



Development and Evaluation of an Intelligent System for Calibrating Karaoke Lyrics Based on Fuzzy Petri Nets

Yi-Nan Lin, Cheng-Ying Yang, Sheng-Kuan Wang, Gwo-Jen Chiou, Victor R.L. Shen, Yi-Chih Tung, Frank H.C. Shen & Hung-Chi Cheng

To cite this article: Yi-Nan Lin, Cheng-Ying Yang, Sheng-Kuan Wang, Gwo-Jen Chiou, Victor R.L. Shen, Yi-Chih Tung, Frank H.C. Shen & Hung-Chi Cheng (2022) Development and Evaluation of an Intelligent System for Calibrating Karaoke Lyrics Based on Fuzzy Petri Nets, Applied Artificial Intelligence, 36:1, 2110699, DOI: [10.1080/08839514.2022.2110699](https://doi.org/10.1080/08839514.2022.2110699)

To link to this article: <https://doi.org/10.1080/08839514.2022.2110699>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 22 Aug 2022.



Submit your article to this journal [↗](#)



Article views: 1673



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Development and Evaluation of an Intelligent System for Calibrating Karaoke Lyrics Based on Fuzzy Petri Nets

Yi-Nan Lin^a, Cheng-Ying Yang^b, Sheng-Kuan Wang^c, Gwo-Jen Chiou^d, Victor R. L. Shen^{e,g}, Yi-Chih Tung^a, Frank H.C. Shen^f, and Hung-Chi Cheng^e

^aDepartment of Electronic Engineering, Ming Chi University of Technology, New Taipei City, Taiwan, ROC; ^bDepartment of Computer Science, University of Taipei, Taipei, Taiwan, ROC; ^cDepartment of Electrical Engineering, Ming Chi University of Technology, New Taipei City, Taiwan, ROC; ^dDepartment of Electrical Engineering, National Formosa University, Yunlin County, Taiwan, ROC; ^eDepartment of Computer Science and Information Engineering, National Taipei University, New Taipei City, Taiwan, ROC; ^fDepartment of Electronic Engineering, Fu Jen Catholic University, New Taipei City, Taiwan, ROC; ^gDepartment of Information Management, Chaoyang University of Technology, Taichung City, Taiwan, ROC

ABSTRACT

In the home entertainment system, karaoke is a popular leisure facility in our daily life. Via the karaoke system, users can sing along with the lyrics based on the recordings of pop songs. However, a lot of karaoke systems can display lyrics semi-automatically. Traditionally, some lyrics are input manually and need to be synchronized with the tonal music stepwise, which is time-consuming. One of the famous musical phrase segmentation theories is a generative theory of tonal music, through which we have implemented a karaoke system in C# programming language. This intelligent system can automatically segment music phrases and use a high-level fuzzy Petri net model to calibrate the lyrics in pop songs. Fifty Chinese pop songs are selected to evaluate its performance. The experimental results have shown that the average calibration precision value (92.78%) and recall value (90.46%) are highly acceptable.

ARTICLE HISTORY

Received 30 May 2022
Revised 11 July 2022
Accepted 3 August 2022

Introduction

In the home entertainment system, karaoke playing has become a popular activity, which is enjoyed by many people. Via a karaoke system, users can sing along with lyrics based on the recordings of pop songs, and receive a score based on their performances (Tsai and Lee 2017). When reading the karaoke lyrics on the screen, users can recognize how to adjust their singing intonation to improve their scores, which is really simple and interesting. However, a difficulty in the traditional karaoke lyrics to be calibrated manually is encountered. It takes a large amount of time to segment the musical phrases, which is quite inefficient.

CONTACT Victor R.L. Shen  rlshen@mail.ntpu.edu.tw; victor.rlshen@msa.hinet.net 

Subject Area: Intelligent System Development

Present address of Victor R. L. Shen is Department of Information Management, Chaoyang University of Technology, Taichung City, Taiwan, ROC.

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Karaoke songs have become more difficult to make and the manufacturing costs are increasing. Thus, the selling price of a karaoke system is harder to be accepted by users.

In recent years, with advances in computer technology, the music albums have gradually been replaced by digital music purchased online. Nowadays, researchers suggest that automatically calibrating karaoke lyrics may be possible. Firstly, the musical phrase segmentation is required. One of the famous music segmentation theories is a generative theory of tonal music (GTTM) (Chan and Hsiao 2016; Frankland and Cohen 2018), which was conceived by a music theorist, Fred Lerdaahl, and a linguist, Ray Jackendoff; and presented in the book with the same title. Grouping preference rules (GPRs) are established in GTTM for the purpose of making a music grouping structure (Goto 2016). Quantification and testing of the local grouping rules for four local GPRs are provided (Frankland and Cohen 2018). In the approach based on the preference rules, the sound onset detection is applied to find all onsets in the song (Wu and Cao 2018).

The high-level fuzzy Petri net (HLFPN) model (Chiang et al. 2020; 2021; Lin et al. 2020a) is now used to perform the music segmentation. It is based on Petri net theory and fuzzy reasoning. Petri net theory, proposed by Dr Carl Adam Petri in 1962, is a graphical and mathematical modeling tool, which is concurrent, asynchronous, distributed, parallel, non-deterministic, stochastic, and so on. It can be used to model and analyze various systems (Sung and Chang 2019). However, along with the advances in information system, the description of Petri nets is becoming more and more complex with use of fuzzy production rules (Matsumoto and Hori 2019). In general, a fuzzy production rule can be used to describe the fuzzy relationship between antecedent and consequent.

Due to the above reasons and motivation, the goal of this paper is to make a karaoke system which can automatically calibrate karaoke songs with lyrics. A C# program has been designed to implement GTTM and HLFPN model for the purpose of segmenting each musical phrase. Musical phrases are used to automatically calibrate the lyrics that form a database system, and music notes are automatically calibrated for a Chinese character which has only one syllable. Then, an automatic calibration system for Chinese karaoke lyrics is completed.

Problem Statement: Although the karaoke technology has advanced with multi-functions for many years, it still lacks a robust database system for musical analysis. The grouping preference rules are diversified (Frankland and Cohen 2018). Also, the technical analysis parameters are hard to make musical decision. Consequently, the original semi-automatic lyrics calibration system needs to be further improved.

The remainder of this paper is organized as follows: [Section 2](#) discusses the literature review of GTTM which is applied to fuzzy reasoning of HLFPN model. The system architecture and music segmentation are described in [Section 3](#). The experimental results are discussed in [Section 4](#). Finally, the conclusion is remarked in [Section 5](#).

Literature Review

First, the background knowledge of a generative theory of tonal music (GTTM) rules and the musical context combination method are described. Second, some basic definitions regarding the HLFPN model, and the related works are presented.

Generative Theory of Tonal Music (GTTM)

The traditional method of musical phrase segmentation is based on psychological perspectives, and their relationship is discovered by psychology in music grouping. In 1920s, German psychologists proposed Gestalt theory to group similar objects. In Gestalt theory, they proposed the cognitive psychological aspects of musical groups, including three basic principles, namely, principle of proximity, principle of similarity, and principle of continuation.

In Gestalt theory, the research work in music grouping, composed of Jackendoff's and Lerdale's ideas, proposed a generative theory of tonal music (GTTM) (Chan and Hsiao 2016; Frankland and Cohen 2018). GTTM constitutes a formal description of the musical intuitions of a listener who is experienced in a musical idiom. It includes the grouping preference rules (GPRs) defined as follows:

Rule 1: *Single Event*: "Avoid analyses with very small groups- the smaller, the less preferable."

Rule 2: *Proximity*: "Consider a sequence of four notes, $n1 - n4$, the transition $n2 - n3$ may be heard as a group boundary if A. *Slur/Rest*: "the time interval from the end of $n2$ is greater than that from the end of $n1$ to the beginning of $n2$ and that from the end of $n3$ to the beginning of $n4$; or if B. *Attack-Point*: "the time interval between the attack points of $n2$ and $n3$ is greater than that between the attack points of $n1$ and $n2$ and that between the attack points of $n3$ and $n4$."

