Retraction

Retracted: Automatic Piano Harmony Arrangement System Based on Deep Learning

Journal of Sensors

Received 19 September 2023; Accepted 19 September 2023; Published 20 September 2023

Copyright © 2023 Journal of Sensors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article’s content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

Research Article
Automatic Piano Harmony Arrangement System Based on Deep Learning

Jianhua Li

College of Music and Dance, Huaihua University, Huaihua 418000, China

Correspondence should be addressed to Jianhua Li; ljh@hhtc.edu.cn

Received 1 June 2022; Revised 21 June 2022; Accepted 5 July 2022; Published 25 July 2022

Academic Editor: Yuan Li

Copyright © 2022 Jianhua Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Harmony, which plays an important role in enriching melody expression, is a combination of multiple notes. Melody coordination involves adding harmony effect to a single note of melody, which involves professional knowledge of basic music theory and harmony rules, and requires a high technical threshold. Under the macro background of deep learning and neural network technology, artificial intelligence is widely used in music retrieval, music creation, and music teaching. In this article, we provide a powerful tool for piano music creation by manually arranging melody and harmony instead of using deep learning. In this paper, harmonic elimination is divided into three subtasks: note detection, measurement, and multifundamental frequency estimation and model training. The music signal is divided into several segments by note detection, and the main notes and harmonic components of each segment are extracted by multifundamental frequency estimation, which are used as the features and labels of the neural network, so as to give a model with the ability of arrangement and harmony.

1. Introduction

Music is a kind of art in which sound is a form of expression, which can express human emotions, regulate emotions, and promote exchange of ideas and culture, and is an indispensable part of people’s daily life [1]. Harmony is a combination of multiple tones, which can effectively enhance expressive and infectious power of music, and plays an important role in musical accompaniment, including vertical structure and horizontal structure [2]. The piano is a multimusical instrument that produces harmonic effects by pressing several keys at the same time, and because of its wide range and many playing techniques, piano is known as “king of musical instruments” and occupies an important position in music performance [3]. Because of the importance of harmony in music, harmonic arrangement is a crucial step in music composition, as a single melody without harmony cannot express musical image well. Harmonization for the main melody is based on the relationship between tones before and after melody, adding harmonic effects to certain single tones in melody; this work involves basic music theory, harmonic theory, and other professional knowledge, which not only requires staff to have a solid basic knowledge of music theory but also needs to have the ability to think and perceive harmonic sound effects, which is a high technical threshold for ordinary people or amateurs. Even those who have received musical training may not be proficient in applying harmonic theory to actual music performance [4–6].

In the period of classical music school, the main tone music became the main system. The simple and clear characteristics of the ideological content and structural form pursued at that time were also reflected in the simplicity of harmonic techniques. The major and minor system became the basis of harmony, and the medieval mode lost its influence. The tonal meaning of harmony is more clear and concentrated, emphasizing the three main chords of master, subordinate, and subordinate. Digital bass is no longer used in creation, and bass has also got rid of the shackles of flowing lines. Due to the squareness of the structure and the absence of the complicated voice parts and rhythms in polyphonic music, the rhythm of harmony is regular and rhythmic, with symmetrical and balanced harmony as the main body. Out of tune, shifting mode forward, decreasing
seventh chord, increasing sixth chord, and contraposition with major and minor tones are all widely used. We began to apply the harmonic major mode and applied the obstruction of the descending VI major triad. With the application of the semitone outside the chord, the decorative semitone has also developed and become a colorful technique. In the musical forms of the main music, especially in large-scale musical forms, such as sonata form, the structural role of harmony has been brought into full play and has become one of the elements of the structure of the main music.

With improvement of computer performance and development of signal processing technology, the connection between computer science and musicology is getting closer and closer, and more and more researchers are using digital signal processing technology to extract information from music, such as pitch, pitch length, chord, and other characteristics [3, 7, 8]. With advent of artificial intelligence, deep learning, neural networks, and other theoretical knowledge related to machine learning further promote development of music systems in direction of intelligence [9]. On the other hand, music signals change with time, different music states or characteristics change with time, and the states at different times are interrelated. In this sense, recurrent neural networks are good at processing time-related sequence data.

