IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 9, 2020, accepted July 28, 2020, date of publication July 31, 2020, date of current version August 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3013339

An Automatic Method to Develop Music With Music Segment and Long Short Term Memory for Tinnitus Music Therapy

JIEMEI CHEN[1], (Member, IEEE), FAN PAN[1], (Member, IEEE), PING ZHONG[2], TIANTIAN HE[1], LEIYU QI[1], JINGZHE LU[2], PEIYU HE[1], AND YUN ZHENG[2]

[1]College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China

[2]Hearing Center, Otorhinolaryngology-Head and Neck Surgery Department, West China School of Medicine/West China Hospital, Sichuan University, Chengdu 610041, China

Corresponding authors: Peiyu He (hpysbsy@163.com) and Yun Zheng (1141679315@qq.com)

This work was supported by the Science and Technology Support Program of Sichuan Province, China, under Grant 2013GZ1043.

ABSTRACT Tinnitus is a perception of sound when no external sound is present. It has seriously affected patients' life. Music is an option to relieve tinnitus in clinic, as it can bring enjoyment to listeners. However, existing music used in tinnitus therapies has limited duration, it is usually repetitively played during the long-term treatment and may not be helpful for relaxation. Moreover, individualized preferences of patients are ignored in most cases. Both of them may hinder tinnitus relief. Although existing methods can synthesize specific music that has unlimited duration and is not repetitively played, the synthesized music has defects with pitch mutations and long pitch durations. Moreover, characteristics of these synthesized music have not been confirmed by tinnitus patients. Therefore, this study presents an automatic method to develop the specific music based on music segments from existing music and long short term memory (LSTM). Numerical results indicate that specific music developed in this study not only retains characteristics of original music, but also overcome the defects above. Besides, a total of 30 tinnitus patients and 10 tinnitus-free volunteers participated the auditory experiment. Auditory results are consistent with numerical results and also suggest that tinnitus patients can perceive feelings that are conducive to tinnitus relief after listening preferred music. Therefore, the developed music presents a possible complement to tinnitus treatment in clinic.

INDEX TERMS LSTM, music segment, tinnitus, unlimited duration.

I. INTRODUCTION

Tinnitus is a perception of sound without an external source [1]. Suffering from tinnitus causes mental distress, sleep disorders, impaired concentration, and other symptoms that can affect quality of life [2]–[6]. Various kinds of options can be used for tinnitus relieving [1], yet none is universally effective. Music therapy is a non-invasive treatment with no negative side effects [7], and has health benefits for people [8]. It has gradually become a research focus in tinnitus treatment, such as Tailor-made Notched Music (TMNM) Therapy [9], Neuromonics Tinnitus Treatment [10] and Neuro-music therapy [11]. They are mainly developed with the intention of alleviating tinnitus by addressing potential problems in the auditory pathways [9]–[11].

In addition, music therapy is thought to reduce the emotional consequences of tinnitus [1]. It may promote habituation to the tinnitus by reducing the contrast between tinnitus and environmental sound [1], provide sounds that are soothing to induce a sense of relief from stress or tension caused by tinnitus [12], or provide sounds that are interesting with the goal of distracting attention of patients away from tinnitus [1], [12]. Based on it, we find that there are generally 2 deficiencies in above therapies that may have an impact on the efficacy of tinnitus relief [13]. On one hand, existing music used in these therapies is common music. The characteristic of the common music is that it can bring listeners enjoyment, due to the self-similarity (balance of predictability and uncertainty) in internal music [14], yet the duration

The associate editor coordinating the review of this manuscript and approving it for publication was Hiram Ponce.

is limited. Treatment is usually long term, for instance, patients are respectively suggested to listen to music for 3 hours, 2 or more hours, and 25 minutes per day in TMNM therapy [9], Neuromonics Tinnitus Treatment [10], and Neuro-music therapy [11]. Duration of each treatment session generally exceeds that of common music. Therefore, common music with limited duration is only played repetitively during long-term treatment. However, repetitive music may have restrictions on stress reduction because it can evoke memories and potentially negative emotions, and created unwanted distraction [13]. On the other hand, tinnitus impact varies greatly for individuals and therefore, any treatment such as music, is suggested to be individualized to satisfy diverse preferences [13], [15]. Especially in stimulating relaxation, degree of liking for the music rather than specific musical style appears to be the most crucial factor [16]. However, single music type is usually used in these treatments, such as classical music [17] or relaxing music [10]. Individualized music preferences of patients are ignored in most cases.

When music is selected for tinnitus treatment, it is recommended that music evoking positive and relaxing feelings [13], and satisfying individualized preferences should be used [13]. Therefore, we need to explore specific music to relieve tinnitus. The specific music should has unlimited duration, retain the property of self-similarity, and be not repetitively played. In this way, no matter how long tinnitus patients need to listen to music, we can provide it. In order to improve the efficacy, we should also consider individualized preferences of tinnitus patients when we provide the specific music to them.

In order to explore specific music, preliminary researches have been conducted to automatically synthesize music based on chaos [18], [19] or fractal theory [20]. However, there are 2 defects existed in the synthesized music. Firstly, there are pitch mutations. Secondly, the durations of pitch are long, because a large number of same pitch values repeat consecutively. Both defects cause an uncomfortable feeling to listeners while listening to music. It is contrary with the purpose of music therapy on tinnitus, which is substituting a less disruptive music for an unpleasant tinnitus perception [21]. Although both defects are overcome by the follow-up research that combines pentatonic scale and chaos theory [22], the synthesized music is limited to only five pitch values within an octave. In addition, though the above methods are able to synthesize diversified and specific music by modifying parameters, they may not satisfy individualized preferences of listeners. On this basis, in order to satisfy individualized preferences of listeners, Jin *et al.* [23] proposed to synthesize music based on the most frequent melody phrase in existing music and hyper-chaos theory. Nonetheless, the auditory results indicated that the perceived repeatability of the synthesized music was higher than that in [22], due to the limited 7 combinations of melody phrases. Besides, there is a risk in this method that it will extract an arbitrary melody phrase from the existing music, if there is no repetitive phrase detected. Moreover, they ignore the correlations of adjacent pitch values, unexpected pitch mutations may exist at the connections between two melody phrases. In summary, defects above hinder the clinical application of these specific music. Consequently, tinnitus patients have not confirmed the characteristics and effectiveness of specific music developed by previous methods [18]–[20], [22], [23].

As previously mentioned, existing music has the property of self-similarity [14]. Imitating characteristics of existing music is a desired option for us to develop specific music. Long short term memory (LSTM) neural network is a specific recurrent neural network (RNN), which can reserve historical sequence information in its model structure for a long period [24]. It has been successfully applied in many real-world problems involving sequence data, for instance, travel time prediction [25], speech recognition [26], and patterns prediction in medical area [27]. In addition, LSTM has also been applied in automatic music composition [28], [29]. Therefore, we try to apply LSTM to explore characteristics of existing music, such as the development of pitch sequence. The aim of this study is to explore a new method to automatically develop music based on music segment (MS) and LSTM. By this method, specific music is developed with unlimited duration and is not repetitively played. At the same time, different specific music is developed to provide to tinnitus patients, accordingly it can satisfy the individualized preferences of them. More specifically, this study plans to achieve the following research objects:

- Segment original music based on predominant melody pitch sequence, extract pitch sequence of each MS and further determine the main pitch sequence that can represent pitch trend of each MS.
- Design and train 2 LSTM neural networks to predict initial pitch and pitch trend of following MS, respectively. Then determine the label of following MS based on the combination of initial pitch and pitch trend. At last, develop specific music with these MS following predicted labels.
- Examine pitch variations, pitch durations and dynamical characteristics of developed pitch sequences, and compare those of pitch sequences developed with methods in previous studies. Evaluate the developed music by auditory experiments with 30 tinnitus patients and 10 tinnitus-free volunteers.

Since listeners report that the major motivational factor for listening to music is its emotional response [13], 4 categories of original music were selected for the specific music development based on 2-dimensional Arousal-Valence (AV) emotional model [30] to satisfy diverse preferences of tinnitus patients.