Rule 3: *Change*: "Consider a sequence of four notes, $n1 - n4$, the transition $n2 - n3$ may be heard as a group boundary if marked by Register, Dynamics, Articulation, Length, and Timbre," which are shown in [Figure 1-5](#).



Figure 1. Register change.



Figure 2. Dynamics change.



Figure 3. Articulation change.



Figure 4. Length change.



Figure 5. Timbre change.

High-Level Fuzzy Petri Net

Based on the original Petri net theory, scholars conduct their research works with evolutionary Petri nets one after another, such as colored Petri nets (Zhang, Zhao, and He 2022), timed Petri nets (Lin et al., 2020b; Lin et al. 2022), fuzzy Petri nets (Chowdhury et al. 2019), high-level fuzzy Petri nets (Chiang et al. 2020; Lin et al. 2020a). The HLFPN model was adopted to make decision on the segmentation of music phrases. It provides the characters of Petri net and fuzzy logic theories, which can be used to model fuzzy production rules and to conduct fuzzy reasoning (Chiang et al. 2021). The basic definitions and fuzzy reasoning approach are presented as follows:

Definition 1. The HLFPN is defined as an 8-tuple:

$$HLFPN = (P, T, F, C, V, \alpha, \beta, \delta)$$

where:

$P = \{p_1, p_2, \dots, p_k\}$: a finite set of places.

$T = \{t_1, t_2, \dots, t_l\}$: a finite set of transitions.

$$P \cup T = \emptyset$$

$F \subseteq (P \times T) \cup (T \times P)$: called the flow relation which is also a finite set of arcs, each representing the fuzzy set (i.e. fuzzy term) for an antecedent or a consequent; where the positive arcs (i.e. THEN parts) are denoted by \rightarrow .

$C = \{X, Y, Z\}$: A finite set of linguistic variables, e.g. X, Y , and Z , where $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_m\}$, $Z = \{z_1, z_2, \dots, z_q\}$.

$V = \{v_1, v_2, \dots, v_l\}$: A finite set of fuzzy truth values known as the fuzzy relational matrix between the antecedent and the consequent of a fuzzy production rule.

$\alpha : P \rightarrow C$: An association function, mapping from places to linguistic variables. $\alpha(p_i) = c_i, i = 1, \dots, I$, where $C = c_i$ is a set of linguistic variables in the knowledge base (KB) and is the number of linguistic variables in the KB;

$\beta : F \rightarrow \{0, 1\}$: An association function, mapping from the flow relations to the fuzzy truth values between zero and one.

$\delta : T \rightarrow V$: An association function, mapping from transitions to fuzzy relational matrices.

Definition 2. Input and Output Functions:

$I(t) = \{p \in P | (p, t) \in F\}$: A set of input places of transition t .

$I(p) = \{t \in T | (t, p) \in F\}$: A set of input transitions of place p .

$O(t) = \{p \in P | (t, p) \in F\}$: A set of output places of transition t .

$O(p) = \{t \in T | (p, t) \in F\}$: A set of output transitions of place p .

Definition 3. Membership Function:

The mapping function $Mem(p) : P \rightarrow [0, 1]$ assigns each place a real value, where $Mem(p) = DOM(\alpha(p))$, DOM represents the degree of membership in the associated proposition and data tokens are available in a set of places P .

Definition 4. Max-Min Compositional Rule:

In the HLFPN model, transition t , $V(t) = \min$ (fuzzy sets in $I(t)$); place p , $V(p) = \max$ (fuzzy sets in $I(p)$). The Max-Min composition operator is denoted by \circ .

Definition 5. Input Place, Hidden Place, and Output Place:

In the HLFPN model, place $p_i \in P$, if $t_j \in T, p_i \notin O(t_j)$, then p_i is called an input place (IP) of t_j , if $t_j \in T, p_i \notin I(t_j)$, then p_i is called an output place (OP) of t_j ; otherwise, p_i is called a hidden place.

Fuzzy Reasoning

For the fuzzy reasoning, fuzzy production rules are used (Lin et al. 2020a). In general, a fuzzy production rule is used to describe fuzzy relationship between the antecedent and the consequent. Let R be a set of fuzzy production rules, where $R = \{R_1, R_2, \dots, R_n\}$. The general form of the i -th fuzzy production rule R_i is shown as follows:

$$R_i : IF d_j(X \text{ is } A), THEN d_k(Y \text{ is } B); ELSE d_w(Z \text{ is } C) \dots (V)$$

where $d_j, d_k,$ and d_w denote propositions; X is called the input linguistic variable; Y and Z are called the output linguistic variables, respectively; A is called the input fuzzy set; B and C are called the output fuzzy sets, respectively; the fuzzy truth values of the propositions “ X is A ”, “ Y is B ” and “ Z is C ” are restricted to $[0, 1]$; d_j is the antecedent of a fuzzy production rule R_i ; d_k and d_w denote the consequents of the fuzzy production rule R_i . Let V represent the fuzzy relational matrix between antecedent and consequent of a fuzzy production rule.

Illustrative Example:

Let us consider the fuzzy production rule R_1 shown as follows:

$R_1 : IF$ the temperature(X_1) is hot(A_1) AND the sky(X_2) is cloudy(A_2), THEN the humidity(Y) is high(B).

Based on the transformation procedure presented in (Casey et al. 2018), the above fuzzy production rule R_1 is transformed into the following first-order logic form:

$$R'_1 : IF X_1(A_1) AND X_2(A_2), THEN Y(B)$$

Then, the HLFPN model is shown in Figure 6.

Assume that the fuzzy sets A_1, A_2 and B are shown as follows:

$$A_1 = \frac{0.20}{a_{11}} + \frac{0.54}{a_{12}} + \frac{0.24}{a_{13}}$$

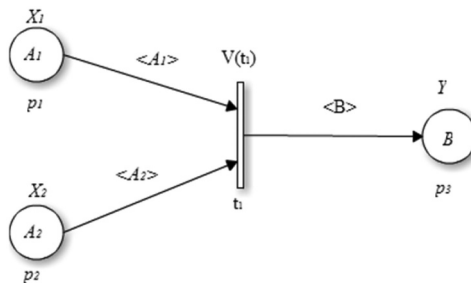


Figure 6. HLFPN model of illustrative example.

$$A_2 = \frac{0.11}{a_{21}} + \frac{0.77}{a_{22}} + \frac{0.42}{a_{23}}$$

$$B = \frac{0.31}{b_1} + \frac{0.66}{b_2} + \frac{0.17}{b_3}$$

By the cylindrical extension operations (Lin et al. 2020a), that is, a Cartesian product, we can obtain the antecedent fuzzy set A , shown as follows:

$$A = A_1 \times A_2 = (0.200.540.24)^{T \wedge} (0.110.770.42) = \begin{bmatrix} 0.11 & 0.20 & 0.20 \\ 0.11 & 0.54 & 0.42 \\ 0.11 & 0.24 & 0.24 \end{bmatrix}$$

Then, the fuzzy relational matrices $V_1(t_1)$, $V_2(t_2)$, and $V_3(t_3)$ between antecedent and consequent of the fuzzy production rule R_1 can be obtained, shown as follows:

$$V_1(t_1) = \begin{bmatrix} 0.11 & 0.20 & 0.20 \\ 0.11 & 0.31 & 0.31 \\ 0.11 & 0.24 & 0.24 \end{bmatrix} \in A \times B \times b_1$$

$$V_2(t_2) = \begin{bmatrix} 0.11 & 0.20 & 0.20 \\ 0.11 & 0.54 & 0.42 \\ 0.11 & 0.24 & 0.24 \end{bmatrix} \in A \times B \times b_2$$

$$V_3(t_3) = \begin{bmatrix} 0.11 & 0.17 & 0.17 \\ 0.11 & 0.17 & 0.17 \\ 0.11 & 0.17 & 0.17 \end{bmatrix} \in A \times B \times b_3$$