2. Related Work

The detection of note starting point is a very critical part in the field of music signal processing, and detection performance will directly affect results of later experiments. AbdelAlGalil et al. [10] extract MFCC coefficient (mel frequency cepstral coefficient, MFCC) from signal using auditory characteristics of the human ear and then detect the starting point based on cepstral distance with an accuracy of 96%. Schönberger [11] records a moment of note onset based on the result of phase difference, while Blaszke and Kostek [12] first preprocess music signal in full phase and then detect note onset using a feature of phase difference mutation, and experimental results show that this type of method is more suitable for note detection of a slow rhythm music. Alqahtani et al. [13] combined wavelet domain and time domain features for note slicing and achieved 96% accuracy for detecting the starting point of piano music, but the number of missed notes was high, resulting in a low recall rate [14]. Firstly, a constant Q-transform was applied to music signal, then energy value of high frequency weighting was calculated, and finally, the energy value was differentiated and peak value was detected to get a detection result, and detection accuracy of this method reached 85%. Other scholars have applied machine learning techniques to note detection, which requires a large number of training samples and high computational complexity, but results are more satisfactory.

The goal of multipitch estimation is to accurately estimate the pitch of all notes in a music signal, which mainly includes two stages: single-pitch estimation and multipitch estimation. The main algorithms of single fundamental frequency estimation include the autocorrelation function, average amplitude difference method, inverse spectrum method, and harmonic product spectrum method. Goto and Dannenberg [15] proposed an autocorrelation algorithm for summation by considering characteristics of the human ear, using a cochlear filter set as well as an autocorrelation algorithm. Wu et al. [16] used a short-time spectrum of music signal for modeling and designed a weighted hybrid tone model, but usually, only the primary fundamental frequency can be estimated. Saghezchi et al. [17] improved the system by using the constrained hybrid Gaussian model to model a set of harmonics as a whole and proposed harmonic time series structured clustering algorithm. In the twenty-first century, multibasic frequency estimation has been incorporated with new techniques and developed more rapidly. For example, Weng and Chen [18] proposed the classical selected generation spectral subtraction method, in which fundamental frequency and its harmonic frequency information are removed from the spectrum by the selected generation, and subsequently improved by using gammatone filter banks, which are more perfect and have a higher accuracy rate. In [19], mel inverse spectral coefficients based on spectral envelope are used as features, and spectral features of signal to be measured are expressed as a weighted sum of spectral features of each note by the least squares method; Radwan et al. [20] propose a nonnegative matrix decompositions algorithm based on average energy spectral envelope for multiple note recognitions; Pachet and Roy [21] build a polyatomic note dictionary, and the dictionary is used for spectral decomposition of music signals. With the rapid development of Internet technology, music has been widely spread. The effective method of extracting, retrieving, and organizing music information, that is, the research of music information retrieval, has been widely concerned by the academic and information circles. Multifundamental frequency estimation is one of the research hotspots in the field of music information retrieval. The basic task is to estimate the number of simultaneous notes in polyphonic music, so as to obtain the fundamental frequency value, start time, and end time of each note. The current multifundamental frequency estimation methods cannot meet the actual needs, so the further research on multifundamental frequency estimation methods is particularly important.

In this paper, we study piano music and combine techniques of basic music theory, digital signal processing, and recurrent neural network to study and improve note detection and multibasic frequency estimation, respectively, and use the main notes and harmonic components obtained by multibasic frequency estimation as features and labels of neural network to train model, so that it has the ability to add harmonic effects to melodic monophonic notes, thus realizing an automatic harmonic arrangement system.

3. Piano Automatic Harmony

System Architecture

As shown in Figure 1, piano harmony automatic arrangement system designed in this paper contains three main sub-tasks: note detection, multibasic frequency estimation, and
model training, and all research and innovation work is centered on these three subtasks.