The paper is organized as follows. Section II presents information of participants and the proposed automatic method for music development based on MS and LSTM. Section III describes data, evaluation methods, and statistical analysis method. Evaluation results are described in

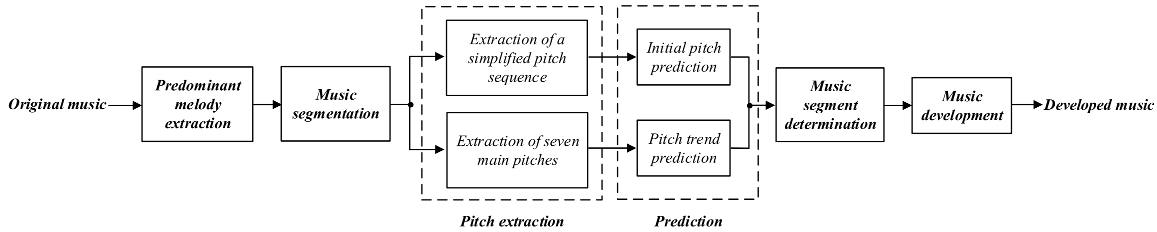


FIGURE 1. Block diagram of the automatic method to develop music.

Section IV, including numerical and auditory evaluation results. Section V provides a discussion about the proposed method by comparisons with previous methods and presents scope for future work. Section VI concludes the proposed method in this study.

II. METHODOLOGY

A. PARTICIPANTS

A total of 30 tinnitus patients were recruited in this study, which was approved by the Ethics Committee of West China Hospital of Sichuan University. All participants signed an informed consent form prior to the experiment and were eligible for this study. All participants attended the experiment at the Speech and Hearing Laboratory, Huaxi Campus of Sichuan University, Chengdu, China. Among them, the ratio of male to female is 11:19, average age is 39.67 years (ranging from 19 to 65), and average time since onset is 65.57 months. The degree of hearing loss was determined using a measured pure tone average of 500, 1000, 2000, and 4000 Hz. In this way, 7 degrees of hearing loss were [31]: no clinically relevant hearing loss (≤ 15 dB HL), slight (≥16-25 dB HL), mild (≥26-40 dB HL), moderate (≥41-55 dB HL), moderately severe (≥56-70 dB HL), severe (≥71-90 dB HL), and profound (91 dB HL or greater). Since music was exposed via a speaker, we only needed to confirm that they could hear the sound, not specifically which ear could hear. Thus, when the degrees of hearing loss on both sides' ears were different, only the lesser degree was taken into consideration. The hearing loss degree distribution of tinnitus participants included in this study is 19/6/3/1/1 (no loss/slight/mild/moderate/severe). The distribution of tinnitus location is 4/11/13/2 (right ear/left ear/both sides of ears or head/inside the head), and distribution of tinnitus severity based on Tinnitus handicap inventory [32] score is 4/5/13/7/1 (slight/mild/moderate/severe/catastrophic). The procedures of clinical trial and efficacy for tinnitus relief will be described in detail in another article.

Besides, 10 tinnitus-free volunteers participated the auditory experiment before the clinic trials. Among them, the ratio of male to female is 4:6 and average age is 34.70 years (ranging from 23 to 63). All tinnitus-free volunteers attended the experiment at the Signal and Information Processing Laboratory, Wangjiang Campus of Sichuan University, Chengdu, China.

B. AUTOMATIC METHOD TO DEVELOP MUSIC WITH MUSIC SEGMENT AND LSTM

In order to retain characteristics of original music and further develop the specific music, an automatic method that combines MS and LSTM is proposed in this study. The proposed approach is comprised of six main blocks, as depicted in Figure 1. Original music is segmented based on predominant melody extraction. Then two forms of pitch values are extracted from every MS and used for prediction. According to the results of prediction, labels of following MS are determined and used to develop music. In the following sections, we describe each block in detail.

1) PREDOMINANT MELODY EXTRACTION

Predominant fundamental frequency (f_0) sequence is extracted from original music with interval 0.01 seconds [33]. It displays zero when only accompany music is performed and corresponding frequency value when predominant melody is performed at that time.

In order to simplify the implementation of music development, conversions [34] from frequency f_0 to Musical Instrument Digital Interface (MIDI) number *pitch* are implemented as

$$pitch = 12 \times \log_2(f_0/440) + 69 \tag{1}$$

where the reference pitch is 69 (A4) and the corresponding frequency is 440 Hz. In this study, the value of pitch is regarded as the corresponding MIDI number.

2) MUSIC SEGMENTATION

After predominant melody extraction, parts of predominant melody and accompaniment are labeled and their durations are calculated. There are two steps to implement music segmentation.

- *Segmentation for predominant melody and accompaniment.*

A continuous of accompaniments are regarded as an accompaniment segment, while a continuous of predominant pitch values are regarded as a predominant melody segment. Accompaniment segments only serve as continuations of predominant melody segments in this study, instead of being used alone.

- *Segmentation based on duration.*

Before experiment, we investigated the shortest duration that people could determine whether they like the music they were listening to. Through the Internet, a total of 2004 volunteers participated in this investigation. Among them, 12.08%, 31.64%, 27.99%, 24.90%, and 3.39% of volunteers respectively reported that they could make a decision within 5, 10, 20, 60 seconds, and more than 1 minute. In brief, although most listeners could make a decision within 10 seconds, almost 88% of them could make a decision after 5 seconds. Therefore, we try to set the minimum duration of a MS to 5 seconds. Furthermore, in order to get a comfortable transition at the connections of adjacent MS, we try to start and end each MS with accompaniments, which respectively last at least 0.01 seconds. If the duration of a predominant melody segment is less than 5 seconds, the segment will be extended with the following segments (including predominant melody segment and accompaniment segment) in original music until the duration is longer than 5 seconds, and then it is regarded as a MS.

3) PITCH EXTRACTION IN MUSIC SEGMENTS

Pitch values (without 0) arranged in chronological order in a MS are obtained and saved as a pitch sequence (PS) with interval 0.01 seconds. Two forms of pitch values are extracted in this section. Firstly, a PS is converted to a simplified pitch sequence (sPS) that only one pitch value is represented when there are same pitch consecutively occurring. It is used to predict the initial pitch of following MS. Secondly, 7 main pitch values, which are used to predict pitch trend of following MS, are extracted from a PS. The process is as follows.

- *Find five main pitch values in a PS.*

The amount of pitch P after proportional computing in the PS is denoted as

$$frag\_num\_P = \left\lfloor \frac{num\_P}{frag\_all} \times 5 \right\rfloor \tag{2}$$

where $\lfloor x \rfloor$ denotes the largest integer no more than x, num_P and $frag\_all$ respectively represent the amount of a consecutive pitch P and total number of pitch values in a PS. The sum of $frag\_num\_P$ in the PS is denoted as $frag\_num\_s$. It may be not equal to 5 sometimes, due to the operation in (2). The difference between them is denoted as $diff$. If $frag\_num\_s < 5$, the minimum value in $frag\_num\_P$ will be added with $diff$. Otherwise, the maximum value in $frag\_num\_P$ will be subtracted with $diff$. At last, the 5 main pitch values are denoted as $Note\_5$.

- *Determine the seven main pitch values in a MS.*

The initial pitch $Note\_i$ and last pitch $Note\_l$ in a PS are significant for measuring pitch trend, however, they will be ignored if their ratios are significantly lower than other pitch values. Thus, the result contains seven pitch values, which are arranged as [$Note\_i$, $Note\_5$, $Note\_l$].

- *Determine pitch trend.*

Three pitch trends are defined in this study. If the maximum value is $Note\_i$ or $Note\_l$, the trend will be respectively labeled as "0.1" or "0.2". Otherwise, it will be labeled as "0.3". After pitch extraction, every MS is labeled with the sum of initial pitch value and pitch trend of the corresponding MS and denoted as LS.

4) LSTM PREDICTION FOR MUSIC SEGMENTS

- *Long short term memory.*

Recurrent neural network (RNN) is one of deep learning model structures that make efficient use of temporal information in the input sequence. After RNN training, the interrelations between current input and internal states are processed to produce the output and to represent the relevant past information in the internal states [35], [36]. Long short term memory (LSTM) neural network is a specific type of RNN which stores the information for long period of time. It has a more complex structure named LSTM cell in its hidden layer. The LSTM cell has 3 gates namely input gate, forget gate, and output gate. With the help of these gates, LSTM can decide which data to keep or remove from the recurring data, which can help to overcome the issue of vanishing or exploding gradient that usually challenges the training of RNN [24], [37]. Therefore, LSTM neural networks that permit the information to persist are used in this study to predict the development of music segments.

