2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics

TuneRank Model for Main Melody Extraction from Multi-Part Musical Scores

Hanqing Zhao[1,2] and Zengchang Qin[1]
[1]*Intelligent Computing and Machine Learning Lab, School of ASEE, Beihang University, China*
[2]*École d'Ingénieur Généraliste, École Centrale de Pékin, Beihang University, Beijing, China*
Emails: zhq@gmx.com, zcqin@buaa.edu.cn

***Abstract*—A musical part is a strand of music played by an individual instrument within a larger music work, main melody is a single musical part which consists of the most significant melodic elements of multi-part musical score. This part is typically related to dissonance between musical parts and musical instruments. In this paper, we propose a novel model for extracting the main melody from multi-part musical scores. This model is referred to as the TuneRank model that has the conceptually similar idea of the PageRank model. If each musical note can be considered like a web page in the Internet, and the dissonance value between two notes is like the quantity of links between two web pages. The TuneRank (rank of becoming main melody) of each note is calculated using Markov transition probability. This model is tested on the ECPK4 database. By comparing to the previous work, we find that this note-based model is more effective for processing scores containing main melody in multiple parts. Also, the accuracy does not change with the increase of the number of parts. In general, this model can be used for extracting the single-part main melody of digital musical scores.**

***Keywords*-TuneRank; PageRank; Melody extraction; Musical score.**

I. Introduction

A musical part is a strand or melody of music played by an individual instrument (or group of identical instruments) within a larger music work. Main melody is a single part of music which consists of the most melodic elements of multi-part music. With the development of information technology, the digital musical score formats (like MusicXML, MSCZ and etc.) are becoming more common for composing music and sharing music on the Internet. However, it is always difficult for users to search musical score files. One of the most efficient approaches is Query by Humming (QbH) [1]. However, The QbH algorithm can only match one input humming audio track with a single part of music melody, while the digital musical score files usually contain multiple parts. In order to use the QbH method to search digital musical score files, we need to extract the main melody part from multi-part digital musical score file in order to build the melodic library.

As far as we concerned, computing-based main melody extraction from multi-part musical files has not been reported in any literatures. But in a similar field, the study of extracting main melody from MIDI files are available in literatures for years [2–5]. Ozcan *et al.* [2] eliminated MIDI channels those do not contain melodic information depending on pitch histogram. Shih *et al.* [3] proposes a modified Lempel-Ziv model to extract MIDI main melody by using a dictionary based approach for extracting repetitive patterns in music. Zhao and Wu [4] use the melodic feature of each parts (including left and right channel balance value, average loudness, rest time and etc.) to extract main melody. Zhang [5] uses overtone significant degree detecting method to extract main melody of .WAV files and transform them into a MIDI library.

However, previous research are mainly based on probability and statistics, but have not consider much from music theory criterions, like consonance and dissonance between different intervals. Another problem is that the existing research often assumes that there exists only one part main melody. But in digital musical score files, the main melody often exists in different parts of the score.

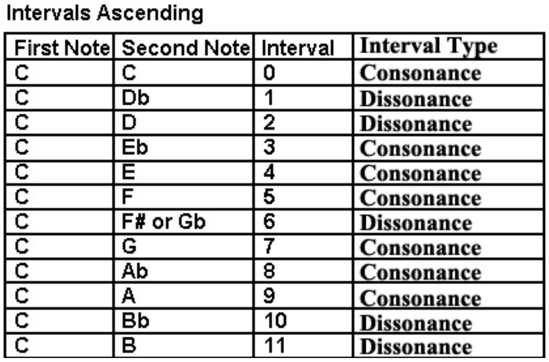
In this paper, we proposed a model for main melody extraction from digital musical score files. With a similar idea of PageRank model in Internet search, each music note is regarded as a web page in the Internet, and extract main melody by calculating TuneRank (rank of becoming main melody) of each note. We deal with the MusicXML format, which is well used in music composing and musical score storing. The remaining paper is consisted by the following sections. Section II introduces the basic knowledge of music theory. Section III introduces the TuneRank model. And Section IV gives test results of TuneRank model under ECPK4 database. Conclusions are given in Section V.