Most of the previous starting point detection methods do not take into account a priori knowledge of instrument pitch. The end of note will be detected between two note start points according to energy reading value, and note start point and note end point will form a note segment, which contains one or more notes. The diagram of note detection described in this paper is shown in Figure 2. This paper takes the polyphonic music of piano, a multipart musical instrument, as the research object. In the framework of multifundamental frequency estimation method based on nonnegative matrix decomposition, the time-frequency representation of music signal, the construction of note dictionary, and spectral decomposition algorithm are analyzed, and \( L_q(P, q) \) norm block sparsity constrained nonnegative matrix decomposition algorithm effectively improves the accuracy of multifundamental frequency estimation of a single-frame signal. Finally, based on the nonnegative matrix decomposition, the multifundamental frequency estimation method is studied directly at the note event level rather than at the signal frame level.

Note start detection refers to the detection of the start time of all notes in a piece of music. Generally, from the beginning to end of note will go through four stages, namely, Attack, Decay, Sustain, and Release (ADSRS), and change of amplitude of each stage is shown in Figure 3.

The structure of neurons is shown in Figure 5.

In Figure 5, \( x_1, x_2, \cdots, x_n \) represents the input signal, and \( w_1, w_2, \cdots, w_n \) represents the weight of each input signal and neuron connection; in general, neural network will have multiple layers, except for input and output layers. Specific structure is shown in Figure 6.

Different from ordinary neural networks, the hidden neurons in recurrent neural networks are interconnected; that is, the output of each hidden layer is not only related to the input at the current time but also related to the output state of the hidden layer at the previous time. Its structure is shown in Figure 7.

In practice, the length of input sequence is not necessarily equal to output sequence, and the length of input and output sequences can be divided into the following three modes: one-to-many mode, many-to-one mode, and many-to-many mode. Figure 8 is a diagram of one-to-many mode, in which there is only one input and multiple outputs, and this mode is generally used in automatic composition, text generation, etc.
Figure 9 is a schematic diagram of the many-to-one model, in which there are multiple inputs and only one output, and this model is generally used in classification-related problems.

Figure 10 shows a schematic diagram of the many-to-many model, in which there are multiple inputs and outputs, and this model has a wide range of applications, such as speech recognition and machine translation. There is another common variant of the many-to-many model, which is called the encoder-decoder model, which works by first encoding input data into an intermediate vector $e$ and then decoding output according to this intermediate vector as shown in Figure 11.

4. Improved Neural Network Algorithm

(1) First define error of each layer:

$$\delta^l = \frac{\partial L}{\partial Z^l}, \quad (l = 1, 2, \cdots, N),$$

(1)
where \( l \) stands for loss and \( \delta^l \) represents the error of layer \( L \).

(2) Calculate error of each layer from the last layer forward according to the chain rule:

\[
\delta^l_i = \sum_j \frac{\partial L}{\partial z^l_j} \frac{\partial z^l_j}{\partial \delta^l_i} = \left( \sum_j \delta^{l+1}_j w_{ij} \right) g'(z^l_i), \quad (2)
\]

where \( \delta^l_i \) represents the error of the neuron in layer.
(3) Calculate gradient of each parameter according to the error of each layer:

\[
\nabla w_{ij}^{l} = \frac{\partial L}{\partial z_{ij}^{l}} \frac{\partial z_{ij}^{l}}{\partial w_{ij}^{l}} = \delta_i a_i^{-1},
\]

\[
\nabla b_i^{l} = \frac{\partial L}{\partial z_i^{l}} \frac{\partial z_i^{l}}{\partial b_i^{l}} = \delta_i.
\]

(4) Update each weight parameter according to the gradient:

\[
\begin{align*}
\begin{align*}
    w_{ij}^{l} &\leftarrow w_{ij}^{l} - \gamma \nabla w_{ij}^{l}, \\
    b_i^{l} &\leftarrow b_i^{l} - \gamma \nabla b_i^{l},
\end{align*}
\end{align*}
\]

where \(w_{ij}^{l}\) represents the \(j\)-th weight parameter of the \(i\)-th neuron in layer, a bias term of the \(i\)-th neuron in layer \(b_i^{l}\), and \(\gamma\) represents the learning rate, also known as step size. The larger the \(\gamma\), the faster the weight parameter updates, but too large to converge. On the contrary, the smaller the \(\gamma\), the newer and the slower fork weight parameter, which can avoid oscillation or divergence, but the number of iterations will increase.