According to Heaton, one hidden layer is enough to approximate the vast majority of non-linear functions [38], [39]. Thus, we only explore two hidden layers at most in this study. To prevent overfitting, dropout is implemented in both layers and optimal dropout between 0.1 and 1 is searched. The number of epochs tested ranges from 10 to 200, and the model training will be automatically stopped if its validation loss hasn't improved in 10 epochs. The number of batch size tested ranges among 2 to the power of b ($0 \leq b \leq 9$). The number of neurons in each layer tested ranges from 10 to 100 with interval 10. Optimal value of these hyper-parameters are determined separately by the validation accuracy (mean squared error) and training duration. The final parameters of models are listed in the relevant section. To ensure the convergence of LSTM model, data is normalized to feed into the network. At the end of prediction process, de-normalization is implemented to get a real world number.

Two kinds of LSTM models are designed and trained in this study to predict initial pitch and pitch trend of following MS, respectively. It should be noted that a set of unique models are designed and trained specifically for each original music. Sixty pieces of existing music (duration = 7.40 ± 4.66 minutes) are applied in this study. The training is made on the previous 80% of each pieces of music and validated on the following 20%.

- *LSTM prediction for initial pitch of following music segments.*

In this section, sPS of original music is input to the corresponding LSTM model. The optimal hyper-parameter setting of the LSTM model is to define 70 neurons in the first hidden layer, which is following with a dropout layer with dropout rate 0.5 and then feed into a fully connected normal layer of 1 neuron with a linear activation function for predicting pitch. The input shape is of 9 time-step with 1 feature (pitch value) and batch-size is 128.

Figure 2 (a) shows an example of pitch value prediction and 500 pitch values are displayed in this figure. The predictions can roughly capture the curve of true testing data. However, it is predicting in a point-by-point way, which only predicts a single point ahead each time, plots this point as a prediction, then takes the next window along with the full true testing data, and predicts the next point along once again [40]. By this way, the network does not need to know much about the pitch sequence itself other than that each next pitch value most likely will not be too far from the last pitch value. Therefore, this LSTM model is only used to predict the initial pitch of following MS. It can avoid pitch mutations at connections of adjacent MS.

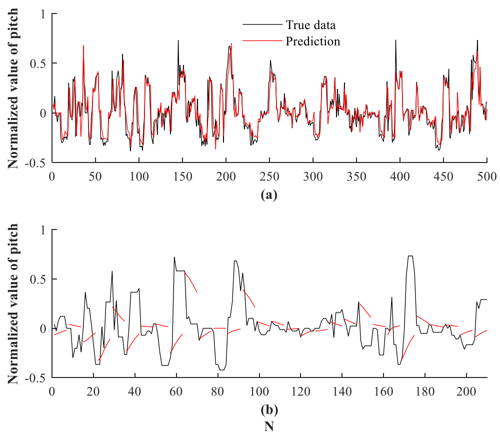


FIGURE 2. **Examples of LSTM predictions of (a) pitch value and (b) pitch trend of MS.**

- *LSTM prediction for pitch trend of following music segments.*

In this section, 7 main pitch values of each MS that follow temporal order of original music are fed in the LSTM in the form of array. The Keras LSTM layer takes the numpy array of 3 dimensions (N, W, F) where N is the number of training sequences, W is the sequence length and F is the number of features in each MS. For this study, W (read window size) is set as 7 which allows the network to get glimpses into the shape of pitch trend at each MS and only main pitch value is the feature. This process enables us to build up a pattern of sequences based on the prior window received. The sequences themselves are sliding windows and hence shift by 1 each time, causing a constant overlap with the prior windows. Batch-size is determined as 128. The optimal LSTM model is defined with 50 neurons in the first layer, which is following with a dropout layer with dropout rate 0.6, and then feeds into a fully connected normal layer of 1 neuron with a linear activation function. These steps will be used to give the prediction of next time step. At each time step, we then pop the oldest entry out of the rear of the window and append the prediction for the next time step to the front of the window, in essence shifting the window along so that it slowly builds itself with predictions, until the window is full of only predicted values. In our cases, as our window is of size 7, it would occur after 7 time steps. We then keep this up indefinitely, predicting the next time step on the predictions of the previous time steps, to see an emerging trend. We limit our prediction sequence to 7 future time steps and then shift the initiation window by 7 each time, in effect creating many independent sequence predictions of 7 time steps.

Figure 2 (b) shows an example of pitch trend prediction. Main pitch values of 30 MS are displayed in this figure. The red curves, which represent predictions of pitch trend, indicate the LSTM model does appear to be correctly predicting the trends for a good majority of MS. Therefore, this LSTM model is used to predict pitch trend of following MS. At last, the predicted label of following MS is denoted as L_p, which is the sum of predicted initial pitch value and pitch trend.

5) MUSIC SEGMENT DETERMINATION

As previous mentioned, MS is originally labeled as LS. Then predicted label L_p is compared with original label LS to determine the order of following MS. The determination process is depicted in Figure 3.

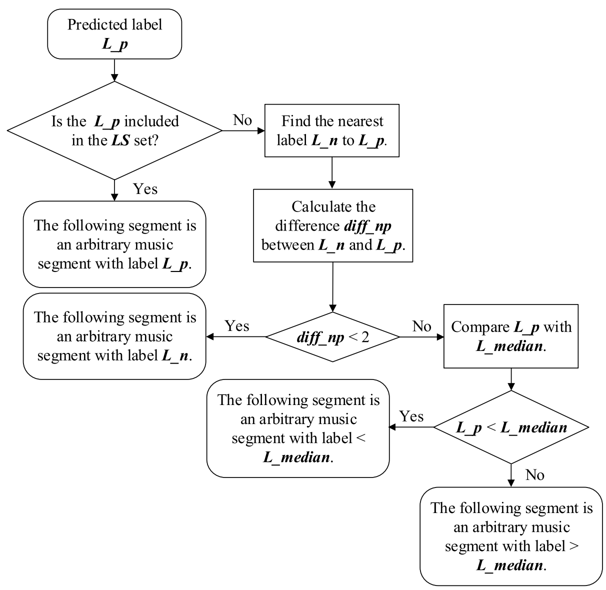


FIGURE 3. **Determination process of following MS.**

Among them, L_median is the median value of LS. In order to avoid unexpected continuous repetition, after determination as shown in Figure 3, we try to compare the determined label with that of adjacent 9 MS before itself. If it is identical with that of 1st, 3rd, 5th, 7th, or 9th MS in front of itself, the following segment will be relocated to an arbitrary segment with label $> L\_median$. Otherwise, if it is identical with that of 2rd, 4th, 6th, 8th segment in front of itself, the following segment will be relocated to an arbitrary segment with label $< L\_median$.

6) MUSIC DEVELOPMENT

There may be amplitude (volume) rapid changes occurring at connections of adjacent MS, if we directly splice them.

It may cause uncomfortable feeling sometimes and prevent listeners from relaxing [15]. A Hamming window is an option to overcome it. The amplitude of a Hamming window reaches the maximum value in the middle, gradually decreases on both sides, and the smallest value does not drop to zero. A length M of generalized Hamming window (w) is designed by

$$w = 0.54 - 0.46 \times \cos(2 \times \pi \times (0 : M - 1)/(M - 1)) \quad (3)$$

The processed MS ($MS\_processed$) is depicted by

$$MS\_processed(n) = MS(n) \times w(n) \quad (4)$$

where MS is the original MS, n is the nth point in the MS or Hamming window.

Figure 4 depicts the time domain wave of splicing before and after processing. If there are rapid amplitude changes at connection of two MS, as shown in subgraph (a) around 9th second, it can be overcome after processing, as shown in subgraph (b). In terms of auditory perception, volume of a MS is getting louder and then getting lower slowly until it is connected to the next MS.

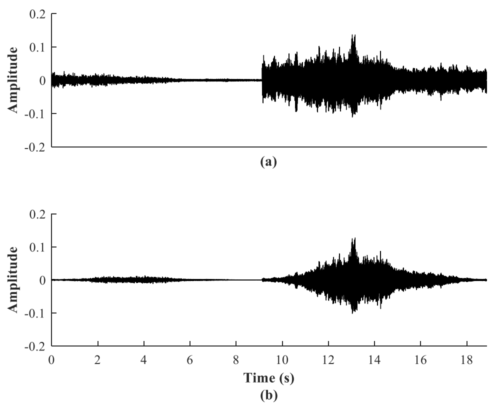


FIGURE 4. **Time domain wave of splicing with (a) 2 original MS, (b) 2 processed MS.**

III. DATA AND STATISTICAL ANALYSIS

Five types of representative sPS are compared in this study. The amount of each type sPS is 60. In addition to the sPS derived from original music (sPS_0) and developed music in this study (sPS_4), 3 types of sPS are generated by previous methods [18], [19], [23]. They are respectively denoted as sPS_1, sPS_2, and sPS_3. There are two settings that need to be explained in advance.

