

Similar Singer Search by Property of Vocal Sound through Weighted Euclidean Distance

Ning-Han Liu

*Department of Management Information Systems
National Pingtung University of Science & Technology
Taiwan, R.O.C.*

gregliu@mail.npust.edu.tw

Abstract

Although using the singer's name to search for songs is the common function of music retrieval systems, people often only remember famous singers. For new singers, people often only remember that the singer's voice is similar to a famous singer, but cannot recall his/her name. Among current studies, there is little discussion on the music search method by voice. Therefore, this study proposed a method of music search by similar vocal sound. Users can enter the name of a singer, and then the system can find the names of other singers with similar features in vocal sound. In the proposed method, the sound signal is processed by Discrete Fourier Transform, and the statistical value of frequency spectrum and formant is calculated as the characteristic value of the singer's vocal sound. Then, the similarity of the singer's vocal sound is calculated by the weighted squared Euclidean distance function to obtain similar singers' names, so as to search for music by using similar vocal sound. The experiment verified that the feature extraction method can realize the vocal sound-based search effectively.

Keywords: *Discrete Fourier Transform, Formant, Vocal Similarity, Music Features*

1. Introduction

When listening to a song, the audiences seldom notice the singer's name or the name of the song, while have stronger impression on the vocal sound or melody. When the audiences want to listen to the song again, they often cannot recall the relevant information of the song. If the audiences can remember the vocal sound and melody, they can hum the memorized melody for search. If the audiences cannot even remember any part of the melody or are unable to hum precisely for the search, it is impossible to search by humming search. They may only associate the singers having similar vocal sound to their preferred singer. However, there is no such search mode. Moreover, there is no study on music classification or music clustering aiming at similar vocal sound of singers. Therefore, this study proposed a new type of search mechanism for users to find their wanted music more conveniently. Currently, there are many studies on singer identification. Although the identification rate is high, there are still errors [1]. Many of the misidentified songs are sung by the singers with similar vocal sound characteristics. These misidentified singers are meaningful to this study for developing a new search mode.

The vocal sound of a singer is a result of personal vocal habit. Each singer has his/her own intonation, *i.e.*, vocal sound while singing songs, the singer's vocal sound is usually described

as grave, deep, loud, clear, *etc.*. The vocal sound defined in this study refers to the vocal sound of singers contains personal vocal particularity.

This study analyzed the singer's vocal sound by features that could identify the voice, such as timbre. Timbre is the feature that could be used to identify participants of a conversation, as well as the singing of singers. The time domain and frequency domain are commonly used for sound feature extraction. The advantage of time domain is that the sound feature can be calculated disregarding the variance in sound signal. The defect is that it is not suitable for calculating very complicated sound, such as pop music containing vocal sound. The advantage of frequency domain is that complicated sound can be calculated, and reasonable feature can be found. As the characteristic value of sound cannot be calculated until the time domain is converted into frequency domain, the computing time is longer. The formant difference in frequency spectrum often presents the difference in speaking, thus, the formant of vocal sound is used as the primary characteristic value of music.

The remainder of this paper is organized as follows: Section 2 introduces voice and vocal sound, discusses singer recognition and music classification, and compares related studies of similarity; Section 3 presents the method of searching for singers with similar vocal sound, and implements the proposed method; Section 4 discusses the experimental analysis; Section 5 gives conclusions and suggestions to future studies.

2. Related Works

Human voice is usually evaluated by formants, which is the frequency of the range with the densest energy. The formants are sequenced from low frequency to high frequency. F_1 , F_2 and F_3 are low frequency range. Generally, F_1 and F_2 can represent the vowels of voice. After pronunciation training, the voice will have a speaker's formant approximate to F_3 , the frequency range is 3-4 kHz [2], but the speaker's formant is not obvious. Professionals (e.g. vocal music teachers) have obvious speaker's formant. In terms of male and female, the formant frequency of female is higher than that of male, because the vocal tract of female is shorter, sounding higher pitch, and the male can sound higher volume. Similarity of voice in daily life is due to similar frequencies, sounding movements, and habits, so as to identify the speaker as one of acquaintances having similar voice.