There are 88 single-tone samples of piano keys. Each sample obtains logarithmic energy spectrum after Fourier transform and uses timbre filter bank for filtering; that is, logarithmic energy spectrum and amplitude spectrum of timbre filter bank are multiplied and accumulated according to the corresponding frequency points, and output is shown in the following equation:

\[
A_{88 \times 88} = [A_{(k,1)} A_{(k,2)} \cdots A_{(k,88)}]
\]

\[
= \begin{bmatrix}
A_{(1,1)} & A_{(1,2)} & \cdots & A_{(1,88)} \\
\vdots & \vdots & \ddots & \vdots \\
A_{(88,1)} & A_{(88,2)} & \cdots & A_{(88,88)}
\end{bmatrix}
\]

For note segment requiring multifundamental frequency estimation, energy matrix is calculated. Similar to training tone matrix, note segment to be measured is first Fourier transformed and then filtered by tone filter bank:

\[
E_k = \sum_{\omega} Y(\omega) G_k(\omega),
\]

where \(Y(\omega)\) is the logarithmic energy spectrum representing musical segments, \(G_k(\omega)\) represents the frequency response of \(k\)-th filter in timbre filter bank. The value range of \(K\) is \([1, 88]\). \(E\) represents the energy value of note segment filtered by the \(k\)-th filter. \(E_1 E_2 \cdots E_{88}\) can be spliced to form \(88 \times 1\) energy matrix \(E_{88 \times 1}\) as shown in the following equation.

\[
E_{88 \times 1} = [E_1 E_2 \cdots E_{88}]^T = \begin{bmatrix}
E_1 \\
E_2 \\
\vdots \\
E_{88}
\end{bmatrix}
\]

Since each element in energy matrix \(E_{88 \times 1}\) represents the sum of energy of fundamental frequency of corresponding piano key and its harmonics and also represents the possibility of corresponding piano key, greater its value, greater possibility of occurrence, and number of notes in harmony generally do not exceed 5, and timbre matrix \(A_{88 \times 88}\) reduces dimension from energy matrix \(E_{88 \times 1}\). Find out first the \(n\) larger values in \(X_1\) to get their corresponding piano key serial numbers \([n_1, n_2, \cdots, n_n]\), and then from timbre matrix \(A_{88 \times 88}\), find out the column of the corresponding key serial number and reduce the dimension of timbre matrix \(a \times 8\) to \(88 \times N\). Obtain timbre...
dimension reduction matrix $B_{88 \times N}$, as shown in the following equation.

$$B_{88 \times N} = \begin{bmatrix} A_{(k,n_1)} & A_{(k,n_2)} & \cdots & A_{(k,n_N)} \\ \vdots & \vdots & \ddots & \vdots \\ A_{(88,n_1)} & A_{(88,n_2)} & \cdots & A_{(88,n_N)} \end{bmatrix}$$  \hspace{1cm} (9)

For the calculated energy matrix $E_{88 \times 1}$ and tone dimension reduction matrix $B_{88 \times N}$, calculate harmony coefficient vector $W_{N \times 1}$ so that the product of dimension tone reduction matrix $B_{88 \times N}$ and harmony coefficient vector $W_{N \times 1}$ approximates energy matrix $E_{88 \times 1}$:

$$E_{88 \times 1} \approx \hat{E}_{88 \times 1} = B_{88 \times N} \times W_{N \times 1},$$  \hspace{1cm} (10)