- *Pitch range.*

Difference between the maximum and minimum value of pitch is regarded as pitch range (PR) in a piece of music in this study. Among the 60 original music, the PR of sPS_0 range from 22 to 53.

Developed process of the 3 types of sPS (sPS_1, sPS_2, and sPS_3) is briefly described in this section. Assume that an output sequence of a chaos system is X_c, the maximum and minimum values are respectively denoted as $X_{c\,\max}$ and $X_{c\,\min}$. The target values of maximum, minimum, and PR are denoted as $P_{\max}$, $P_{\min}$, and PR, respectively. Following reference [18], pitch sequence is generated as

$$PS = \langle (X_c - X_{c\,\min})/(X_{c\,\max} - X_{c\,\min}) \times PR \rangle + P_{\min} \quad (5)$$

where $\langle x \rangle$ denotes an operation of round off x to an integer.

Following reference [19], X_c is firstly converted to a sequence D_c of integers from 0 to 9 based on values at the decimal place (the decimal point right side first) of X_c. Then pitch values P_s to be selected are from $P_{\min}$ to $P_{\max}$ with interval $\langle PR/9 \rangle$. Pitch sequence are generated as

$$PS(i) = P_s(D_c(i) + 1) \quad (6)$$

where i denotes serial number in the pitch sequence.

Following reference [23], pitch sequence is developed based on sPS_0. The sPS_0 is firstly divided into phrases with 4 seconds duration, because they assumed that a phrase consisted 4 sections and a section lasted 1 second. Then we find the most frequent melody phrase, develop 7 forms of pitch combinations based on the most frequent melody phrase, convert output of a hyper-chaos system to a sequence of integers from 1 to 7, map different outputs to different forms of pitch combinations, and obtain the corresponding pitch sequence.

For comparison, the maximum value, minimum value, and pitch range of sPS_0 are target values. Thus, PR of sPS_1, sPS_2, and sPS_4 range from 22 to 53, which are same as that of sPS_0. Whereas the 60 sequences of sPS_3 are developed with main phrase from original music, they have smaller PR, which range from 8 to 50.

- *Length of sequence.*

Pitch values repeated continuously are considered as one pitch in sPS. The continuity is reflected in the pitch duration. The length of sPS_0 is determined by the duration of original music. As duration of each therapy session is set at 30 minutes in clinical protocol. The length of sPS_4 is determined by the corresponding 30-minute developed music. In order to compare outcome, length of sPS_1, sPS_2, and sPS_3 are same as that of corresponding sPS_4, which is developed from the same PR or original music with them.

We mainly evaluate the developed music from four aspects. Firstly and secondly, adjacent pitch variations and pitch durations of these sPS are compared to examine the existence of pitch mutations and long pitch durations. The absolute value of difference between adjacent pitch values in sPS is regarded as pitch variation (PV) and total duration of continuously repetitive pitch values in PS is regarded as pitch duration (PD).

Thirdly, dynamic characteristics of sPS are compared to measure the self-similarity and complexity. Music is a kind of $1/f$ fluctuation, it can merge the randomness and orderliness into a naturally pleasant and attractive whole [41], [42]. In other words, it is self-similar and not repetitive. The power spectrum $S(f)$ of pitch fluctuations in music exhibits $1/f$ power spectra, varying as $f^{-\alpha}$ ($0.5 \le \alpha \le 1.5$) [43], [44]. To examine the self-similarity of pitch fluctuations in these

TABLE 1. Adjacent pitch variations of different type sequences.

Sequence types	*PV=1*	*1 < PV ≤ 5*	*5 < PV ≤ 10*	*10 < PV ≤ 20*	*PV >20*
sPS_0 (%)	58.37 ± 14.21	31.44 ± 10.91	6.57 ± 4.04	3.36 ± 2.48	0.26 ± 0.31
sPS_1 (%)	88.09 ± 5.78**	11.91 ± 5.78**	0**	0**	0**
sPS_2 (%)	0**	15.68 ± 6.50**	16.11 ± 5.05**	31.64 ± 4.78**	36.57 ± 11.80**
sPS_3 (%)	56.31 ± 17.91	30.66 ± 12.33	6.64 ± 4.43	5.40 ± 5.08*	0.99 ± 1.88*
sPS_4 (%)	57.97 ± 15.13	31.69 ± 11.55	6.53 ± 4.23	3.50 ± 2.78	0.31 ± 0.35

Data is expressed as Mean ± SD.
$^{*}p < 0.05$ and $^{**}p < 0.001$ represent significant difference between sPS_4 and other pitch sequences.

TABLE 2. Pitch duration distributions of different type sequences.

Sequence types	*PD ≤ 0.25*	*0.25 <PD ≤ 0.5*	*0.5 <PD ≤ 1*	*1 <PD ≤ 1.5*	*PD >1.5*
sPS_0 (%)	83.10 ± 10.96	8.80 ± 5.71	5.32 ± 4.21	1.47 ± 1.38	1.31 ± 2.23
sPS_1 (%)	18.84 ± 2.83**	23.12 ± 3.18**	26.93 ± 1.05**	14.82 ± 1.57**	16.29 ± 5.45**
sPS_2 (%)	28.14 ± 0.19**	50.98 ± 0.33**	15.14 ± 0.22**	4.19 ± 0.17**	1.55 ± 0.11
sPS_3 (%)	81.75 ± 15.34	9.08 ± 7.45	5.13 ± 5.55	1.44 ± 2.71	2.59 ± 4.90
sPS_4 (%)	82.47 ± 11.98	9.24 ± 6.59	5.59 ± 4.43	1.43 ± 1.41	1.26 ± 2.06

Data is expressed as Mean ± SD.
$^{*}p < 0.05$ and $^{**}p < 0.001$ represent significant difference between sPS_4 and other pitch sequences.

sequences, values of α are calculated. Besides, complexity of pitch fluctuation is measured. It mainly reflects the mutual information and distribution in pitch fluctuation [45]. Pitch value sequences in existing music are examined to have high complexity [46]. Although traditional entropy-based algorithms had been applied to quantify complexity of pitch fluctuation [45], an increase in the entropy may not always be associated with an increase in dynamical complexity [47]. Therefore, multiscale entropy (MSE) [47] is applied to measure complexity of pitch fluctuations in multiple time scales in this study.

At last, 60 pieces of developed music in this study were subjectively evaluated by auditory experiment.

All data was recorded in Microsoft Excel 2013. Statistical analysis was performed using IBM Statistical Package for the Social Sciences (SPSS) Version 19. Possible differences between two types of sPS were assessed by independent sample t-test sample. A p-value of <0.05 was considered statistically significant in this study.

IV. RESULTS

A. PITCH VARIATIONS

In this section, PV distributions of 5 types of sPS are compared and displayed in Table 1. To simplify the results, PV is divided into 5 categories: 1 (PV = 1), 2 ($1 < PV \leq 5$), 3 ($5 < PV \leq 10$), 4 ($10 < PV \leq 20$), and 5 ($PV > 20$). As shown in Table 1, although PV of sPS_0, sPS_3, and sPS_4 are mainly distributed in small value (such as $PV \leq 5$), they are distributed in all categories of PV. While the PV of sPS_1 is only distributed in small value, which may make the synthesized music sound dull. The PV of sPS_2 is mainly distributed in large value (such as $PV > 10$), which may cause more pitch mutations than others. More specifically, although PR of sPS_3 is smaller than that of sPS_0 and sPS_4, PV of sPS_3 significantly increases the proportion in large PV. Whereas PV of sPS_4 is basically consistent with that of sPS_0.

In order to examine PV in different PR, 3 sPS_0 with smallest (22), moderate (42), and widest (53) PR are selected for reference. The PV distributions of sPS_0 and the corresponding 4 types of sequences are depicted in Figure 5 and consistent with results in Table 1. Furthermore, we observe that as PR increases, proportion of category 1 in sPS_1 decreases and proportion of category 5 in sPS_2 increases.