The most widely used fuzzy reasoning method is the max–min composition inference (Yang et al. 2019). Assume that the input fuzzy sets A_1' and A_2' are shown as follows:

$$A_1' = \frac{0.09}{a_{11}} + \frac{0.85}{a_{12}} + \frac{0.29}{a_{13}}$$

$$A_2' = \frac{0.29}{a_{21}} + \frac{0.89}{a_{22}} + \frac{0.45}{a_{23}}$$

Then, we can get

$$A_1' \circ V_1(t_1) = (0.090.850.29) \circ V_1(t_1) = (0.110.310.31)$$

$$A_1' \circ V_2(t_1) = (0.090.850.29) \circ V_2(t_1) = (0.110.540.42)$$

$$A_1' \circ V_3(t_1) = (0.090.850.29) \circ V_3(t_1) = (0.110.170.17)$$

Finally, we can obtain

$$\begin{aligned}
 B' &= (0.290.890.45) \circ \begin{bmatrix} 0.11 & 0.11 & 0.11 \\ 0.31 & 0.54 & 0.17 \\ 0.31 & 0.42 & 0.17 \end{bmatrix} \\
 &= (0.310.540.17) \\
 &= \frac{0.31}{b_1} + \frac{0.54}{b_2} + \frac{0.17}{b_3}
 \end{aligned}$$

The above description is the fuzzy reasoning process of HLFPN model.

Fuzzy Reasoning Algorithm

In this sub-section, a fuzzy reasoning algorithm (FRA) (Chiang et al. 2020) is briefly reviewed to determine whether there exists a fuzzy relational matrix between antecedent and consequent of a fuzzy production rule or not.

INPUT: Input Place (IP), $Mem(p_i)$ $p_i \in IP$, (or fuzzy set), where IP denotes a set of input places.

OUTPUT: Output Place (OP), $Mem(p_i)$ $p_i \in OP$, (or fuzzy set), where OP denotes a set of output places.

Procedure

Step 1.: Initially, assume that only the Degree of Memberships (DOMs) in the propositions operating on input variables are available. Consequently, the initial marking function is shown as follows:

$$M(p_i) = 0, \text{ if } p_i \notin IP$$

$$M(p_i) = \text{the number of data tokens, if } p_i \in IP$$

Step 2.: Calculate the fuzzy relational matrices $V(t_i)$ of the current transition, and use the input to perform cylindrical extension.

$t_j \in T$, compute $V(t_j) = W_a \times W_c = (w_{a1}, w_{a2}, \dots, w_{am})^T \wedge (w_{c1}, w_{c2}, \dots, w_{cn})$, where T denotes a set of transitions; $V(t_j)$ is a fuzzy relational matrix between the antecedent and the consequent of rule t_j ; $W_a = (w_{a1}, w_{a2}, \dots, w_{am})$ is a fuzzy set of weights in the antecedent; $W_c = (w_{c1}, w_{c2}, \dots, w_{cn})$ is a fuzzy set of weights in the consequent; and each element of a fuzzy set is denoted as a fuzzy weight interval.

Step 3.: Input a data pattern to W_a -input.

Step 4. Fire the enabled transitions and execute Zadeh's max-min operation. Let t_j be any enabled transition. Then, compute $t_j \in T / p_k, M(p_k) =$ the number of data tokens

$$W'_a = W_a\text{-input}$$

$$W'_c = W'_a *_{\circ} V(t_j) \text{ or } \neg W'_a *_{\circ} V(t_j)$$

if an ELSE part is available.

Step 5. Send the results to the output place. For every output variable O , its associated membership distribution is $W'_c = \{w_{ci}'\} = \bigvee w_{ci}', i = 1, 2, \dots, l$, where l denotes the in-degree of output variable O . Then, W'_c becomes an actual output.

Step 6. Go back to Step 4, while $\exists t_j \in T / M(p_i) = 1, p_i \in IP(t_j)$ (That is, while the enabled transitions still exist, go to Step 4).

Step 7. The weighted average defuzzification method is applied, and the real operation value is obtained.

IEEE 1599

“Music is much more than listening to audio encoded in some unreadable binary format” (Wick, Hartelt, and Puppe 2020). So, we need to integrate each karaoke song, score, information, and musical phrase. The standard is named as IEEE1599 (Baggi 2015), proposed by Denis L. Baggi and Goffredo M. Haus in 2009. IEEE 1599 is a standard to encode music with XML symbols. The main distinguished characteristics use blocks to represent music and the concept of layers, as shown in Figure 7.

General Layer: The layer provides a general description of the music. It contains the following items: music title, authors, number, data, type, and so on. *Structural Layer:* The layer stores information of musical structure. *Logical Layer:* The layer provides music description from a symbolic point and includes scores with symbols, as well as the spine with logically organized symbols, and a sorted list of music events. *Notational Layer:* The layer includes notations of music, like a great variety of pictures. *Performance Layer:* The layer contains the information about musical performance, including midi, Csound, etc. *Audio Layer:* The audio layers including media contents are still available in their original encoding.

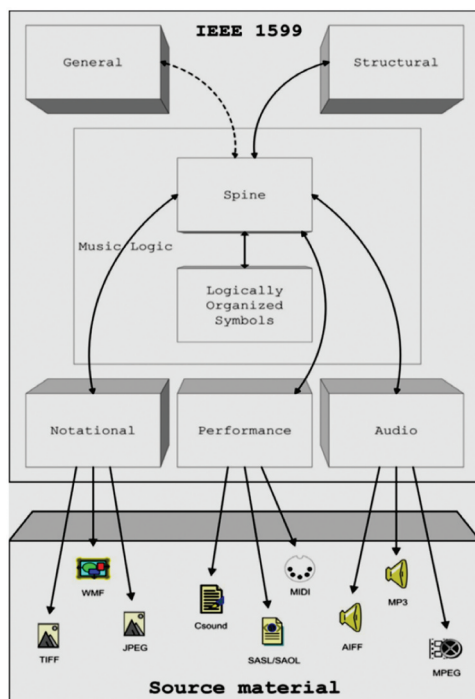


Figure 7. All layers in IEEE 1599 (Wick, Hartelt, and Puppe 2020).

Related Works

According to reference (Meier 2013; Wu and Cao 2018), for the problem of song sentiment classification based on Chinese lyrics, there are two methods that have been widely used at present, namely, a sentiment-dictionary-based method and a machine-learning-based method. The former sentiment dictionary is difficult to expand, while the latter may cause “dimensional disaster” when generating feature vectors. It proposes a song sentiment classification method that combines the advantages of these two ideas. The method obtains the positive and negative tendency values by the process of word segmentation and weight calculation based on sentiment dictionary, which are used as feature vectors. Then, we use Logistic Regression algorithm to carry binary classification of Chinese songs. According to reference (Shen and Chen 2014), it was the initial idea to develop this lyrics calibration system, which had lower precision value (71.43%) and no recall values available. More useful parameters and detailed discussions are needed to improve its system performance. The fuzzy reasoning algorithm also needs to be further improved. According to reference (Bilbao et al. 2020; Ruangsang and Assawinchaichote 2019), as the computational costs for physical modeling synthesis of sound are often much greater than those for conventional sound synthesis methods, most techniques

currently rely on simplifying assumptions. They include digital waveguides and modal synthesis methods. However, it can be difficult to approach some of the more detailed behavior of musical instruments.