where $E_{88 \times 1}$ represents the energy matrix, $B_{88 \times N}$ represents the dimension reduction matrix of timbre, and $W_{N \times 1}$ represents the harmonic coefficient vector. Its physical meaning is that energy matrix $E_{88 \times 1}$ of symbol segment is expressed as a combination of energy distribution of $N$ piano key single-tone samples, and the elements in harmony coefficient vector $W_{N \times 1}$ represent the volume of corresponding piano key. In this paper, gradient descent method is used to solve harmonic coefficient vector $W_{N \times 1}$. Since volume of piano keys cannot be negative, that is, elements in $W_{N \times 1}$ must be greater than or equal to zero, multiplicative update rule is used to ensure that elements in $W_{N \times 1}$ are positive in the process of solving. Firstly, a loss function $J$ based on Euclidean distance is defined. The process of solving is to change value of $W_{N \times 1}$ so that $J$ decreases continuously. The definition of $J$ is shown in the following equation:

$$J = \frac{1}{2} \sum_{ij} \left( E_{ij} - \hat{E}_{ij} \right)^2.$$  \hspace{1cm} (11)

According to Equation (11),

$$\frac{\partial J}{\partial E_{ij}} = -\sum_i \left( E_{ij} - \hat{E}_{ij} \right),$$

$$\hat{E}_{ij} = \sum_k B_{ik},$$

$$W_{kj} \frac{\partial \hat{E}_{ij}}{\partial W_{kj}} = B_{ik}.$$  \hspace{1cm} (12)

According to Equation (12),

$$\frac{\partial J}{\partial W_{kj}} = \frac{\partial J}{\partial E_{ij}} \frac{\partial \hat{E}_{ij}}{\partial W_{kj}} = -\sum_i \left( E_{ij} - \hat{E}_{ij} \right) B_{ik} = \sum_i \hat{E}_{ij} B_{ik} - \sum_i E_{ij} B_{ik},$$

$$= (B^T \hat{E})_{kj} - (B^T E)_{kj} = (B^T BW)_{kj} - (B^T E)_{kj},$$  \hspace{1cm} (13)

Then, update the parameters according to the descending direction of the gradient:

$$W_{kj} = W_{kj} - \gamma \times \left[ (B^T BW)_{kj} - (B^T E)_{kj} \right],$$  \hspace{1cm} (14)

where $\gamma$ is the updated step size and $W_{kj}$ is negative when iterating according to the following equation so that

$$\gamma = \frac{W_{kj}}{(B^T BW)_{kj}}.$$  \hspace{1cm} (15)

Substitute Equation (15) into Equation (14) to obtain the following:

$$W_{kj} = W_{kj} \frac{(B^T E)_{kj}}{(B^T BW)_{kj}},$$  \hspace{1cm} (16)

where the value of $K$ is $[1, n]$ and the value of $J$ is 1, because the dimensions of matrix $E$ and matrix $W$ are $88 \times 1$ and $N \times 1$, respectively. Updating parameters according to Equation (16) ensures that each element in the entire update process $w_{N+1}$ is a positive number. You need to set a threshold before updating $\delta$. When loss/less than threshold $\delta$, stop the parameter update.

After parameter update is completed, there will be a positive number very close to zero in $w_{N+1}$. At this time, it is necessary to set a threshold and set elements less than threshold to zero. The specific method is to take the maximum value from $w_{N+1}$ and multiply the maximum value by a coefficient $\mu$, and the obtained product result is used as a threshold to filter result. Because elements in $w_{N+1}$ represent the volume of corresponding keys, keys corresponding to the filtered nonzero elements are combined into a harmony, and the note of the key corresponding to the maximum element value is the main note. In the experiment, $N = 20.6 \times 10^{-3}$, and $\mu = 0.1$.

### 5. Results

Therefore, we can find the following conclusions: on the one hand, due to the characteristics of the piano’s own structure, the basic components of the sound emitted by the keys in the bass area are few; on the other hand, there are many harmonic components, while the keys in the high-pitched area have more basic components and less harmonic components. This difference is shown in Figure 12.

Loss of fundamental frequency means that in spectrum of a piano note, fundamental frequency component is very small compared to harmonic component, or even completely lost, which occurs mostly in bass region. Take bass note A0 for example, loss of fundamental frequency is shown in Figure 13.