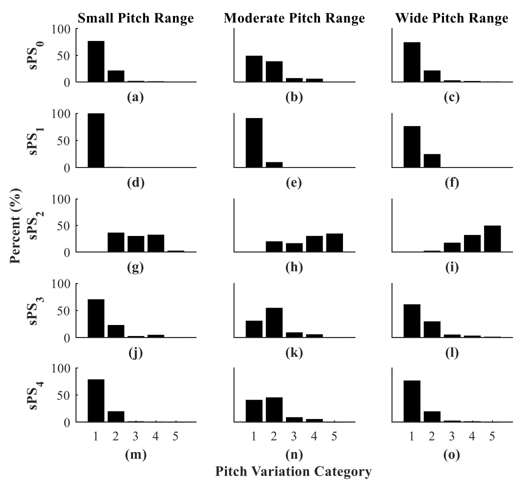


FIGURE 5. Adjacent pitch variation distribution. Distributions in small, moderate, and wide PR are respectively depicted in the 1st, 2nd, and 3rd columns. Meantime, distributions of sPS_0, sPS_1, sPS_2, sPS_3, and sPS_4 with corresponding PR are respectively depicted in the 1st, 2nd, 3rd, 4th, and 5th lines.

B. PITCH DURATIONS

Pitch duration distributions of 5 types of sPS are compared and depicted in Table 2. To simplify the results, PD is divided into 5 categories: 1 (PD 0.25 seconds), 2 (0.25 < PD 0.5 seconds), 3 (0.5 < PD 1 second), 4 (1 < PD 1.5 seconds), and 5 (PD > 1.5 seconds). As shown in Table 2, PD of sPS_0, sPS_3, and sPS_4 mainly distributed in category 1. In category 4 and 5,

TABLE 3. The dynamic characteristics of different type sequences.

Sequence types	α	*MSE_1_10* [a]	*MSE_11_20* [b]	*MSE_21_30* [c]
sPS_0	1.08 ± 0.17	1.10 ± 0.34	$1.46 \pm 0.39^{*}$	1.52 ± 0.41
sPS_1	1.03 ± 0.00	$0.70 \pm 0.09^{**}$	$1.14 \pm 0.11^{**}$	$1.15 \pm 0.06^{**}$
sPS_2	1.00 ± 0.01	$1.44 \pm 0.02^{**}$	$0.96 \pm 0.01^{**}$	$0.75 \pm 0.02^{**}$
sPS_3	1.03 ± 0.00	$0.90 \pm 0.22^{*}$	$1.00 \pm 0.17^{**}$	$0.96 \pm 0.17^{**}$
sPS_4	1.03 ± 0.00	1.01 ± 0.27	1.31 ± 0.29	1.42 ± 0.29

Data expressed as Mean ± SD.
[a] MSE_1_10 = average value of MSE from scale 1 to 10. (MSE in small time scales.)
[b] MSE_11_20 = average value of MSE from scale 11 to 20. (MSE in moderate time scales.)
[c] MSE_21_30 = average value of MSE from scale 21 to 30. (MSE in large time scales.)
$^{*}p < 0.05$ and $^{**}p < 0.001$ represented significant difference between sPS_4 and other pitch sequences.

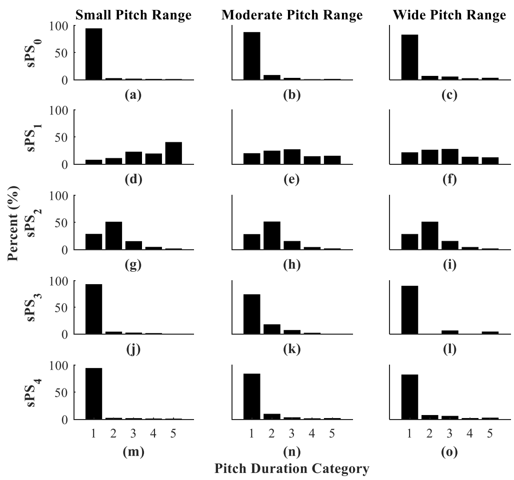


FIGURE 6. Pitch duration distribution. Distributions in small, moderate, and wide PR are respectively depicted in the 1st, 2nd, and 3rd columns. Meantime, distributions of sPS_0, sPS_1, sPS_2, sPS_3, and sPS_4 with corresponding PR are respectively depicted in the 1st, 2nd, 3rd, 4th, and 5th lines.

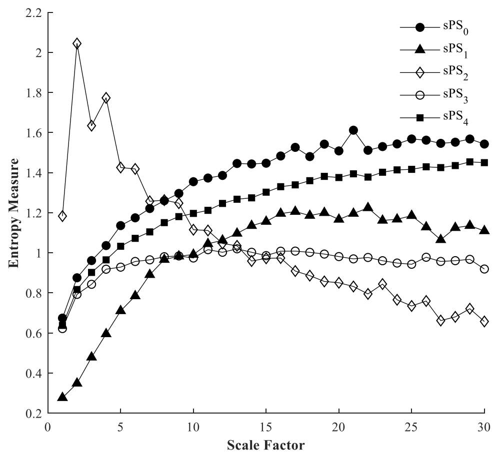


FIGURE 7. MSE analysis of different type sequences. Data is given as average value.

PD of sPS_1 has the largest proportion among these types of sequences. It indicates that long duration of pitch can be avoided in these sequences except sPS_1.

Same as the previous section, PD distribution in different PR (small - 22, moderate - 42, wide - 53) are examined. As PR increases, distributions of PD in sPS_0, sPS_2, sPS_3, and sPS_4 do not change much. However, we observe that the proportion of category 5 decreases in sPS_1.

C. DYNAMIC CHARACTERISTICS OF PITCH FLUCTUATIONS

Dynamic characteristics of pitch fluctuations in these sequences are examined in this section. As shown in Table 3, the values of α indicate that pitch fluctuations in all sequences are according with the fundamental principle of music that is the balance between repetition and contrast. Besides, values of entropy in sPS_1, sPS_2, and sPS_3 are significantly different with that in sPS_4 in small, moderate, and large time scales. Whereas the values of entropy in sPS_0 and sPS_4 are close, only different in moderate time scales ($p = 0.016$).

More specifically, the values of entropy in each time scales are depicted in Figure 7. The scale factor specifies the number of data points averaged to obtain each element of the coarse-grained pitch sequence [47]. Compared with others, the entropy values in sPS_4 are closest to those in sPS_0 and nearly monotonically increases as time scale increases. Although the values corresponding to sPS_0, sPS_3, and sPS_4 almost overlap at scale 1, the entropy value in sPS_3 is significantly lower than those in sPS_0 and sPS_4 as scale increases. Besides, we observe that for scale 1-5, highest values of entropy are assigned to sPS_2 in comparison with other pitch sequences. Yet the values in sPS_2 nearly monotonically decrease after scale 4. The values in sPS_1 are always significantly smaller than that in sPS_0 and sPS_4 in all time scales.

D. AUDITIONS

Only music developed with sPS_4 in this study is auditioned by participants to achieve comfortably auditory experience.

1) AUDITORY PROCESS

Each participant completed 15 audition sessions and there were totally 450 sessions completed by tinnitus patients and 150 sessions completed by tinnitus-free volunteers in this study.

As previously mentioned, we provided listeners 4 representative categories (happiness, anger, sadness, and contentment) of music, according to the 2-dimensional AV emotion

TABLE 4. Results of subjective evaluation for music.

Questions	*Answers*	
	Sessions completed by Tinnitus patients (%)	*Sessions completed by Tinnitus-free volunteers (%)*
Can you repeat the music just listen to?	8/92 *(Yes/No)*	7/93 *(Yes/No)*
Is there a sudden pitch change in the music you just listened to that makes you feel unpleasant?	5/95 *(Yes/No)*	14/86 *(Yes/No)*
Do you hear similar segments from music you just listened to?	91/9 *(Yes/No)*	73/27 *(Yes/No)*
Do you think the music you just listened to is positive or negative?	81/19 *(Positive/Negative)*	84/16 *(Positive/Negative)*
Do you think the music you just listened to is nervous or relaxing?	6/94 *(Nervous/Relaxing)*	6/94 *(Nervous/Relaxing)*

Data was expressed as the number of audition sessions.

model [30]. Before each audition session, the first 10 seconds of 4 different categories of music were provided to participants in a random order. They chose the most favorite one as preferred music after listening based on their own preferences and used the preferred music for the 30-minute listening. Music was free of vocals. In order to prevent participants from falling asleep during the process and not hearing the music, technicians would wake them up briefly every 10 minutes. Immediately after listening in each session, participants were asked several questions about music. The answers to these questions are used for subjective evaluation for music. The detail evaluation results is shown in Table 4.