II. Fundamental Music Knowledge

In music theory, the Twelve-Tone Equal Temperament (12-TET)[12] divides an octave into 12 different notes, there can be up to 11 intervals (semitones) between any two notes. Comparing to the consonance, the dissonance is an interval between two notes which sound harsh or unpleasant to most people. In music theory, 11 intervals between each two different notes can be divided into 6 consonance and 5 dissonance matching between notes. Consonance/dissonance matching to the note C are shown in Fig. 1.

In general, consonance are usually used between accompaniment melodies, and the main melody usually changes more frequently than accompaniment melodies. As a result,

978-1-4799-4955-7/14 $31.00 © 2014 IEEE
DOI 10.1109/IHMSC.2014.145



Intervals Ascending

First Note	Second Note	Interval	Interval Type
C	C	0	Consonance
C	Db	1	Dissonance
C	D	2	Dissonance
C	Eb	3	Consonance
C	E	4	Consonance
C	F	5	Consonance
C	F# or Gb	6	Dissonance
C	G	7	Consonance
C	Ab	8	Consonance
C	A	9	Consonance
C	Bb	10	Dissonance
C	B	11	Dissonance

Figure 1. Types of intervals in Twelve-Tone Equal Temperament

the dissonances appears more frequently between main melody and accompaniment melody than between accompaniment melodies themselves. An example is given in Fig. 2, where the first part (trumpet) is the main melody and other three parts (flute, saxophone and trombone) are accompaniment melodies.

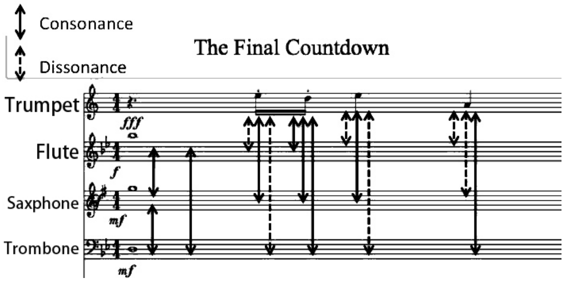


Figure 2. An example of consonances and dissonances in multi-part musical score

A. Dissonance Values

In digital musical score files, we can acquire exact pitch values of each notes. In the Twelve-Tone Equal Temperament (12-TET)[1] system, there exists 11 intervals between any two notes. As a result, in order to identify the dissonance degree between each two different notes, we need to acquire dissonance values. Chen and Lu [6] proposed a model of measuring dissonance values in each interval. We use I to represent the dissonance value of each interval in Twelve-Tone Equal Temperament. The normalized I as shown in Table I.

Table I
Dissonance values in different intervals

Interval (in semitones)	0	1	2	3	4	5
Dissonsnce Value	0.1	0.9	0.67	0.65	0.58	0.43
Interval (in semitones)	**6**	**7**	**8**	**9**	**10**	**11**
Dissonsnce Value	0.65	0.34	0.62	0.56	0.69	0.81

B. Weights of Instruments

The musical instrument weight represents the saliency of a type of instrument in the orchestra. To determine this value, we considered the Average Measured Equivalent Continuous Sound Level (L_{eq}) of each instrument, Moore [7] measured L_{eq} of different instruments. By normalizing L_{eq}, the musical instrument weights (W) are shown in Table II based on [7].

Table II
Instruments weights (normalized L_{eq}) of each instrument

Instrument	Violin	Viola	Cello	Bass	Flute
Instrument weight (W)	0.507	0.700	0.497	0.06	0.363
Instrument	**Oboe**	**Clarinet**	**Bassoon**	**French Horn**	**Trumpet**
Instrument weight (W)	0.543	0.737	0.577	0.867	0.883

III. TuneRank Model

A. Three Criterions of Main Melody

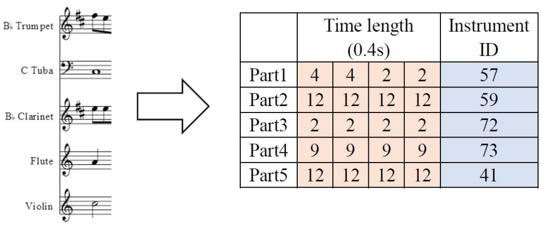
Based on the fundamental music knowledge which has been introduced in section II, we proposed three basic criterions to define the main melody.