Therefore, compared with other studies, the innovations of this study are as follows: (1) the common multiresolution time-frequency representation constant Q-transform (CQT) in music signal analysis is studied. It is found that although CQT has a high frequency resolution at low frequencies, it also leads to a reduction in time resolution. The variable
Q-transform is introduced for the first time as a tool for estimating the time-frequency representation of music signals with multifundamental frequencies, which has a better time resolution than CQT at the same frequency resolution and efficient coefficient calculation. (2) The spectral decomposition method based on monatomic and polyatomic note dictionary is studied. Study uses $L_{p,q}$ norm sparse constrained multifundamental frequency estimation. The
The experiment of spectral decomposition of monatomic dictionary shows that the effect of $PPL$ norm is better than that of common $1L$ norm. In view of the obvious spectrum changes of notes at different times, it is pointed out that the single-atom note dictionary does not take into account the dynamic changes of the base atoms of the note spectrum, and then, the construction methods of multiatom note dictionary are introduced from the perspectives of modeling and learning. Finally, based on the multiatom note dictionary, $L(P, q) \sim q$ norm block sparsity constrained nonnegative matrix decomposition algorithm. The experimental results show that when the number of atoms is 2, the $F$ value of the algorithm for multiple fundamental frequency estimations of single-frame signal of music clips in maps database reaches nearly 78%. (3) The multifundamental frequency estimation methods based on nonnegative matrix decomposition are all for single-frame signal processing. It does not detect the note start point in advance, but obtains the note start point through postprocessing detection results, which may lead to false start points and the error of dividing a note into multiple notes between two note start points.

Harmonic overlap means that fundamental frequency or harmony of a note overlaps with that of other notes. For example, when the ratio of fundamental frequencies of two
notes satisfies \( f_1/f_2 = m/n \) relationship, \( N \)th harmonic of note \( f_1 \) will overlap with \( N \)th harmonic of note \( f_2 \). The smaller the ratio of \( m \) to \( N \), the more serious overlap. Taking note A3 and note A4 as examples, a phenomenon of harmonic overlap is shown in Figure 14:

The amplitude-frequency characteristics of 49th note filter are shown in Figure 15 for piano key number 49, A4.

As shown in Figure 15, the 49th tone filter has only one passband, because fundamental frequency of note A4 with key number 49 is 440 Hz, so center frequency of passband is 440 Hz, and according to relative pitch of twelve mean rhythms, frequencies of semitones before and after 440 Hz can be calculated as 415.3 Hz and 466.1 Hz, which correspond to boundary frequencies of passband.

The amplitude-frequency characteristics of 88 band-pass filters of tone rhythm filter set were superimposed together to obtain amplitude-frequency characteristics shown in Figure 16.

Select a piece of music from 30 pieces of music and intercept segment for about 2 seconds. Extract spectrum features through the three methods, respectively. The detection results are shown in Figure 17.

In Figure 17, time domain diagrams of music fragments and detection diagrams of the three methods are presented, respectively. One of the reasons is that the MFCC method can be interfered by frequencies outside pitch range of keys, and these redundant frequency information may lead to appearance of pseudo-peaks, and the result of treating pseudo-peaks as peaks is false detection. The detection diagram based on the MFCC inverse spectral distance method has clearer changes. False peaks may appear in the spectrum, some of which are not very obvious, and the amplitude between the peaks is very different. It is easy to ignore the peaks, which may easily lead to missed detection and errors.

The results are shown in Table 1.

Rand subset contains 300 wav format files, each of which has only one group of multiple fundamental frequencies composed of random notes. Therefore, there are 300 groups in total. The number of combinations ranges from 2 to 7. According to the number, it can be divided into six categories, and the number in each category is 50. The results are shown in Table 2.

From Table 2, it can be seen that recall and precision rates decrease gradually with an increase of number of notes,
which leads to decrease of $F$-measure. In general, the proposed method for Rand subset data with multibasic frequency estimation experiments results in 84.07%, 78.71%, and 81.30% recall, accuracy, and $F$-measure, respectively. The results were proposed in [13], and comparison results are shown in Figure 18.
Considering the large WAV, I extracted about 30 seconds of music clips from the synthetic WAV, compressed them into MP3 format, and then uploaded them to the ECS. The score is 5 levels, with a minimum of 1 point and a maximum of 5 points. From low to high, it corresponds to poor quality, barely acceptable, acceptable, relatively satisfied, and very satisfied. After listening to the music, the listener can evaluate the effect. I invited 20 music lovers to evaluate the effect, as shown in Table 3.