Music was played via a speaker. Tinnitus and tinnitus-free participants were instructed to lay supine on a bed and listen in a soundproof room (IAC Acoustic Technology (Shenzhen) Co., LTD) at Huaxi campus of Sichuan University and in a quiet room at Wangjiang campus of Sichuan University, respectively. To ensure that the volume received by each ear is consistent, the speaker was put on the ground and below the head. In this study, 5 audition sessions were regarded as a phase. In the beginning of each phase, volume of music was determined via listening to white noise. The volume, which was just out of hearing when patients listening to a clip of white noise, was regarded as their hearing threshold. Then in a range above hearing threshold 10 dB A and 15 dB A, a comfortable and safe sensation level was determined by participants themselves. It ensured that music was exposed at low levels where music blends with tinnitus. The sensation level was tested by a sound level meter (Larson Davis 824). The volume of the following 4 audition sessions in a phase remained unchanged, unless participants required.

2) AUDITORY RESULTS

We observed that after most preferred music sessions, both tinnitus patients and tinnitus-free volunteers reported that they could not repeat music immediately after listening, did not hear pitch mutations, and heard similar segments. The corresponding proportions of sessions were respectively 92%, 95%, and 91% in tinnitus patients and respectively 93%, 86%, and 73% in tinnitus-free volunteers. Although they chose different categories of music, they could perceive positive and relaxing feeling after most preferred music sessions. It indicates that developed music in this study is self-similar, not repetitively played, and free of pitch mutations. It might evoke positive and relaxing feelings.

V. DISCUSSION

This study proposes an automatic method to develop specific music by learning regularity of pitch fluctuations in existing music. The specific music has unlimited duration and is not repetitively played. To the best of our knowledge, this is the first study that combines music segments and LSTM to develop specific music for tinnitus relief. Moreover, by learning characteristics of different existing music, corresponding developed music that satisfies individualized preference of patients can be provided to them. Compared with previous synthesized music [18]–[20], [22], [23], we observe that developed music in this study avoids unexpected pitch mutations and long durations of pitch. Both numerical and auditory results indicate that the developed music satisfies the requirements of tinnitus music therapy [13].

There are 3 advantages in this study. Firstly, both unexpected pitch mutation and long duration of pitch are avoided in the developed music. Furthermore, various music can be developed to satisfy individualized preferences of tinnitus patients. On one hand, the music is developed based on the inherent regularity of pitch fluctuations in original music. At the connections of adjacent MS, intrinsic relationship between adjacent pitch values is considered by LSTM prediction in this study. Thus, we do not observe difference in PV distributions between sPS_4 and sPS_0. It indicates that there is no unexpected pitch mutation in developed music. On the other hand, the proposed method is developing music based on MS from existing music. Through splicing, the durations of pitch in MS are directly used instead of re-designing by any algorithms or methods. Therefore, there is no difference in PD distributions between sPS_4 and sPS_0. And we do not observe a large proportion of long pitch durations in sPS_4 like that in sPS_1. Whereas in previous researches, sPS_1 are developed by directly mapping with the value of a chaotic system after linear processing [18]. Accordingly, in sPS_1, PV between adjacent pitch values are small, and the PV are generally long, because there are a large amount of same pitch continuously repeating in the original form of sPS_1. As shown in Table 1 and 2, there are 88.09% of PV distributing in $PV = 1$ and 16.29% of PD distributing in $PD > 1.5$ in sPS_1, whereas there are only respectively 57.97% and 1.26% in sPS_4. Consequently, music synthesized with sPS_1 sounds dull. As PR increases, although the defects in PV and PD of sPS_1 will improve, it is not obvious. Moreover, pitch mutations will exist if sPS_1 is in a large PR, and too large PR is not to be true. At the same time, although PD distribution is improved in sPS_2 through mapping rule adjustment, there are obvious pitch mutations as PR increases. As shown in Table 1, there are 36.57% of PV distributing in PV > 20 in sPS_2, whereas there are only 0.31% in sPS_4. Thus, sPS_2 are only applicable for synthesizing music with small PR. Another pitch sequence sPS_3, which are developed with transformations of main phrase from existing music [23], can satisfy individualized preference requirement. However, it still has pitch mutations. This is because that though splicing orders of segments in sPS_3 are according with chaotic regularity, the correlation between adjacent segments (or pitch values) is ignored. Thus, it may easily cause pitch mutations at the connection of adjacent segments. Besides, due to the limited number of pitch values and corresponding duration elements, there may be missing pitch and duration elements in the music developed with sPS_3, as shown in Figure 6.

Secondly, compared with music synthesized by previous methods, developed music in this study can bring more comfortable experience to listeners. In addition to the improvement in PV and PD, we also find sPS_4 has the higher complexity than pitch sequences developed by previous methods. According to Table 3 and Figure 7, we observe that only entropy values of sPS_4 increase steadily as time scales increase. It indicates that sPS_4 has the more complex structure than sPS_1, sPS_2, and sPS_3. Moreover, complexity of sPS_4 is closest to that in existing music (sPS_0). Specifically, we find that for time scale 1, the highest value of entropy is assigned to sPS_2, while the smallest is assigned to sPS_1. It indicates that variability between adjacent pitch values in sPS_2 is largest and that in sPS_1 is smallest. It is consistent with the results of PV between adjacent pitch values, which indicates that sPS_2 has obvious pitch mutations and the PV in sPS_1 are small. The PV in sPS_3 and sPS_4 are in a moderate state and similar with that in existing music. Entropy values of sPS_3 and sPS_4 are smaller than those of sPS_2, and larger than those of sPS_1 in small scales (from scale 1 to 7 as shown in Figure 7). It indicates that sPS_3 and sPS_4 are more regular than sPS_2, yet are not as regular as sPS_1 in short term. In large scales, entropy values of sPS_4 are always the largest. It indicates that sPS_4 is most uncertain in long term and also suggests that the developed music is not repetitive. Although entropy values of sPS_4 are generally slightly smaller than those of sPS_0, no significant statistical differences are found at most scales. Therefore, developed music based on sPS_4 has high complexity, has the balance between predictability in short term and uncertainty in long term. These characteristics are consistent with those in existing music. In addition, as the proposed method in this study is developing music based on MS, which are obtained from existing music. The coordination of multitrack is already included in these MS, yet it had not been researched in previous methods. Therefore, we speculate that developed music in this study can achieve more comfortable auditory experience than previous synthesized music.

Thirdly, this study is an effective advance over previous related researches, as it is the first time to recruit tinnitus patients to evaluate developed music. Through auditory experiments, we not only confirm that developed music is non-repetitive, self-similar, and free of pitch mutations, but also observe that tinnitus patients could perceive positive and relaxing feelings after preferred music listening. Although we only provided participants with music developed in this study to tinnitus patients for audition, we could speculate the functional differences between this music and other music in tinnitus treatment. As previously mentioned, original music is common music, which is time limited. If we directly repetitively play it in therapy, it may have restrictions on stress reduction [13]. Besides, music developed by sPS_1 and sPS_2 could not satisfy diverse preferences of tinnitus patients. It may affect the effectiveness of tinnitus relief and even be counterproductive [21]. Additionally, pitch mutations and long pitch durations in sPS_1, sPS_2, and sPS_3 reduce comfort degree of developed music and also may interfere with tinnitus patients. It is contrary with the role of music therapy in tinnitus, which is substituting a less disruptive background sound for an unpleasant tinnitus perception [21]. Therefore, we did not provide music generated by previous methods and original music to participants. And it can be speculated that music developed in this study is more effective than them.

The developed music in this study may be an alternative to tinnitus treatment. On one hand, it may be applied to tinnitus treatment separately to reduce the emotional consequences of tinnitus. On the other hand, due to the requirements for music not to be played repeatedly to reduce stress [13] and individualized preference consideration [9], [11], it may be integrated into other music therapies, such as TMNM, Neuromonics, and Neuro-music therapy.

However, there are also limitations existed in this study. In some developed music, the volume of music slowly drops. It may be detected by some patients as it may be contrary to their expectations. This study mainly focus on pitch development of music but ignores the development of rhythm and volume. Further research on the combination development for pitch, rhythm, and volume is necessary for the better application of music therapy in tinnitus.

