classify the wood based on the obtained features. Figure 12 shows the average contribution of each order cepstrum component for wood classification of HMFCC, LPHMFCC and WLPHMFCC on the tapping sound.

According to the increasing and decreasing component method, we select 31 orders HMFCC cepstrum components with the highest contribution, 29 orders LPHMFCC cepstrum components and 23 orders WLPHMFCC cepstrum components. Figure 13 reflects the tendency of the accuracy rate of the features before and after processing on the tapping sound.

Table 2 shows the maximum values and the average values of the accuracy rate of the original features and processed features on the tapping sound.

5.2.3. Feature combination
In the experiment, we tried the following combinations of features. Table 3 shows the forms of the feature combination.

Table 4. The accuracy rate of the combination features.

| Feature origin | Feature combinations | Average | Max |
|----------------|----------------------|---------|-----|
| Slide          | HMFCC + WLPHMFCC     | 52.32   | 55.56|
| Tap            | HMFCC + LPHMFCC      | 77.37   | 81.11|
|                | HMFCC + WLPHMFCC     | 57.47   | 62.00|
|                | LPHMFCC + WLPHMFCC   | 72.63   | 76.67|
| Slide + tap    | WLPHMFCC + LPHMFCC   | 68.79   | 73.33|

Figure 14. The accuracy of different combinations of features.
The classification of these feature combinations is as follows. Figure 14 shows the accuracy of different combinations of features. The maximum values and the average values of the accuracy rate of these combination features is shown in Table 4.

Figure 15 shows the confusion matrix of the best accuracy of the three feature origins. From the above analysis, we can see that the feature combination of HMFCC and LPHMFCC extracted from tapping sound signal can get the best classification effect with the highest accuracy rate of 81.11% and the highest average accuracy rate of 77.37%. When using the sliding sound characteristics to classify, the best accuracy rate is only 55.56%. Among them, the classification of tzumu has the worst accuracy with only 10%. When we combine the WLPHMFCC features of sliding sound with the LPHMFCC features of tapping sound, the best accuracy rate is only 73.33%, which is lower than the feature combination of HMFCC and LPHMFCC extracted from tapping sound. Therefore, the sliding sound is not beneficial for the recognition of wood material.

6. Conclusions
In this study, using the sound to recognize the wood material has certain feasibility on the recognition of wood material. It can also be seen that the wood’s tapping sound is more suitable for the recognition of wood material than the sliding sound of wood. The feature combination is more advantageous than the single feature in the recognition process. Nonetheless, further research is needed on the recognition of wood material.
Acknowledgements
This work was supported in part by the National Natural Science Foundation of China under Grant 61673238.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This work was supported in part by the National Natural Science Foundation of China [grant number 61673238].

ORCID
Huaping Liu http://orcid.org/0000-0002-4042-6044

References
Ahmad, K. S., Thosar, A. S., & Nirmal, J. H. (2015). A unique approach in text independent speaker recognition using MFCC feature sets and probabilistic neural network. In Eighth international conference on advances in pattern recognition (pp. 1–6). Kolkata, India: IEEE.

Astuti, W., Sediono, W., & Albinu, A. M. (2012). Adaptive short time Fourier transform (STFT) analysis of seismic electric signal (SES): A comparison of Hamming and rectangular window. In Industrial electronics and applications (pp. 372–377). Singapore: IEEE.

Cao, J., & Lin, Z. (2015). Extreme learning machines on high dimensional and large data applications: A survey. Mathematical Problems in Engineering, 2015, p1, 1–13.

Cao, J., Lin, Z., & Huang, G. B. (2012). Voting based extreme learning machine. Information Sciences, 185(1), 66–77.

Cao, J., Zhao, T., & Wang, J. (2017). Excavation equipment classification based on improved MFCC features and ELM. Neurocomputing, 261, 231–241.

Dixit, A., Vidwans, A., & Sharma, P. (2016). Improved MFCC and LPC algorithm for bundelkhandi isolated digit speech recognition. In International conference on electrical, electronics, and optimization techniques (pp. 3755–3759). Chennai, India: IEEE.