Proposed Approach

For a karaoke system, it is important to make musical phrases, which are concepts and practices related to grouping consecutive melodic notes in performance and composition. A musical phrase is a unit that has a complete musical sense of its own length, in which a singer or instrumentalist can sing or play in one breath. We have developed a C# program to implement GTTM and HLFPN model to segment musical phrases. Musical phrases are used to automatically calibrate the lyrics that form a database system. And the music notes are automatically calibrated for a Chinese character, which has only one syllable.

Data Preprocessing System

A phrase segmentation technique is highly required because a phrase is the basic unit for organizing and analyzing the music signals. In music scores, it is usually marking a phrase with a slur, as shown in Figure 8.

For phrase segmentation, as rule 3 in grouping preference, “Consider a sequence of four notes, $n_1 - n_4$, the transition $n_2 - n_3$ may be heard as a group boundary if marked by register, dynamics and length.” (Haus and Ludovico 2017; Haus, Longari 2019; Hu, Chen, and Yin 2020), the score function (Liu 2010) is defined for the grouping preference rule 3 of GTTM as follows:

The register change rule is assumed that the pitches (in *cent*) of the four music notes $n_1 n_2 n_3 n_4$ are $c_1 c_2 c_3 c_4$. If $|c_3 - c_2| > |c_2 - c_1|$ and $|c_3 - c_2| > |c_4 - c_3|$, then it is defined as:

$$f_{register} = \left\{ 1 - \frac{|c_2 - c_1| + |c_4 - c_3|}{2|c_3 - c_2|} \right\} \quad (1)$$

else, defined as $f_{register} = 0$. It represents that the pitch has not been changed.

Similarly, the dynamics change rule is assumed that the intensities (in *dB*) of the four music notes $n_1 n_2 n_3 n_4$ are $d_1 d_2 d_3 d_4$. If $|d_3 - d_2| > |d_2 - d_1|$ and $|d_3 - d_2| > |d_4 - d_3|$, then it is defined as



Figure 8. A phrase.

$$f_{dynamics} = \left\{ 1 - \frac{|d_2 - d_1| + |d_4 - d_3|}{2|d_3 - d_2|} \right\} \quad (2)$$

else, defined as $f_{dynamics} = 0$. It represents the intensity not being changed.

Also, the length change rule is assumed that the lengths (in *second*) of the four music notes $n_1n_2n_3n_4$ are $l_1l_2l_3l_4$ defined as:

$$f_{length} = \left\{ 1 - \frac{Min(l_2, l_3)}{Max(l_2, l_3)} \right\} \quad (3)$$

First, the membership function can be defined in accordance with the input parameters. The fuzzy reasoning rule and the corresponding HLFPN model can be established. The $f_{register}$, $f_{dynamics}$ and f_{length} associated with the membership functions for the fuzzifier are input into the HLFPN model. With the characters of fuzzy reasoning, each input parameter is reasoned by the fuzzy rule. Finally, all indices are integrated to make decision on the musical phrase segmentation and to perform automatic calibration of karaoke lyrics (Bastanfard, Amirkhani, and Naderi 2020; Chen 2021).

HLFPN Model Based on GTTM

To make more precise in the musical phrase segmentation, it is a good way to implement grouping preference rules of GTTM. The decision based on input parameters uses the calculated values to set the position of segmentation. However, an input parameter needs to be considered as a fuzzy term, high or low. Grouping preference rules of GTTM including three parameters are used to determine where the position of performance or segmentation is. In these unclear, obscure, and ambiguous cases, it is appropriate to use the fuzzy logic theory. Thus, the HLFPN model is applied to make decision (Shen et al. 2017), as shown in Figure 9.

Calibration Mechanism in Membership Degrees

According to the corresponding membership functions and fuzzy sets defined, we input each parameter to a fuzzifier and calculate the membership degrees in each set. Based on the size of the input parameter, each parameter is transformed to 「If... then...」 statement and the fuzzy rules are established. In the decision method, three parameters are used, namely, the $f_{register}$, $f_{dynamics}$ and f_{length} . As a result, each index is defined in membership degrees of *Low* and *High*. Then, three sets of technical indices include the S-Type and Z-Type membership degrees, as shown in Figure 10.

In addition, the decision is also divided into two cases, namely, “Performance” and “Segmentation.” Therefore, two sets of performance and segmentation belong to the Λ -Type of membership degree, as shown in Figure 11.

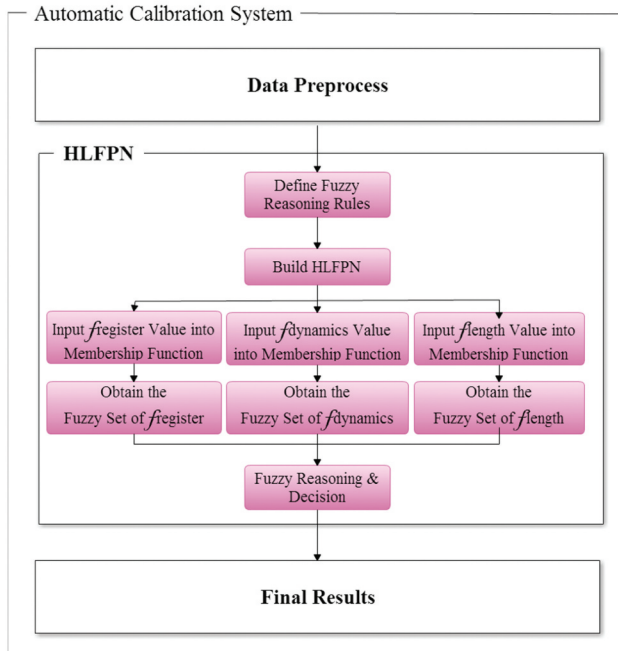


Figure 9. Block diagram of HLFPN model.

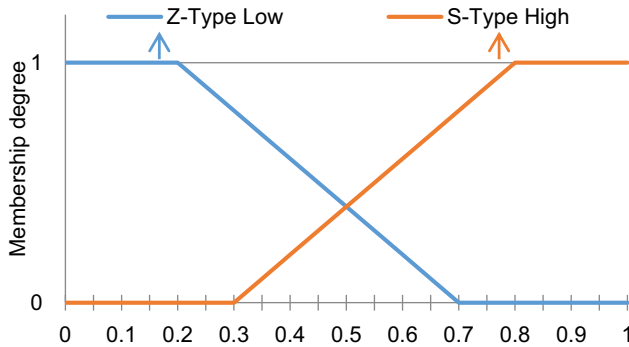


Figure 10. The type of membership degrees for input parameter.

In this paper, three input parameters are used, namely, the $f_{register}$, $f_{dynamics}$ and f_{length} . As a result, the membership functions of “high” and “low” are defined for each input parameter. Therefore, all of them are defined in Table 1. The membership degrees of the decision on music performance and segmentation are listed in Table 2.

In the analysis, the values of each input parameter are usually set between 0 and 1. When $f_{register}$, $f_{dynamics}$, and $length$ values larger than 0.3 are high, and those values less than 0.7 are low. However, due to the membership degree as defined above, whose scope of a variable is set

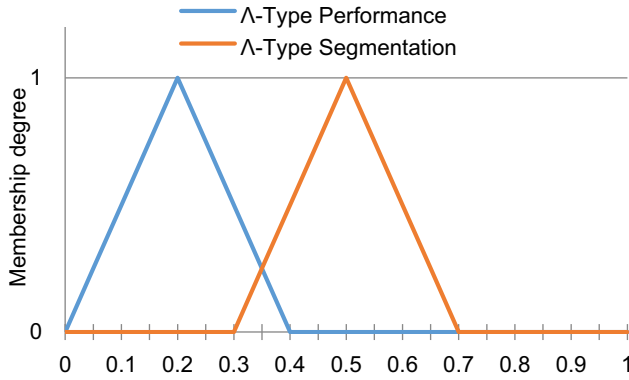


Figure 11. The type of membership degrees for technical decision.