The proposed method in this study is automatic and user-friendly, it may contribute to the effective implementation in clinic. The developed music in this study is without time limitation, it may be helpful to tinnitus music therapy for long term. At the same time, it may be conducive to personalized music therapy, as it is developing music based on existing music segments, instead of generating music randomly. It may offer a potential supplement to existing tinnitus music therapies, such as TMNM [9], Neuromonics [10], and Neuro-music therapy [11].

VI. CONCLUSION

In summary, we apply a new method to develop specific music automatically for tinnitus music therapy. This method develops music by learning the mutual information among pitch sequence in existing music. The developed music can retain characteristics of existing music. At the same time, it is without time limitation and not repetitive. Tinnitus patients can be provided with music that satisfy their individualized preferences, and they can perceive feelings that are conducive to tinnitus relief after listening. It may provide possible complement for tinnitus music therapy in clinic, and is worth further exploration.

ACKNOWLEDGMENT
The authors would like to thank the staff and technicians of Hearing Center at Department of Otolaryngology/Head and Neck Surgery, West China Hospital of Sichuan University in China, and tinnitus patients for their willingness to participate in this study.

REFERENCES
[1] D. E. Tunkel *et al.*, "Clinical practice guideline: Tinnitus," *Otolaryngology–Head Neck Surgery*, vol. 151, no. 2, pp. S1–S40, Oct. 2014, doi: 10.1177/0194599814545325.
[2] A. R. Møller, "Sensorineural tinnitus: Its pathology and probable therapies," *Int. J. Otolaryngol.*, vol. 2016, Feb. 2016, Art. no. 2830157, doi: 10.1155/2016/2830157.
[3] F. T. Husain, "Neural networks of tinnitus in humans: Elucidating severity and habituation," *Hearing Res.*, vol. 334, pp. 37–48, Apr. 2016, doi: 10.1016/j.heares.2015.09.010.
[4] J. A. Henry, K. C. Dennis, and M. A. Schechter, "General review of tinnitus: Prevalence, mechanisms, effects, and management," *J. Speech Lang. Hear. R.*, vol. 48, no. 5, pp. 1204–1235, Oct. 2005, doi: 10.1044/1092-4388(2005/084).
[5] A. N. Mezei, R. Enache, and C. Sarafoleanu, "Tinnitus in elderly population: Clinic correlations and impact upon QoL," *J. Med. Life*, vol. 4, no. 4, pp. 412–416, Nov. 2011.
[6] A. O. Lasisi and O. Gureje, "Prevalence of insomnia and impact on quality of life among community elderly subjects with tinnitus," *Ann. Otology, Rhinol. Laryngol.*, vol. 120, no. 4, pp. 226–230, Apr. 2011, doi: 10.1177/000348941112000402.
[7] C. A. Browning, "Using music during childbirth," *Birth*, vol. 27, no. 4, pp. 272–276, Dec. 2000, doi: 10.1046/j.1523-536x.2000.00272.x.
[8] S. Koelsch, "A neuroscientific perspective on music therapy," *Ann. New York Acad. Sci.*, vol. 1169, no. 1, pp. 374–384, Jul. 2009, doi: 10.1111/j.1749-6632.2009.04592.x.
[9] A. Stein, A. Engell, M. Junghoefer, R. Wunderlich, P. Lau, A. Wollbrink, C. Rudack, and C. Pantev, "Inhibition-induced plasticity in tinnitus patients after repetitive exposure to tailor-made notched music," *Clin. Neurophysiol.*, vol. 126, no. 5, pp. 1007–1015, May 2015, doi: 10.1016/j.clinph.2014.08.017.
[10] P. J. Hanley, P. B. Davis, B. Paki, S. A. Quinn, and S. R. Bellekom, "Treatment of tinnitus with a customized, dynamic acoustic neural stimulus: Clinical outcomes in general private practice," *Ann. Otology, Rhinol. Laryngol.*, vol. 117, no. 11, pp. 791–799, Nov. 2008, doi: 10.1177/000348940811701101.
[11] H. Argstatter, M. Grapp, E. Hutter, P. K. Plinkert, and H.-V. Bolay, "The effectiveness of neuro-music therapy according to the heidelberg model compared to a single session of educational counseling as treatment for tinnitus: A controlled trial," *J. Psychosomatic Res.*, vol. 78, no. 3, pp. 285–292, Mar. 2015, doi: 10.1016/j.jpsychores.2014.08.012.
[12] J. A. Henry, T. L. Zaugg, P. J. Myers, and M. A. Schechter, "Using therapeutic sound with progressive audiologic tinnitus management," *Trends Amplification*, vol. 12, no. 3, pp. 188–209, Jul. 2008, doi: 10.1177/1084713808321184.
[13] D. Hann, G. D. Searchfield, M. Sanders, and K. Wise, "Strategies for the selection of music in the short-term management of mild tinnitus," *Austral. New Zealand J. Audiol.*, vol. 30, no. 2, pp. 129–140, Nov. 2008, doi: 10.1375/audi.30.2.129.
[14] D. J. Levitin, "Why music moves us," *Nature*, vol. 464, no. 7290, pp. 834–835, Apr. 2010, doi: 10.1038/464834a.
[15] S. L. Robb, R. J. Nichols, R. L. Rutan, B. L. Bishop, and J. C. Parker, "The effects of music assisted relaxation on preoperative anxiety," *J. Music Therapy*, vol. 32, no. 1, pp. 2–21, Mar. 1995, doi: 10.1093/jmt/32.1.2.
[16] L. A. Mitchell and R. A. R. MacDonald, "An experimental investigation of the effects of preferred and relaxing music listening on pain perception," *J. Music Therapy*, vol. 43, no. 4, pp. 295–316, Dec. 2006, doi: 10.1093/jmt/43.4.295.
[17] S.-A. Li, L. Bao, and M. Chrostowski, "Investigating the effects of a personalized, spectrally altered music-based sound therapy on treating tinnitus: A blinded, randomized controlled trial," *Audiol. Neurotol.*, vol. 21, no. 5, pp. 296–304, Nov. 2016, doi: 10.1159/000450745.
[18] J.-M. Chen, P.-Y. He, and F. Pan, "Research on synthesizing music for tinnitus treatment based on chaos," in *Proc. 12th Int. Conf. Signal Process. (ICSP)*, Hangzhou, China, Oct. 2014, pp. 2286–2291.
[19] C. Jiemei, H. Peiyu, and P. Fan, "A new method of synthesizing chaotic music for tinnitus sound therapy," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Singapore, Jul. 2015, pp. 278–282.
[20] L. Wang, P. He, and F. Pan, "Research on fractal tones generating method for tinnitus rehabilitation based on musical instrument digital interface technology," *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi.*, vol. 31, no. 4, pp. 888–893, Aug. 2014.
[21] R. S. Tyler, "Neurophysiological models, psychological models, and treatments for tinnitus," in *Tinnitus Treatment: Clinical Protocols*, R. S. Tyler, Ed. New York, NY, USA: Thieme, 2006, pp. 1–22.
[22] J. Chen, P. He, and F. Pan, "A method of synthesizing tinnitus rehabilitation sound based on pentatonic scale and chaos," *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi.*, vol. 32, no. 6, pp. 1329–1334, Dec. 2015.
[23] N. Jin, "A method of personalized synthesizing tinnitus rehabiliation sound based on hyper-chaos," *J. Chengdu Univ. Inf. Technol.*, vol. 33, no. 4, pp. 359–364, Aug. 2018.
[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[25] Y. Duan, Y. L. V., and F.-Y. Wang, "Travel time prediction with LSTM neural network," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Rio de Janeiro, Brazil, Nov. 2016, pp. 1053–1058.
[26] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 6645–6649.
[27] H. Rathore, A. K. Al-Ali, A. Mohamed, X. Du, and M. Guizani, "A novel deep learning strategy for classifying different attack patterns for deep brain implants," *IEEE Access*, vol. 7, pp. 24154–24164, Mar. 2019, doi: 10.1109/ACCESS.2019.2899558.
[28] Y. Huang, X. Huang, and Q. Cai, "Music generation based on convolution-LSTM," *Comput. Inf. Sci.*, vol. 11, no. 3, pp. 50–56, Jun. 2018, doi: 10.5539/cis.v11n3p50.
[29] D. Eck and J. Schmidhuber, "A first look at music composition using LSTM recurrent neural networks," Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale, Manno, Switzerland, Tech. Rep. IDSIA-07-02, Mar. 2002.
[30] J. A. Russell, "A circumplex model of affect.," *J. Personality Social Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980, doi: 10.1037/h0077714.
[31] A. Working Group on Update of Genetics Evaluation Guidelines for the Etiologic Diagnosis of Congenital Hearing Loss; for the Professional Practice and G. Committee, "American college of medical genetics and genomics guideline for the clinical evaluation and etiologic diagnosis of hearing loss," *Genet. Med.*, vol. 16, no. 4, pp. 347–355, Mar. 2014, doi: 10.1038/gim.2014.2.
[32] A. McCombe, D. Baguley, R. Coles, L. McKenna, C. McKinney, and P. Windle-Taylor, "Guidelines for the grading of tinnitus severity: The results of a working group commissioned by the British association of otolaryngologists, head and neck surgeons, 1999," *Clin. Otolaryngology Allied Sci.*, vol. 26, no. 5, pp. 388–393, Oct. 2001.
[33] J. Salamon and E. Gomez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 6, pp. 1759–1770, Aug. 2012, doi: 10.1109/TASL.2012.2188515.
[34] M. Puckette, *The Theory and Technique of Electronic Music*. Hackensack, NJ, USA: World Scientific, 2007, pp. 7–8.
[35] M. Huesken and P. Stagge, "Recurrent neural networks for time series classification," *Neurocomputing*, vol. 50, pp. 223–235, Jan. 2003, doi: 10.1016/S0925-2312(01)00706-8.
[36] C. W. Omlin and C. L. Giles, "Extraction of rules from discrete-time recurrent neural networks," *Neural Netw.*, vol. 9, no. 1, pp. 41–52, 1996, doi: 10.1016/0893-6080(95)00086-0.
[37] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000, doi: 10.1162/089976600300015015.