Feizifar, B., Haghigham, M. R., & Soleyman, S. (2012). Application of continuous wavelet transform for fault location in combined overhead line and cable distribution networks. In International conference on electrical and electronics engineering (pp. 142–146). Macau, China: IEEE.

Fujisaki, W., Tokita, M., & Kariya, K. (2015). Perception of the material properties of wood based on vision, audition, and touch. Vision Research, 109, 185–200.

Gao, P. (2012). Study on the optimization of backpropagation neural network classifier. Wuhan, China: Huazhong University of Science and Technology.

Ghosal, A., & Chakraborty, R. (2009). Speech/music classification using occurrence pattern of ZCR and STE. In International symposium on intelligent information technology application (pp. 435–438). Shanghai, China: IEEE.

Huang, J., Xiao, S., & Zhou, Q. (2015). A robust feature extraction algorithm for the classification of acoustic targets in wild environments. Circuits, Systems, and Signal Processing, 34(7), 2395–2406.

Huang, G. B., Zhou, H., & Ding, X. (2012). Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42(2), 513–529.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2004). Extreme learning machine: A new learning scheme of feedforward neural networks. Proceedings of the International Joint Conference on Neural Networks, 2, 985–990.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. Neurocomputing, 70(1C3), 489–501.

Kanaya, S., Kariya, K., & Fujisaki, W. (2016). Cross-modal correspondence among vision, audition, and touch in natural objects: An investigation of the perceptual properties of wood. Perception, 45(10), 1099–1114.

Li, X., Dai, B., & Fang, S. (2001). The recognition performance and robustness of high order MFCC for speaker recognition. Signal Processing, 17, 124–129.

Liu, H., Sun, F., & Zhang, X. (2019). Robotic material perception using active multi-modal fusion. IEEE Transactions on Industrial Electronics, in press. doi:10.1109/TIE.2018.2878157

Mohammadi, H. R. S., & Saeidi, R. (2008). Speaker identification performance enhancement using Gaussian mixture model with GMM classification post-processor. In IEEE international conference on signal processing and communications (pp. 504–507). Dubai, United Arab Emirates: IEEE.

Pai, P. F., Deng, H., & Sundaresan, M. J. (2015). Time-frequency characterization of Lamb waves for material evaluation and damage inspection of plates. Mechanical Systems and Signal Processing, 62, 183–206.

Strese, M., Schuwerk, C., & lepure, A. (2016). Multimodal feature-based surface material classification. IEEE Transactions on Haptics, PP(99), 1–1.

Wei, J., Liu, H. P., Yan, G., & Sun, F. C. (2016). Multi-modal deep extreme learning machine for robotic grasping recognition. In Proceedings in adaptation, learning and optimization (Vol. 7, pp. 223–233). Cham: Springer.

Yang, Y., & Wu, Q. M. J. (2016). Multilayer extreme learning machine with subnetwork nodes for representation learning. IEEE Transactions on Cybernetics, 46(11), 2570–2583.

Yusnita, M. A., Paulraj, M. P., & Yaacob, S. (2011). Malaysian English accents identification using LPC and formant analysis. In IEEE international conference on control system, computing and engineering (pp. 472–476). Penang, Malaysia: IEEE.

Zhang, X. F., Wang, F., & Xia, P. (2011). Feature factor fused on LPC and MFCC. Computer Engineering, 37(4), 216–217.

Zhang, Y. J., Yang, J. L., & Sun, D. F. (2011). Application of wavelet MFCC and HMM in low altitude acoustic target recognition. Information and Electronic Engineering, 9(6), 744–748.

Zhang, X. M., Zhang, X. Y., & Liang, W. Z. (2008). A noise-robust speech recognition based on wavelet MFCC. Journal of China Academy of Electronics and Information Technology, 3(2), 187–189.

Zhou, P., Li, X. P., & Li, J. (2013). Speech emotion recognition based on mixed MFCC. Applied Mechanics and Materials, 249–250, 1252–1258.