From the score, the highest score of 1 is 4.6, and the lowest score of 9 is 3.1. Most of the average scores are above 3.5, and the overall average score is about 4, which means that listeners are generally satisfied with the effect of automatic harmony arrangement.

Table 1: Test results of note starting point detection.

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall rate $R$</th>
<th>Accuracy $p$</th>
<th>F-measure</th>
<th>Correct number</th>
<th>Number of missed inspections</th>
<th>Number of false inspections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper method</td>
<td>86.54%</td>
<td>97.28%</td>
<td>91.60%</td>
<td>35455</td>
<td>5520</td>
<td>994</td>
</tr>
<tr>
<td>Spectral difference method</td>
<td>85.49%</td>
<td>94.77%</td>
<td>89.90%</td>
<td>35033</td>
<td>5940</td>
<td>1935</td>
</tr>
<tr>
<td>Based on MFCC cepstrum distance</td>
<td>78.33%</td>
<td>96.07%</td>
<td>86.32%</td>
<td>32609</td>
<td>8366</td>
<td>1328</td>
</tr>
</tbody>
</table>

Table 2: Multifundamental frequency estimation results of Rand subset.

<table>
<thead>
<tr>
<th>Detection result</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct number</td>
<td>91</td>
<td>135</td>
<td>176</td>
<td>211</td>
<td>247</td>
<td>274</td>
<td>1136</td>
</tr>
<tr>
<td>Number of missed inspections</td>
<td>9</td>
<td>15</td>
<td>21</td>
<td>40</td>
<td>55</td>
<td>75</td>
<td>215</td>
</tr>
<tr>
<td>Number of false inspections</td>
<td>11</td>
<td>25</td>
<td>39</td>
<td>50</td>
<td>82</td>
<td>100</td>
<td>306</td>
</tr>
<tr>
<td>Recall</td>
<td>91.09</td>
<td>90.66</td>
<td>89.49</td>
<td>84.40</td>
<td>82.10</td>
<td>78.29</td>
<td>84.06</td>
</tr>
<tr>
<td>Accuracy rate</td>
<td>84.39</td>
<td>82.50</td>
<td>80.54</td>
<td>75.22</td>
<td>72.90</td>
<td>78.72</td>
<td>78.90</td>
</tr>
<tr>
<td>F-measure</td>
<td>87.15</td>
<td>85.84</td>
<td>82.40</td>
<td>78.45</td>
<td>75.41</td>
<td>76.12</td>
<td>81.30</td>
</tr>
</tbody>
</table>

Table 3: Scoring results.

<table>
<thead>
<tr>
<th>Example</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.5</td>
</tr>
<tr>
<td>2</td>
<td>4.2</td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
</tr>
<tr>
<td>4</td>
<td>4.4</td>
</tr>
<tr>
<td>5</td>
<td>3.8</td>
</tr>
<tr>
<td>6</td>
<td>4.3</td>
</tr>
<tr>
<td>7</td>
<td>3.9</td>
</tr>
<tr>
<td>8</td>
<td>3.2</td>
</tr>
<tr>
<td>9</td>
<td>3.2</td>
</tr>
<tr>
<td>10</td>
<td>3.8</td>
</tr>
</tbody>
</table>

6. Conclusions

This paper studies a piano music automatic harmony arrangement system based on deep learning. The system uses machine instead of human as the core design idea and adds a harmony effect to melody monophones to realize automatic harmony arrangement. The recursive neural
network structure is adopted, and the model mechanism based on codec is introduced to build a network scoring platform. The original music and voice-arranged music are uploaded to the cloud server for fans to listen to through the browser, and the effect is evaluated. Most of the songs have high scores. The results show that the listener is very satisfied with the performance of the system, which verifies the effectiveness of the proposed harmonic automatic configuration system and provides a theoretical basis for machine learning.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declared that there are no conflicts of interest regarding this work.

References