[38] J. Heaton, *Introduction to Neural Network With Java*. Chesterfield, MO, USA: Heaton Research, Inc., 2005, pp. 126–128.

[39] S. McNally, J. Roche, and S. Caton, "Predicting the price of bitcoin using machine learning," in *Proc. 26th Euromicro Int. Conf. Parallel, Distrib. Network-based Process. (PDP)*, Cambridge, U.K., Mar. 2018, pp. 339–343.

[40] Accessed: Sep. 1, 2018. [Online]. Available: https://www.altumintelligence.com/articles/a/Time-Series-Prediction-Using-LSTM-Deep-Neural-Networks

[41] P. Campbell, "Is there such a thing as fractal music?" *Nature*, vol. 325, no. 6107, p. 766, Feb. 1987, doi: 10.1038/325766a0.

[42] R. F. Voss, "Fractals in nature: From characterization to simulation," in *The Science of Fractal Images*. New York, NY, USA: Springer, Aug. 1988, pp. 21–70.

[43] R. F. Voss and J. Clarke, "'1/f noise' in music and speech," *Nature*, vol. 258, pp. 317–318, Nov. 1975.

[44] Y. Shi, "Correlations of pitches in music," *Fractals*, vol. 4, no. 4, pp. 547–553, Dec. 1996, doi: 10.1142/S0218348X96000662.

[45] J. P. Boon, "Complexity, time and music," *Adv. Complex Syst.*, vol. 13, no. 2, pp. 155–164, Apr. 2010, doi: 10.1142/S0219525910002529.

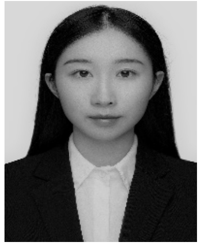
[46] W. Ebeling, T. Poschel, and K.-F. Albrecht, "Entropy, transinformation and word distribution of information-carrying sequences," *Int. J. Bifurcation Chaos*, vol. 05, no. 01, pp. 51–61, Feb. 1995, doi: 10.1142/S0218127495000041.

[47] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, Jul. 2002, Art. no. 068102, doi: 10.1103/PhysRevLett.89.068102.

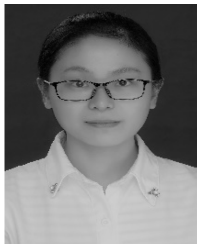
JIEMEI CHEN (Member, IEEE) received the B.S. degree in electronic and information engineering from the Chengdu University of Information Technology, in 2012, and the M.S. degree in signal and information processing from Sichuan University, Chengdu, China, in 2015, where she is currently pursuing the Ph.D. degree in signal and information processing. Her research interests include acoustic signal processing, biomedical signal processing, and machine learning.

FAN PAN (Member, IEEE) received the B.S. degree in electronic and information engineering, the M.S. and the Ph.D. degrees in signal and information processing from Sichuan University, Chengdu, China, in 2004, 2007, and 2017, respectively. He is currently a Lecturer with the College of Electronics and Information Engineering, Sichuan University. His research interests include biomedical signal processing, physiological measurement, acoustic signal processing, and machine learning.

PING ZHONG received the B.S. degree in clinical of integrated traditional Chinese and western medicine from the Chengdu University of Traditional Chinese Medicine (TCM), in 2008, the M.S. degree in otolaryngology of traditional Chinese medicine from the Chengdu University of TCM, in 2011, and the Ph.D. degree in otolaryngology of traditional Chinese medicine from the Guangzhou University of Chinese Medicine, in 2014. She is currently a Lecturer and an Attending Physician with the Hearing Center of Otorhinolaryngology-Head and Neck Surgery (ORL-HNS) Department, West China School of Medicine/West China Hospital, Sichuan University, Chengdu, China. Her research interests include the basic and clinical study of tinnitus and deafness. She is also a Committee Member of the Otolaryngology Branch of the China Association of Chinese Medicine (CACM) and a Council Member of the Otolaryngology Branch of World Federation of Chinese Medicine societies (WFCMS).



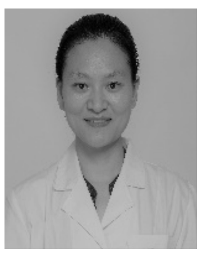
TIANTIAN HE received the B.S. degree in electronic information science and technology and the M.S. degree in signal and information processing from Sichuan University, Chengdu, China, in 2017 and 2020, respectively. Her research interest is acoustic signal processing.



LEIYU QI received the B.S. degree in electronic and information engineering from China West Normal University, in 2018. She is currently pursuing the M.S. degree in signal and information processing with Sichuan University, Chengdu, China. Her research interest is acoustic signal processing.

JINGZHE LU received the B.S. degree in clinical medicine of traditional Chinese and western medicine from the Henan University of Chinese Medicine, in 2011, the M.S. and the Ph.D. degrees in otolaryngology of traditional Chinese medicine from the Guangzhou University of Chinese Medicine, in 2014 and 2017, respectively. She is currently a Lecturer and an Attending Physician with the Hearing Center of ORL-HNS Department, West China School of Medicine/West China Hospital, Sichuan University, Chengdu, China. Her research interests include the basic and clinical study of tinnitus and deafness. She is also a Committee Member of the Otolaryngology Branch of CACM and a Council Member of the Otolaryngology Branch of WFCMS.

PEIYU HE received the B.S. degree in engineering from Tsinghua University, in 1986, the M.S. degree and the Ph.D. degree in communication and information system from Sichuan University, Chengdu, China, in 1989 and 2005, respectively. She is currently a Professor with the College of Electronics and Information Engineering, Sichuan University. Her research interests include acoustic signal processing, biomedical signal processing, array signal processing, and machine learning. She is also a Committee Member of the Sichuan Institute of Electronics and a Senior Member of the Chinese Institute of Electronics.



YUN ZHENG received the B.S. degree in medicine (otorhinolaryngology) from the West China University of Medicine, in 1997, the M.S. degree in audiology from Dalhousie University, in 2001, and the Ph.D. degree in medicine (otorhinolaryngology) from the West China Medical School, Sichuan University, Chengdu, China, in 2009. She is currently a Professor with the Hearing Center of ORL-HNS Department and the Director of the Hearing and Speech Science Laboratory, West China School of Medicine/West China Hospital, Sichuan University. Her research interests include the basic and clinical study of tinnitus and deafness. She is also the Chair of the Chinese Academy of Audiological Rehabilitation and a member of the standing committee of the Otolaryngology Branch of WFCMS.