The formant is often used to analyze the vocal sound of singers, professional singers have a singer's formant at special frequency F_3 [3]. The Perceptual Linear Prediction (PLP) is used to obtain the singer's formant and to analyze the vocal sound [4]. The formant is also used to analyze the singer's timbre, generally, the formant is the multiple of fundamental frequency, and the fundamental frequency is the reciprocal of fundamental period. It is the pitch of voice. The formant can be used to analyze the quality of timbre. The slight gaps in the multiples among formants (frequencies F_1 , F_2 and F_3) are the particular vocal sound characteristics of singers, so that people can identify different vocal sounds of singers.

The singer recognition is also known as singer identification. Many studies have proposed different methods for singer recognition. Youngmoo and Brian [1] proposed automatic identification of singers, and used feature extraction methods of linear predictive coding (LPC) and Warped Linear Prediction (WLP). They also used frequency as characteristic value to compare the classification effects of the two characteristic values, and found that the accuracy rate of LPC was higher than that of WLP. Tsai and Wang [5] used MFCC (Mel-scale Frequency Cepstral Coefficients) to capture the characteristic values of vocal sound, and classified singers according to vocal sound.

In related studies of music classification, the music feature extraction method has been able to express music features effectively, i.e. studies extending from general music classification

to automatic music classification. The classification method emphasizes the accuracy rate, and many studies have proposed effective feature extraction methods to increase the accuracy rate of music classification. Some studies have proposed methods advantageous to identify vocal sound or music, such as MFCC and LPC feature extraction methods [6, 7, 8]. MFCC obtains the energy of sound and the high frequency part of sound frequency. LPC is the feature method for obtaining the sound energy all-pole. The two feature extraction methods usually involve the frequency range insensible to human ears. They are more suitable for singer identification. The classifier is also one of subjects being discussed by researchers, in order to obtain the best accuracy rate of classifier, effective characteristic values are tried continuously. Therefore, some classifiers can increase the accuracy rate of music classification, and multiple classifiers can remedy other classifiers' defects. The classifiers often used for music classification include back-propagation neural network, k Nearest Neighbor Algorithm, Gaussian Mixture Model, *etc.* The decision trees and Discriminant Analysis [9] are often used for classification.

Among the studies comparing similarity, the most representative one is Euclidean Distance [10], *i.e.*, the method for calculating distance often used in clustering algorithms (e.g. SOM[11], K-means[12]). The Euclidean Distance is usually used in image recognition, voice identification, music recommendation, document recommendation, medical research domains, *etc.*

3. Method for Searching for Singers with Similar Vocal Sound

This study analyzed the vocal sound of singers in pop songs, and used the timbre of vocal sound to find singers with similar timbre. Professional singers sing different timbres for different songs, meaning that the singing style and vocal sound vary with the songs. The proposed mechanism aimed to find singers with similar vocal sound from the songs of different singers.

This study selected pop songs for analysis, and captured vocal sound segments from the refrains of songs. It then used two feature extraction methods, Discrete Fourier Transform (DFT) and Formants, to obtain the timbre of vocal sound of singers. The characteristic values of vocal sound were captured from the frequency spectrum which showed the timbre change. The weighted squared Euclidean distance function was used to calculate the distance between characteristic values of vocal sound of different singers. The similarity of vocal sound of different singers was determined by the calculated distance.

3.1 Vocal sound capture from refrain

This study obtained almost pure vocal sound by capturing a refrain segment with low background music and loud vocal sound from a music file. There is a lot of vocal sound extraction tools can be found in Internet (*e.g.*, Audacity). We manually extracted the segment of vocal sound. Although there is still background music in the extracted frames, it is enough to represent the vocal's sound characteristics according to the below experiments. Figure 1 (left) shows the oscillogram of a complete song, and Figure 1 (right) shows the captured vocal sound segment from refrain.

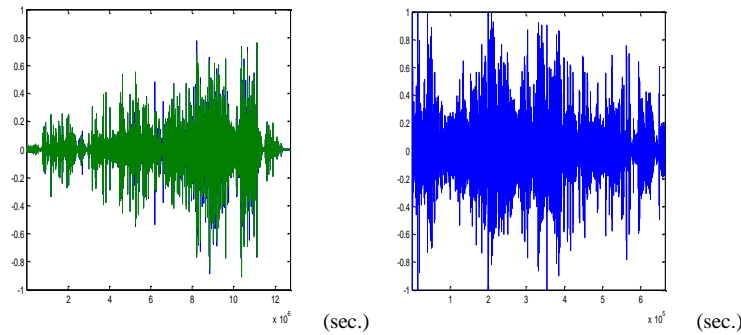


Figure 1. Captured entire song (left) and refrain vocal sound (right)

3.2 Vocal sound feature extraction

When the refrain was captured, the feature was extracted from the refrain. This study used the statistical value of frequency spectrum and formants of refrain as the feature of vocal sound. The sampling rate of the vocal sound from the audio files was set to 16 kHz.