Table 1. Membership degrees of input parameter.

Input Parameter	Low		High	
$f_{register}$	0.2	0.7	0.3	0.8
$f_{dynamics}$	0.2	0.7	0.3	0.8
f_{length}	0.2	0.7	0.3	0.8

Table 2. Membership degrees of technical decision.

Performance			Segmentation		
0.0	0.2	0.4	0.3	0.5	0.7

between 0 and 1. Thus, the values of each input parameter must be converted to the values between 0 and 1 before they are substituted into the calculation of basic rules.

Assume that v_H and v_L values represent membership degrees of “high” and “low,” respectively. The fuzzification procedure is defined as follows:

$$v_H(x) = \begin{cases} 1, & x \geq 0.8 \\ 2(x - 0.4), & 0.3 < x < 0.8 \\ 0, & x \leq 0.3 \end{cases} \tag{4}$$

$$v_L(x) = \begin{cases} 0, & x \leq 0.2 \\ 2(x - 0.1), & 0.7 > x > 0.2 \\ 1, & x \geq 0.7 \end{cases} \tag{5}$$

$$v_{Segmentation}(x) = \begin{cases} 0, & x \geq 0.7 \\ -\frac{1}{0.3}(x - 1.0), & 0.5 \leq x < 0.7 \\ \frac{1}{0.3}(x - 0.4), & 0.3 < x < 0.5 \\ 0, & x \leq 0.3 \end{cases} \tag{6}$$

$$v_{Performance}(x) = \begin{cases} 0, & x \geq 0.4 \\ -5(x - 0.5), & 0.2 \leq x < 0.4 \\ 5(x - 0.0), & 0.0 < x < 0.2 \\ 0, & x \leq 0.0 \end{cases} \quad (7)$$

Fuzzy Reasoning and Building HLFPN Model

According to fuzzy sets and their corresponding membership degrees as defined previously, each input parameter is imported to a fuzzifier and the technical index is calculated to find its own membership degree. Based on the size of an input parameter, it has been transformed into an [If... then ...] statement in order to establish a rule base. The activity diagram of fuzzy reasoning is shown in Figure 12.

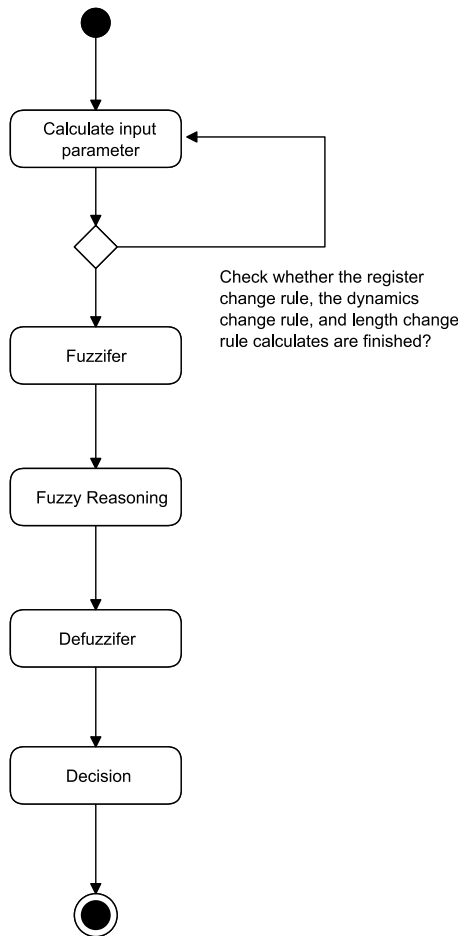


Figure 12. The activity diagram of fuzzy reasoning.

Assume that $f_{register}$, $f_{dynamics}$, and f_{length} denote input linguistic variables with fuzzy terms: high (H) and low (L). And assume that the decision (D) is an output linguistic variable with fuzzy terms: strong (S) and weak (W). The fuzzy production rules are defined as follows:

- R1: If $f_{register}$ is H , then D is S .
- R2: If $f_{register}$ is L , then D is W .
- R3: If $f_{dynamics}$ is H , then D is S .
- R4: If $f_{dynamic}$ is L , then D is W .
- R5: If f_{length} is H , then D is S .
- R6: If f_{length} is L , then D is W .

Then, the fuzzy production rules are transformed into the HLFPN model, as shown in Figure 13, and the parameters are listed in Table 3.

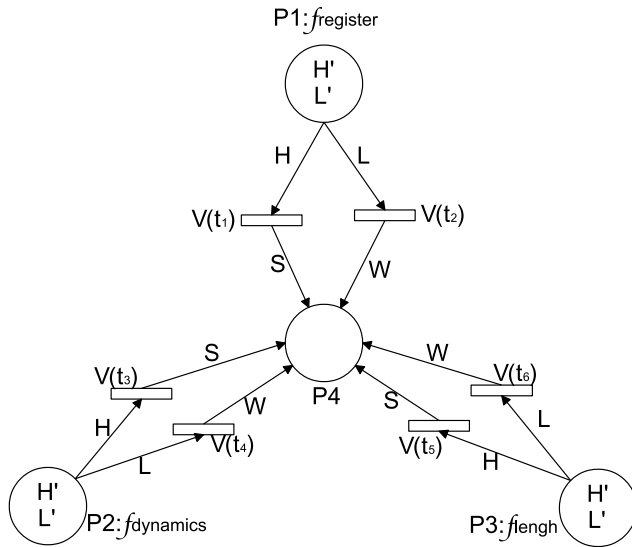


Figure 13. The HLFPN model representing six fuzzy production rules.

Table 3. Description of parameters.

Name of Parameter	Description of Parameter
$f_{register}$	Denotes the value of $f_{register}$ in input place P_1 .
$f_{dynamics}$	Denotes the value of $f_{dynamics}$ in input place P_2 .
f_{length}	Denotes the value of f_{length} in input place P_3 .
D	Denotes a decision in an output place P_4 .
H, L	Denote high and low fuzzy sets, respectively.
S, W	Denote strong and weak fuzzy sets, respectively.
$V(ti), i = 1, 2, 3, \dots, 6$	Denotes the fuzzy relational matrix of $f_{register}, f_{dynamics}, f_{length}$ and decision.
H', L'	Denote high and low fuzzy sets of input value.

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ieee1599 SYSTEM
"http://standards.ieee.org/downloads/1599/1599-2008/ieee1599.dtd">
<ieee1599 creator="IEEE1599 Generator" version="1.0">
<general>...</general>
< logic >...</ logic >
< structural >...</ structural >
< notational >...</ notational >
< performance >...</ performance >
< audio >...</ audio >
</ieee1599>

```

Figure 14. Layers in IEEE 1599.