1) Dissonance value (Longitudinal TuneRank): If a note has a larger sum of dissonance values than other notes at the same time point (column), which means this note sounds more dominant than others, so this note should have a high probability of becoming the main melody.
2) Distance between important notes (Horizontal TuneRank): A tune is an ordered set of notes, a note has a higher probability of becoming the main melody, if it is close to notes with larger probabilities of being the main melody.
3) Musical instrument weight (W): If a part is performed by a more important instrument, all the notes in this part will have higher probabilities to become the main melody.

B. Mathematical Representation of Musical Scores

A music score can be transformed into a description matrix S (E.g., Fig.3). In this matrix, elements in a row represent parts of the score, and the numbers of columns are the length of time. Each column contains the notes at a certain time point, the sampling interval between each two columns is 0.1 second.

[1] An equal temperament is a musical temperament, or a system of tuning, in which every pair of adjacent notes has an identical frequency ratio. As pitch is perceived roughly as the logarithm of frequency, this means that the perceived "distance" from every note to its nearest neighbor is the same for every note in the system [12].



	Time length (0.4s)				Instrument ID
Part1	4	4	2	2	57
Part2	12	12	12	12	59
Part3	2	2	2	2	72
Part4	9	9	9	9	73
Part5	12	12	12	12	41

Figure 3. A 5-parts score can be represented by a description matrix S

In matrix S, according to the Twelve-Tone Equal Temperament, each note is represented by an integer between 1 and 12. And the last column of the matrix contains the instrument ID.

In order to acquire dissonance value between each two notes, we firstly need to construct a matrix E including dissonance values in Table. I. E.g.:

$$\begin{bmatrix} 0.1 & 0.9 & 0.67 & 0.65 & \cdots & 0.62 & 0.56 & 0.69 & 0.81 \end{bmatrix} \quad (1)$$

Assuming note $A = S(a, j)$ and note $N = S(n, j)$ in a musical score S, the longitudinal dissonance value $D(a, n)$ between these two notes can be calculated by the following:

$$D(a, n) = E(1, |a - n| + 1) \quad (2)$$

C. Generating Transition Probability Matrix and Calculating TuneRank

For a note $A = S(a, j)$ in an M-parts musical score, the sum of longitudinal dissonance of note A, $D_L(A)$ depends on impacts from the other $M - 1$ notes at the same time (i.e., the column j):

$$D_L(A) = \sum_{n=1}^{a-1}(D(a,n) + W(N) \times \alpha) + \sum_{n=a+1}^{M}(D(a,n) + W(N) \times \alpha) \quad (3)$$

While $D(a, n)$ is the dissonance value between note $A = S(a, j)$ and note $N = S(n, j)$ calculated by Equation 2, and $W(N)$ is the instrument weight of the part which contains note N, and the parameter α controls the influence of $W(N)$ on the result. The sum of longitudinal dissonance $D_L(A)$ is the sum of dissonance values between note A and the other $M - 1$ notes (notes in same time point j except note A itself).

After calculating the sum of longitudinal dissonance of each notes, we start the process of transition probabilities. In this process, note A in an M-parts musical score transfers its probability of becoming main melody of the rest $M - 1$ notes in the same time point. The longitudinal transition probability $P_L(a, n)$ between note A and N is based on dissonance values calculated by each note.

$$P_L(a, n) = \frac{D(a, n) + W(N) \times \alpha}{D_L(A)} \quad (4)$$

Where $P_L(a, n)$ is the transition probability from note $A = S(a, j)$ to note $N = S(n, j)$, $D(a, n)$ refers to the dissonance value between note A and N, and $D_L(A)$ represents the longitudinal dissonance value of note A. $W(N)$ is the instrument weight of the part which contains the note N. For all the notes in the first column ($j = 1$) in an M-parts musical score, eventually we can build up a longitudinal transition probability matrix B, where $B(a, n) = P_L(a, n)$ as shown below.