3.2.1 Discrete Fourier Transform (DFT)

This study used Fourier analysis and the frequency of aural perception range to analyze the vocal sound change based on the timbre of frequency spectrum. The sound was cut into multiple segments before calculating DFT, and the audio signal values of vocal sound were preprocessed. The input vocal signal was segmented into frames of 20ms, with an optional overlap of 1/2 of the frame size. The procedure was multiplying pre-emphasis (set $\alpha=0.95$) and frame blocking by Hamming windows.

DFT was carried out after preprocessing, and the preprocessed audio signal value was substituted in DFT equation for calculation. The DFT equation is shown below:

$$F(c) = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi}{N} cn}, \quad 0 \leq c \leq N-1 \quad (1)$$

where, $F(c)$ is the DFT coefficient of the obtained plural, N is the frame size, $x(n)$ is the preprocessed audio signal value, j is the imaginary sign, and c is the frame index value.

When the plural of DFT was obtained, the frequency spectrum energy was calculated by using the Pythagorean theorem. The Pythagorean theorem equation is shown below.

$$r(n) = \sqrt{a_n^2 + b_n^2} \quad (2)$$

where, $r(n)$ is the n th energy value of DFT, a is the real coefficient of DFT, b is the imaginary coefficient of DFT.

As the sound we hear is sound change approximating to logarithm, the energy was multiplied by $20 \cdot \log_{10}$ to change into decibels (dB), the hearing range was represented by dB, and the music hearing range was about 20 dB to 100 dB. The frequency spectrum was obtained after DFT, as shown in Figure 2(a).

The bilateral frequency spectrum after DFT is symmetric, thus, this study used a half of frequency spectrum (see Figure 2(b)). The vocal sound frequency spectrum was used as the vocal sound characteristic. According to the sampling rate, each frequency spectrum had 512 values (dB); one half was 256 values (dB). This half frequency spectrum was equally divided into 4 intervals, and the maximum value and minimum value of each interval were obtained.

However, the refrain was separated into many frames; there were too many frequency spectra. Thus, the average value and standard deviation of maximum value and minimum value of each interval of all spectra were used. Therefore, each singer's vocal sound had 8 characteristic values in frequency spectrum.