```

<general>
<description>
<main_title>安靜(Silence)</main_title>
<author type="作詞者">周杰倫(Jay Chou)</author>
<author type="作曲者">周杰倫(Jay Chou)</author>
<other_title/>
<number>1</number>
<work_title>范特西(Fantasy)</work_title>
<work_number/>
<date type="發行日期">2001</date>
<genres>
<genre weight="100" description="流行音樂" name="POP"/>
</genres>
</description>

```

Figure 15. Description in general layer of IEEE 1599.

According to the proposed HLFPN model, the fuzzy inference is performed with the technical indices of a fuzzier and the fuzzy rules in the rule base. Within the process of fuzzy inference, the standard operators for calculations are used. Finally, the crisp outputs are obtained to decide performance or segmentation by the weighted average defuzzification method, and the real operation value of “Decision” can be computed.

Preprocessing Results

However, the music lyrics, score, and other music information are integrated into an automatic calibration system. A new music standard named *IEEE 1599* and a software system are used to show the preprocessing results. The IEEE

```

<related_files>
<related_file description="MP3" notes="" copyright="" encoding_format="audio_mp3"
file_format="audio_mp3" file_name="Silence.mp3"/>
<related_file description="MID" notes="" copyright=""
encoding_format="audio_x_midi" file_format="audio_x_midi"
file_name="Silence.mid"/>
<related_file description="COVER" notes="" copyright=""
encoding_format="image_jpeg" file_format="image_jpeg" file_name="Silence.jpg"/>
<related_file description="MV" notes="" copyright="" encoding_format="video_mpeg"
file_format="video_mpeg" file_name="Silence.flv"/>
<related_file description="SCORE" notes="" copyright=""
encoding_format="text_html" file_format="text_html" file_name="Silence.xml"/>
</related_files>
<analog_media>
<analog_medium description="CD" notes="None" copyright=""/>
</analog_media>
<notes/>
</general>

```

Figure 16. The related files in general layer of IEEE 1599.

1599 standard (Baggi 2015) is used to encode music with XML symbols. This is an example in IEEE 1599 from a song named “Silence,” composed by Jay Chou, as shown in Figures 14–17.

Experimental Results

To demonstrate that the proposed HLFPN-based calibration system is feasible, this Section is intended to evaluate the proposed system performance. A fuzzy reasoning algorithm (FRA) (Chiang et al. 2020) is used to determine whether a fuzzy relational matrix exists between antecedent and consequent of a fuzzy rule. Then, our proposed approach is applied to the illustrative example.

Example of HLFPN Model for Fuzzy Reasoning

In this sub-section, an example is used to illustrate the viability of the fuzzy reasoning process. Two digits of significant numbers represent the result of each calculation.

Step 1: Initially, assume that only the DOMs in the propositions operating input variables are available. Assume that four fuzzy sets are shown as follows:

$$H = \frac{0.00}{h_L} + \frac{0.70}{h_H}, L = \frac{0.30}{l_L} + \frac{0.00}{l_H}, S = \frac{0.00}{s_L} + \frac{0.70}{s_H}, W = \frac{0.30}{w_L} + \frac{0.00}{w_H}$$

Step 2: Compute the fuzzy relational matrices, shown as follows:

```

- <los>
- <lyrics voice_ref="null" part_ref="piano">
  <syllable hyphen="no" end_event_ref="P1" start_event_ref="P1"/>
  <syllable hyphen="no" end_event_ref="P2" start_event_ref="P2">
    只剩下鋼琴陪我彈了一天</syllable>
  <syllable hyphen="no" end_event_ref="P3" start_event_ref="P3">
    睡著的大提琴 安靜的舊舊的</syllable>
  <syllable hyphen="no" end_event_ref="P4" start_event_ref="P4">
    安靜的舊舊的</syllable>
  <syllable hyphen="no" end_event_ref="P5" start_event_ref="P5">
    我想你已表現的非常明白</syllable>
  <syllable hyphen="no" end_event_ref="P6" start_event_ref="P6">
    我懂我也知道 你沒有捨不得</syllable>
  <syllable hyphen="no" end_event_ref="P7" start_event_ref="P7"/>
  <syllable hyphen="no" end_event_ref="P8" start_event_ref="P8">
    你說你也會難過我不相信</syllable>
  <syllable hyphen="no" end_event_ref="P9" start_event_ref="P9">
    牽著你陪著我 也只是曾經</syllable>
  <syllable hyphen="no" end_event_ref="P10"
    start_event_ref="P10"> 希望他是真的比我還要愛你</syllable>
  <syllable hyphen="no" end_event_ref="P11"
    start_event_ref="P11"> 我才會逼自己離開</syllable>
  <syllable hyphen="no" end_event_ref="P12"
    start_event_ref="P12"/>
  <syllable hyphen="no" end_event_ref="P13"
    start_event_ref="P13"> 你要我說多難堪 我根本不想分開</syllable>
  <syllable hyphen="no" end_event_ref="P14"
    start_event_ref="P14"> 你要我說多難堪 我根本</syllable>
  <syllable hyphen="no" end_event_ref="P15"
    start_event_ref="P15"> 為什麼還要我用淚來帶過</syllable>
  <syllable hyphen="no" end_event_ref="P16"
    start_event_ref="P16"> 我沒有這種天份 包容你也接受他</syllable>
  <syllable hyphen="no" end_event_ref="P17"
    start_event_ref="P17"> 不用熱心的太多 我會一直好好過</syllable>
  <syllable hyphen="no" end_event_ref="P18"
    start_event_ref="P18"> 你已經遠遠離開 我也會慢慢走開</syllable>
  <syllable hyphen="no" end_event_ref="P19"
    start_event_ref="P19"> 為什麼我連分開都還就著你</syllable>
  <syllable hyphen="no" end_event_ref="P20"
    start_event_ref="P20"> 我真的沒有天份 安靜的沒這麼快</syllable>
  <syllable hyphen="no" end_event_ref="P21"
    start_event_ref="P21"> 我會學著放棄你 是因為我太愛你</syllable>
  <syllable hyphen="no" end_event_ref="P22"
    start_event_ref="P22"/>
  <syllable hyphen="no" end_event_ref="P23"
    start_event_ref="P23"> 只剩下鋼琴陪我彈了一天</syllable>
  <syllable hyphen="no" end_event_ref="P24"
    start_event_ref="P24"> 睡著的大提琴 安靜的舊舊的</syllable>
  <syllable hyphen="no" end_event_ref="P25"
    start_event_ref="P25"> 我想你已表現的非常明白</syllable>
  <syllable hyphen="no" end_event_ref="P26"
    start_event_ref="P26"> 我懂我也知道 你沒有捨不得</syllable>
  <syllable hyphen="no" end_event_ref="P27"
    start_event_ref="P27"/>
  <syllable hyphen="no" end_event_ref="P28"
    start_event_ref="P28"> 你說你也會難過我不相信</syllable>
  <syllable hyphen="no" end_event_ref="P29"
    start_event_ref="P29"> 牽著你陪著我 也只是曾經</syllable>
  <syllable hyphen="no" end_event_ref="P30"
    start_event_ref="P30"> 希望他是真的比我還要愛你</syllable>

```

Figure 17. Layers in IEEE 1599.

$$V(t_1) = H * S = (0.00 \ 0.70)^T \wedge (0.00 \ 0.70) = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.70 \end{bmatrix}$$

$$V(t_3) = V(t_1) = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.70 \end{bmatrix}$$

$$V(t_5) = V(t_1) = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.70 \end{bmatrix}$$

$$V(t_2) = L * W = (0.30 \ 0.00)^T \wedge (0.30 \ 0.00) = \begin{bmatrix} 0.30 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$$

$$V(t_4) = V(t_2) = \begin{bmatrix} 0.30 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$$

$$V(t_6) = V(t_2) = \begin{bmatrix} 0.30 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$$

Step 3: Input data pattern, shown as follows:

$$H'_{register} = \frac{0.00}{h'_L} + \frac{0.20}{h'_H}$$

$$L'_{register} = \frac{0.80}{l'_L} + \frac{0.00}{l'_H}$$

$$H'_{fdynamics} = \frac{0.00}{h'_L} + \frac{0.80}{h'_H}$$

$$L'_{fdynamics} = \frac{0.00}{l'_L} + \frac{0.00}{l'_H}$$

$$H'_{flength} = \frac{0.00}{h'_L} + \frac{0.40}{h'_H}$$

$$L'_{flength} = \frac{0.20}{l'_L} + \frac{0.00}{l'_H}$$

Step 4: Fire the enabled transitions:

$$S'_{register} = H'_{register} \circ V(t_1) = [0.00 \ 0.20] \circ \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.70 \end{bmatrix} = [0.00 \ 0.20]$$

$$W'_{register} = L'_{register} \circ V(t_2) = [0.80 \ 0.00] \circ \begin{bmatrix} 0.30 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} = [0.30 \ 0.00]$$

$$\begin{aligned} S'_{fdynamics} &= H'_{fdynamics} \circ V(t_3) = [0.00 \ 0.80] \circ \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.70 \end{bmatrix} \\ &= [0.00 \ 0.70] \end{aligned}$$

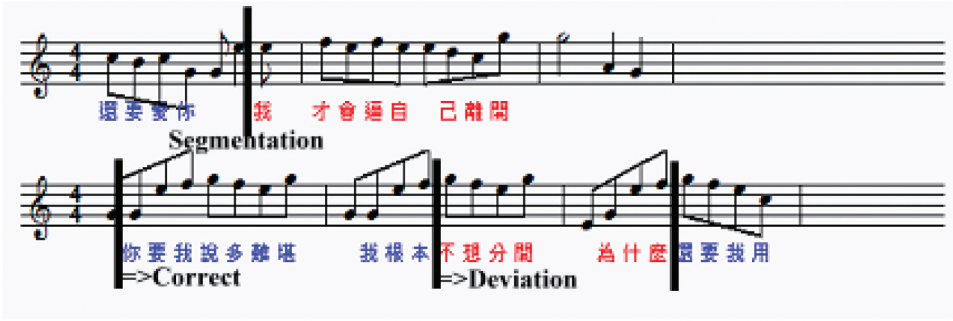


Figure 18. Deviation in automated system.



Figure 19. The actual score.

$$W'_{dynamics} = L'_{dynamics} \circ V(t_4) = [0.00 \quad 0.00] \circ \begin{bmatrix} 0.30 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} = [0.00 \quad 0.00]$$

$$S'_{flength} = H'_{flength} \circ V(t_5) = [0.00 \quad 0.40] \circ \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.70 \end{bmatrix} = [0.00 \quad 0.40]$$

$$W'_{flength} = L'_{flength} \circ V(t_6) = [0.20 \quad 0.00] \circ \begin{bmatrix} 0.30 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} = [0.20 \quad 0.00]$$

Steps 5–6: Finally, the fuzzy reasoning result is shown as follows:

$$\begin{aligned} D &= S'_{fregister} \cup W'_{fregister} \cup S'_{fdynamics} \cup W'_{fdynamics} \cup S'_{flength} \cup W'_{flength} \\ &= \frac{0.30}{d_L} + \frac{0.70}{d_H} \end{aligned}$$

Step 7: If the weighted average defuzzification method is applied, then the real operation value of “Decision” can be computed, shown as follows:

$$decision = \frac{0.30 \times d_L + 0.70 \times d_H}{0.30 + 0.70} = \frac{0.30 \times 0.30 + 0.70 \times 0.70}{0.30 + 0.70} = 0.58$$

Table 4. Experimental results.

No.	Number of Phrases	Number of Phrases Correctly Detected	Number of Phrases Deviated	Number of Phrases Detected	Recall (%)	Precision (%)
1	43	39	3	42	90.70	92.86
2	46	39	4	43	84.78	90.70
3	49	45	3	48	91.84	93.75
4	48	42	3	45	87.50	93.33
5	22	19	2	21	86.36	90.48
6	34	31	2	33	91.18	93.94
7	27	25	2	27	92.59	92.59
8	34	31	3	34	91.18	91.18
9	49	45	3	48	91.84	93.75
10	44	40	3	43	90.91	93.02
11	24	22	1	23	91.67	95.65
12	40	35	3	38	87.50	92.11
13	39	36	2	38	92.31	94.74
14	27	24	2	26	88.89	92.31
15	45	39	3	42	86.67	92.86
16	43	40	2	42	93.02	95.24
17	48	43	3	46	89.58	93.48
18	36	33	2	35	91.67	94.29
19	39	35	3	38	89.74	92.11
20	32	29	2	31	90.63	93.55
21	48	44	2	46	91.67	95.65
22	44	39	4	43	88.64	90.70
23	42	38	3	41	90.48	92.68
24	43	39	3	42	90.70	92.86
25	53	47	5	52	88.68	90.38
26	36	33	2	35	91.67	94.29
27	49	45	3	48	91.84	93.75
28	31	28	2	30	90.32	93.33
29	35	32	2	34	91.43	94.12
30	53	47	4	51	88.68	92.16
31	60	54	4	58	90.00	93.10
32	44	40	3	43	90.91	93.02
33	36	32	3	35	88.89	91.43
34	59	53	4	57	89.83	92.98
35	45	41	3	44	91.11	93.18
36	38	34	3	37	89.47	91.89
37	40	36	3	39	90.00	92.31
38	42	38	3	41	90.48	92.68
39	46	42	3	45	91.30	93.33
40	50	45	4	49	90.00	91.84
41	61	56	4	60	91.80	93.33
42	46	42	4	46	91.30	91.30
43	47	43	3	46	91.49	93.48
44	22	20	2	22	90.91	90.91
45	40	36	3	39	90.00	92.31
46	37	33	3	36	89.19	91.67
47	32	32	0	32	100.0	100.0
48	36	33	2	35	91.67	94.29
49	58	52	4	56	89.66	92.86
50	34	30	2	32	88.24	93.75
Average					90.46	92.78

Table 5. Results of functional comparison.

Functions Approaches	Correct Phrase Detection	Deviated Phrase Detection	Missing Phrase Detection	Recall (%)	Precision (%)
Our Proposed	√	√	√	90.46	92.78
X. Wu's et al	√	√			√
S. Bilbao's et al	√	√			√

Due to $0.3 \leq decision < 0.7$, the decision is made as “Segmentation.”

Main Results

The experimental results illustrate the effectiveness of the proposed system. An automatic calibration system for karaoke lyrics is useful, but its musical phrase segmentation may have a little deviation. The comparison results are shown in Figures 18 and 19.

According to manual experiments, the automatic music segmentations which are measured by the recall and the precision values defined as follows:

$$recall = \left\{ \frac{\text{number of phrases correctly detected}}{\text{number of phrases}} \right\} \times 100\% \quad (11)$$

$$precision = \left\{ \frac{\text{number of phrases correctly detected}}{\text{number of phrases detected}} \right\} \times 100\% \quad (12)$$

Finally, in the experiment, 50 Chinese pop songs were used. In total, there are 2076 musical phrases manually labeled. As shown in Table 4, there are 1463 musical segmentations that are correctly detected by the proposed system. Therefore, the average recall is 90.46% and the average precision is 92.78%.

Functional Comparison

Based on the above experimental results, we have made a functional comparison with each other among different approaches, including ours, Wu's (Wu and Cao 2018), and Bilbao's (S. Bilbao et al. 2020). The results of functional comparison are listed in Table 5. In summary, our proposed approach is more feasible and powerful than others.