$$B = \begin{bmatrix} 0 & P_L(1,2) & \cdots & P_L(1,M) \\ P_L(2,1) & 0 & \ddots & \vdots \\ \vdots & \ddots & 0 & P_L(M-1,M) \\ P_L(M,1) & \cdots & P_L(M,M-1) & 0 \end{bmatrix} \quad (5)$$

We can calculate the longitudinal TuneRank value of each note by using a Markov chain transition probability process.

$$X^t = BX^{t-1} = X^{t+1}e \quad (6)$$

While e is a column vector with M entries of 1. According to the Perron-Frobenius theorem, when B is an irreducible stochastic matrix, the largest eigenvector v with positive entries will determine longitudinal TuneRank value of each note.

$$Bv = v \quad (7)$$

When t gets large enough, X^t will approach v. In actual operation, we use the vector norm of the differences between X^t and X^{t-1} to determine convergence.

$$\|X^t - X^{t-1}\| \leq 0.001 \quad (8)$$

While each entries in vector X^t is the longitudinal TuneRank (R_L) of each notes in certain time point (column). The R_L of note $A = S(a, j)$ is as shown below.

$$R_L(A) = X^t(a, 1) \quad (9)$$

After calculating longitudinal TuneRank (R_L), each note affects its vicinity notes in the same part (row i) based on its R_L. The horizontal impact $E_H(a, b)$ from note $A = S(i, a)$ to note $B = S(i, b)$ is decreasing by the distance (L) between note A and B:

$$\begin{gathered} L = |a - b| \\ E_H(a, b) = \beta^L R_L(A) \end{gathered} \quad (10)$$

While β is the horizontal attenuation coefficient. We also introduce γ by controlling the quantity of vicinity notes for each note could have an effect on.

Once β and γ are defined, we can build up a horizontal-impact matrix C where $C(b, a) = E_H(a, b)$.For all the notes in the part of matrix S, the horizontal-impact matrix of notes in each row of matrix S can be shown below. In this model, each note has influence on 2γ vicinity notes, including γ notes forward and γ notes backward.

In order to avoid the first γ notes which cannot receive sufficient amount of influence, we firstly add γ time points of pauses (columns with all entries of 0) to the leftmost of matrix S. The dimension (L_C) of matrix C equals to the length (L_S) of matrix S plus $\gamma : L_C = L_S + \gamma$.

$$C = \begin{bmatrix} \cdots & 0 & \cdots \\ \cdots & \vdots & \cdots \\ \cdots & 0 & \cdots \\ \cdots & E_H(S(i, n-\gamma), S(i,n)) & \cdots \\ \cdots & \vdots & \cdots \\ \cdots & E_H(S(i, n-1), S(i,n)) & \cdots \\ \cdots & 0 & \cdots \\ \cdots & E_H(S(i, n+1), S(i,n)) & \cdots \\ \cdots & \vdots & \cdots \\ \cdots & E_H(S(i, n+\gamma), S(i,n)) & \cdots \\ \cdots & 0 & \cdots \\ \cdots & \vdots & \cdots \\ \cdots & 0 & \cdots \end{bmatrix} \qquad (11)$$

Then we firstly normalize the sum of each rows of matrix C into 1. And then normalize the sum of each column of matrix C into 1.

After normalization we can get a horizontal transition probability L_C-dimensions matrix C_1. Where the transition probability from a fundamental note to a target note is not only related to the absolute size of longitudinal TuneRank values (R_L) of target note, but also related to the relative R_L size between the fundamental note and other fundamental notes which have transition probabilities on the same target note.

According to the Perron-Frobenius theorem,the largest eigenvector v of matrix C_1 with positive entries will determine the horizontal TuneRank of the tunes.

Just as similar as the transition probability process in Equation.6 ,in this stage, we use the longitudinal TuneRank X^t vector calculated in (6) as each note's initial probability of becoming the main melody. The largest eigenvector X^k which contains the horizontal TuneRank (R_H) of each notes.

$$X^k = C_1 X^{k-1} = {C_1}^{k+1} X^t \qquad (12)$$

When each entries in vector X_k becomes the horizontal TuneRank (R_H) of each notes in certain part (row), the R_H of note $A = S(i, a)$ is as shown below.

$$R_H(A) = X^k(a, 1) \qquad (13)$$

We calculate the final TuneRank (R_F) of each note as a result by multiplying longitudinal TuneRank (R_L) and horizontal TuneRank (R_H).

$$R_F = R_L \times R_H \qquad (14)$$

Finally, We consider the note which has a higher TuneRank than the others in the same time point (column) as the main melody.

In this model, we assumed three parameters: α, β, and γ. α controls the influence conducting by musical instrument weights $W(N)$ towards the longitudinal TuneRank. β is the horizontal attenuation coefficient, which identifies the intension exerted on horizontal vicinity notes. γ refers to the maximum horizontal affect length determining the quality of horizontal vicinity notes that a note could be influenced on.

IV. Experimental Studies

To verify the performance of the new proposed model, we firstly build up a test set named ECPK4[2] (the École Centrale de Pékin, ECPK musical score data set with 4 types), consists of 40 musical score files in 4 different quantities of parts 5, 6, 7 and 8 where each of them contains 10 score files. All the musical scores in this test set are in MusicXML format having the same length of 9.9 seconds i.e. 99 time points.

In order to facilitate the statistical accuracy, we build the ECPK4 database according to the following 2 standards: (1) Each score has, and only has one clear main melody part at each time point. (2) There is no part in musical score that pauses from the beginning to the end.

The data set ECPK4 contains 3960 time points i.e. 40 scores with 99 time points, in terms of 4 different types: 5-parts, 6-parts, 7-parts and 8-part of each contains 990 time points. We calculate the accuracy (the quantity of time points whose main melody has been extracted correctly divided by the sum of time points) to evaluate the performance of this model. In order to study the influence of the three parameters. We tested accuracies in different numerical value of parameter α, β and γ the results are shown in tables III to V:

Table III
Accuracy in different sizes of parameter α

Parameter α	10%	20%	30%	40 %	50%
5 parts	39.19%	47.88%	52.32%	55.25%	57.88%
6 parts	53.73%	56.87%	56.46%	56.97%	**57.47%**
7 parts	**63.83%**	61.61%	60.71%	60.10%	58.89%
8 parts	50.20%	60.71%	65.76%	**67.98%**	66.87%
Average accuracy	51.74%	56.77%	58.81%	60.08%	**60.28%**
Parameter α	60%	70%	80%	90%	100%
5 parts	59.19%	61.41%	63.54%	64.44%	**64.95%**
6 parts	57.37%	55.65%	55.86%	56.36%	56.36%
7 parts	56.97%	54.75%	53.94%	52.63%	51.52%
8 parts	66.06%	64.65%	63.64%	62.42%	61.41%
Average accuracy	59.90%	59.12%	59.24%	58.96%	58.56%

[2] http://icmll.buaa.edu.cn/members/Hanqing Zhao/

Table IV
ACCURACY IN DIFFERENT SIZES OF HORIZONTAL AFFECT WEIGHT β

Parameter β	10%	20%	30%	40 %	50%
5 parts	**58.99%**	57.88%	56.67%	57.17%	56.46%
6 parts	**58.59%**	57.78%	54.55%	55.56%	55.56%
7 parts	53.94%	53.54%	53.23%	54.24%	54.95%
8 parts	57.47%	60.00%	61.62%	61.62%	63.84%
Average accuracy	57.25%	57.30%	56.52%	57.15 %	57.71%

Parameter β	60%	70%	80%	90%	100%
5 parts	55.66%	55.76%	57.88%	57.58%	55.86%
6 parts	55.66%	56.87%	57.47%	57.58%	57.58%
7 parts	56.67%	58.28%	**58.89%**	58.18%	57.47%
8 parts	65.35%	66.46%	**66.87%**	65.96%	65.45%
Average accuracy	58.33%	59.34%	**60.28%**	59.82%	59.09%