“√” denotes “Yes” or “Available.”

Conclusion

In this paper, those datasets including karaoke lyrics, songs, titles, and author names downloaded from the Internet have been used, and a database system of Chinese pop songs for automatic calibration has been built. The grouping preference rules of GTTM are adopted to perform the parameter analysis as a basis for

determination. The appropriate decisions are made based on the above parameters by applying the reasoning algorithm of HLFPN model. A C# program has been developed to implement this automatic calibration system. Users can use this system to sing a pop song from which the singer's voices are removed, and the lyrics can be calibrated automatically. The contributions of this paper are stated as follows:

- (1) A robust database system with Chinese pop songs has been successfully built, providing a useful environment for musical analysis.
- (2) The grouping preference rules of GTTM are easily used to group each song.
- (3) The parameters obtained from the register change rule, the dynamics change rule, and the length change rule are all used to easily conduct technical analysis for musical decision.
- (4) Through the fuzzy reasoning algorithm of HLFPN model, a fast decision can be made on music performance or segmentation.
- (5) This system can provide users with an easy way to sing a pop song and to get the lyrics automatically.

In the future, the learning algorithms of the HLFPN model will be adopted to accurately detect music segmentations. Also, a specific tool of HLFPN model will be created to predict the music segmentation and to make a difference between HLFPN model and GTTM. Since this system is now designed for Chinese karaoke songs only, it has been limited to some users. Therefore, it will be enhanced for English songs or songs in other languages to provide more applications.

Acknowledgments

The authors are grateful to the anonymous reviewers for their constructive comments, which have improved the quality of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Ministry of Science and Technology, Taiwan MOST 107-2221-E-845-001-MY3, ROC, under grants MOST 110-2637-E-131-005- and MOST 110-2221-E-845-002-.

References

Baggi, D. L. 2015. An IEEE standard for symbolic music. *IEEE Computer*, 100–02.

- Bastanfard, A., D. Amirkhani, and S. Naderi. 2020. A signing voice separation method from Persian music based on pitch detection methods, *Procs. of 2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, Mashhad, Iran, 1–6.
- Bilbao, S., C. Desvages, M. Ducceschi, B. Hamilton, R. Harrison-Harsley, A. Torin, and C. Webb. 2020. Physical modeling, algorithms, and sound synthesis: The NESS Project. *Computer Music Journal* 43 (2–3):15–30. doi:10.1162/comj_a_00516.
- Casey, M. A., R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. 2018. Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE* more author. 96 (4):668–96. doi:10.1109/JPROC.2008.916370.
- Chan, A. B., and J. H. Hsiao. 2016. Information distribution within musical segments. *Music Perception: An Interdisciplinary Journal* 34 (2):218–42. doi:10.1525/mp.2016.34.2.218.
- Chen, G.-F. 2021. Music sheet score recognition of Chinese Gong-che notation based on deep learning. *Proceedings of 2021 International Conference on Big Data Analysis and Computer Science (BDACS)*, Kunming, China, 1–6.
- Chiang, D.-L., S.-K. Wang, Y.-N. Lin, C.-Y. Yang, V. R. L. Shen, T. T.-Y. Juang, and T.-Y. Liao. 2021. Development and evaluation of a novel investment decision system in cryptocurrency market. *Applied Artificial Intelligence* 35 (14):1169–95. doi:10.1080/08839514.2021.1975380.
- Chiang, D.-L., S.-K. Wang, Y.-Y. Wang, Y.-N. Lin, T.-Y. Hsieh, C.-Y. Yang, V. R. L. Shen, and H.-W. Ho. 2020. Modeling and analysis of Hadoop MapReduce systems for big data using Petri nets. *Applied Artificial Intelligence* 35 (1):80–104. doi:10.1080/08839514.2020.1842111.
- Chowdhury, J. I., D. Thornhill, P. Soulatiantork, Y. Hu, N. Balta-Ozkam, L. Varga, and B. K. Nguyen. 2019. Control of supercritical organic rankine cycle based waste heat recovery system using conventional and fuzzy self-tuned PID controllers. *International Journal of Control, Automation and Systems* 17 (4):2969–81. doi:10.1007/s12555-018-0766-6.
- Frankland, B. W., and A. J. Cohen. 2018. Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff's a generative theory of tonal music. *Music Perception: An Interdisciplinary Journal* 21 (4):499–543. doi:10.1525/mp.2004.21.4.499.
- Goto, M. 2016. A chorus-section detecting method for musical audio signal and its application to a music listening station. *IEEE Transactions on Audio, Speech, and Language Processing* 14 (5):1783–94. doi:10.1109/TSA.2005.863204.
- Haus, G., and M. Longari. 2019. Time-based music description approach based on XML. *Computer Music Journal* 29 (1):70–85. doi:10.1162/comj.2005.29.1.70.
- Haus, G., and L. A. Ludovico. 2017. Music segmentation: An XML-oriented approach. *Lecture Notes in Computer Science* 33 (10–1):330–46.
- Hu, D., Z. Chen, and F. Yin. 2020. Passive geometry calibration for microphone arrays based on distributed damped Newton optimization. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (11):118–31. doi:10.1109/TASLP.2020.3037532.
- Lin, Y.-N., T.-Y. Hsieh, C.-Y. Yang, V. R. L. Shen, T. T.-Y. Juang, and W.-H. Chen. 2020a. Deep Petri nets of unsupervised and supervised learning. *Measurement and Control Online*. 53 (7–8):1–11. doi:10.1177/0020294020923375.
- Lin, Y.-N., S.-K. Wang, G.-J. Chiou, C.-Y. Yang, V. R. L. Shen, T. T.-Y. Juang, and T.-J. Huang. 2022. Novel deadlock control for smartphone manufacturing systems using Petri nets. *International Journal of Control, Automation and Systems* 20 (3):877–87. doi:10.1007/s12555-020-0239-6.
- Lin, Y.-N., S.-K. Wang, C.-Y. Yang, V. R. L. Shen, T. T.-Y. Juang, and C.-S. Wei. 2020b. Novel JavaScript malware detection based on fuzzy Petri nets. *Journal of Intelligent & Fuzzy Systems Online*:1–26.
- Liu, -C.-C. 2010. Automatic phrase segmentation of MP3 songs based on the technique of breath sound detection. *Proceedings of IEEE 7th International Conference on Information Technology and Applications*, Taipei, Taiwan, 285–90.

- Matsumoto, M., and J. Hori. 2019. Classification of silent speech using support vector machine and relevance vector machine. *Applied Soft Computing* 20 (1):95–102. doi:10.1016/j.asoc.2013.10.023.
- Meier, W. 2013. eXist: An open source native XML database. *Lecture Notes in Computer Science* 2593 (1):169–83.
- Ruangsang, S., and W. Assawinchaichote. 2019. Control of nonlinear Markovian jump system with time varying delay via robust H fuzzy state feedback plus state-derivative feedback controller. *International Journal of Control, Automation and Systems* 17 (3):2414–29. doi:10.1007/s12555-019-0044-2.
- Shen, V. R. L., and H.-C. Chen. 2014. An automatic calibration system for Chinese karaoke lyrics based on high-level fuzzy Petri nets. *Procs. of 2014 International Conference on Machine Learning and Cybernetics (ICMLC)*, Lanzhou, China, 1–6.
- Shen, V. R. L., R.-K. Shen, C.-Y. Yang, and W.-C. Chen. 2017. A novel fall prediction system on smartphones. *IEEE Sensors Journal* 17 (6):1865–71. doi:10.1109/JSEN.2016.2598524.
- Sung, W.-T., and K.-Y. Chang. 2019. Health parameter monitoring via a novel wireless system. *Applied Soft Computing* 22 (1):667–80. doi:10.1016/j.asoc.2014.04.036.
- Tsai, W. H., and H.-C. Lee. 2017. Automatic evaluation of karaoke singing based on pitch, volume, and rhythm features. *IEEE Transactions on Audio, Speech, and Language Processing* 20 (4):1233–43. doi:10.1109/TASL.2011.2174224.
- Wick, C., A. Hartelt, and F. Puppe. 2020. Lyrics recognition and syllable assignment of medieval music manuscripts, *Procs. of 2020 17th. International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Dortmund, Germany, 1–6.
- Wu, X., and Y. Cao. 2018. Research on song sentiment binary classification based on Chinese lyrics. *Proceedings of 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, Singapore, 22–29.
- Yang, S.-H. Y.-N. L., G.-J. Chiou, M.-K. Chen, V. R. L. Shen, H.-Y. Tseng, and H.-Y. Tseng. 2019. Novel shot boundary detection in news streams based on fuzzy Petri nets. *Applied Artificial Intelligence* 33 (12):1035–57. doi:10.1080/08839514.2019.1661118.
- Zhang, Y., H. Zhao, and C. He. 2022. Robust control design with optimization for uncertain mechanical systems: Fuzzy set theory and cooperative game theory. *International Journal of Control, Automation and Systems* 20 (2):1377–92. doi:10.1007/s12555-020-0874-y.