Table V
ACCURACY IN DIFFERENT SIZES OF HORIZONTAL AFFECT LENGTH γ

Parameter γ	2	3	4	5	6
5 parts	56.67%	56.87%	56.26%	56.06%	56.46%
6 parts	56.97%	56.26%	55.96%	57.17%	57.47%
7 parts	54.65%	55.15%	56.57%	57.88%	58.48%
8 parts	61.52%	62.42%	64.14%	64.34%	65.45%
Average accuracy	57.45%	57.68%	58.23%	58.86%	59.47%

Parameter γ	7	8	9	10	11
5 parts	57.47%	57.47%	57.89%	58.59%	**58.99%**
6 parts	**57.88%**	57.78%	57.47%	56.87%	56.87%
7 parts	58.79%	**59.09%**	58.89%	58.48%	57.98%
8 parts	66.57%	66.77%	66.87%	66.77%	**67.17%**
Average accuracy	60.18%	60.278%	**60.28%**	60.18%	60.25%

By comparing the above results, we can conclude that when three parameters settled as $\alpha = 0.5$, $\beta = 0.9$, $\gamma = 9$ the model reaches its highest accuracy of 60.28% in main melody extraction. Under this condition, the melody in scores with 5 parts can be extracted at the rate 57.89%; the melody in scores with 6 parts can be extracted at the rate 57.47%; the melody in scores with 7 parts can be extracted at the rate 58.89%; and the melody in scores with 8 parts can be extracted at the rate 66.87%.

V. CONCLUSION

In this paper, we have presented a simple but novel model for main melody extraction from multi-part musical scores. Each piece of music can be represented into score description matrix. We use the dissonance value, distance between important notes, and music instrument weights as criterions. Following the idea of PageRank, we use transition probability to calculate the TuneRank of each note. We consider the note with a higher TuneRank than others as the main melody at a time. By using such a note-based function, this model can deal with scores that a main melody exists in different parts of the score in different time. The performance of the model has been evaluated on the dataset ECPK4. The experimental results show that the accuracy is not decreasing with the increasing of number of parts. The general performance is comparable or even better than some classical methods [8].

ACKNOWLEDGEMENTS

Zengchang Qin is supported by the National Science Foundation of China No. 61305047.

REFERENCES

[1] Ghias A., Logan J., Chamberlin D., ”Query by Humming: Musical Information Retrieval in an Audio Database” *Proceedings of the third ACM international conference on Multimedia*, 1995.

[2] Ozcan G., Isikhan C., Alpkocak A.. “Melody Extraction on MIDI Music Files” *Seventh IEEE International Symposium on Multimedia*, 2005.

[3] Shih H. H., Narayanan S. S., Kuo C. C. J. “Automatic Main Melody Extraction From MIDI Files With A Modified Lempel-Ziv Algorithm” *IEEE Intelligent Multimedia, Video and Speech Processing*, pp.9-12, 2001.

[4] Zhao F. and Wu Y.“Melody Extraction Method from Polyphonic MIDI Based on Melodic Features,” *Computer Engineering*, 2007.

[5] Zhang J. “Harmonic Overture Detection Based Main Melody Extraction” *Master thesis of Shanghai Jiaotong University*, 2007.

[6] Chen Q. and Lu Z. “Calculating Consonance” *Journal of the Central Conservatory of Music*,vol. 4, 1994.

[7] Moore A. “Student Musicians’ Perception of Loudness and How it Correlates to the Measured Level,” *Washington University School of Medicine*, 2010.

[8] Durrieu J. L., Richard G., Blei D., Fevotte C. “Source/Filter Model for Unsupervised Main Melody Extraction from Polyphonic Audio Signals” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 3, 2010.

[9] Fuentes B., Liutkus A., Badeau R., Richard G. “Probabilistic Model for Main Melody Extraction Using Constant-Q transform” *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.5357-5360, 2012.

[10] Cook N. D., Hayashi T. “The Psychoacoustics of Harmony Perception,” *American Scientist*, vol. 4, 2008.

[11] Zhao H. “Music Performance Assistant Syatem,” *Patent Gazette WO/2013/139022*, The International Bureau of WIPO, 2013.

[12] *http://en.wikipedia.org/wiki/Equal_temperament*,Last modified on 10 April 2014 at 19:14.