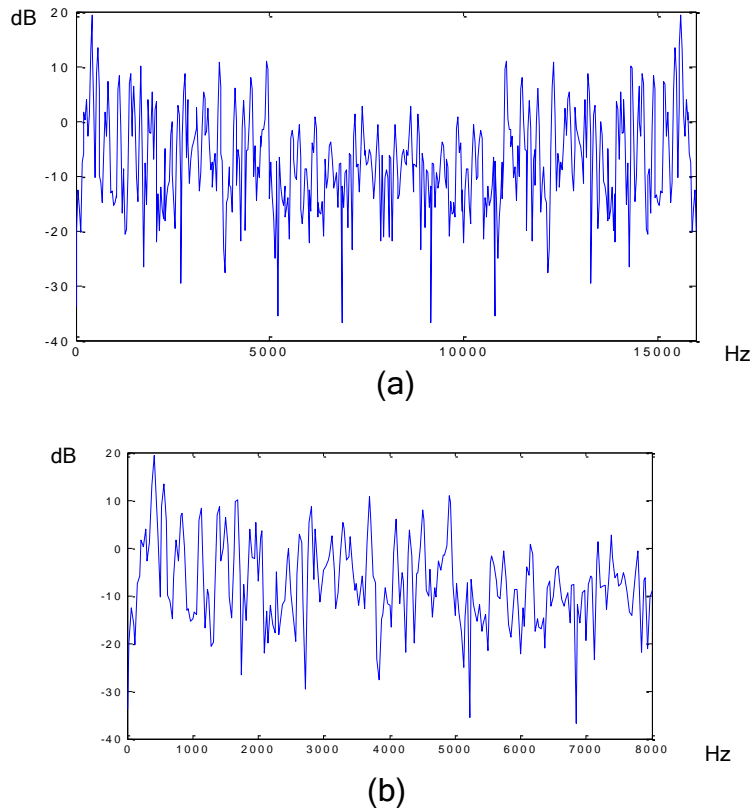


Figure 2. Frequency spectrum example of refrain frame

3.2.2 Formants

The formant is the regional frequency of sound energy concentration, based on which the low frequency feature of auditory sensitivity range can be obtained. The rough high frequency beyond aural perception was neglected. The preprocessing of formant was identical with that of DFT and the preprocessed audio signal value was the beginning of calculation. The formant frequency spectra were converted, and each frequency spectrum of formant had multiple peak values. Generally, multiple peak values were named $F_1, F_2, F_3, \dots, F_N$, there are N formants, each peak value represents a different frequency and energy, highlighting the feature of sound, especially low frequency formants. This study used formants to determine the timbre of vocal sound, and obtained the property of vocal sound from timbre. It then used the vocal sound frequency as the characteristic value of vocal sound.

This study used vocal sound energy to determine the envelope line of vocal sound, by the method of Linear Predictive Coding (LPC). The all-pole mode of vocal sound was defined before determining the envelope line. The equation is shown below:

$$H(z) = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (3)$$

where, the $H(z)$ is the vocal sound frequency component after DFT, G is the constant 1, P is the prediction order number, a_k is the linear prediction coefficient.

In order to obtain the linear prediction coefficient, the Inverse Filter is used for calculation, the equation is shown below.

$$x(n) = \sum_{k=1}^p a_k x(n-k) \quad (4)$$

where, $x(n)$ is the preprocessed audio signal value, the same as preprocessing of DFT, p is the prediction order number, a_k is the linear prediction coefficient.

The autocorrelation coefficient was determined before calculating the linear prediction coefficient. The autocorrelation coefficient used fewer data points to obtain the change in the whole frame signal. The calculation procedure was that the number of frame points moves rightwards k times, multiplied by original frame and averaged, so as to obtain k coefficients. The equation is shown below.

$$r(k) = \frac{1}{N} \sum_{n=0}^N x(n)x(n+k), \quad k = 0, 1, 2, \dots, p \quad (5)$$

where, $r(k)$ is the autocorrelation coefficient, $x(n)$ is the preprocessed audio signal value, N is the frame size, k is the number of autocorrelation coefficients, p is the prediction order number.

The autocorrelation coefficient $r(k)$ was entered in the L-D Recursion (Levinson-Durbin Recursion) to obtain a_k . This calculation was recursive, the calculated a_k was recurred to a_{k+1} for calculation, and the autocorrelation coefficient results in $r(0)=1$, so $r(0)=a_0=1$. The algorithm is shown below:

$$\begin{aligned} & E^{[0]} = r(0) \\ & \text{for } k=1 \text{ to } P \\ & \quad a_0^{[i-1]} = 1 \\ & \quad m_k = \left[r(k) - \sum_{j=0}^{j-1} a_j^{[k-1]} r(k-j) \right] / E^{[k-1]} \\ & \quad a_k^{[k]} = m_k \\ & \quad \text{for } j=1 \text{ to } k-1 \\ & \quad \quad a_j^{[k]} = a_j^{[k-1]} - m_k a_{k-j}^{[k-1]} \\ & \quad \text{end} \\ & \quad E^{[k]} = (1 - m_k^2) E^{[k-1]} \\ & \text{end} \end{aligned}$$

where, p is the prediction order number, m_k is the reflection coefficient, finally $a_1, a_2, a_3, \dots, a_k$ are figured out, there are k linear prediction coefficients, the a_k is substituted in Eq.4, multiplied by z transformation.

This transformation is Eq.1 after DFT, and then substituted in Eq.2 to obtain the energy, when Eq. 4 is calculated, in order to express the change in the volume of sound heard practically. The $H(z)$ is multiplied by $20 \cdot \log_{10}$ to obtain the dB, so as to determine the envelope line of vocal sound. This envelope line can be used to determine the formants, the

peak value of each formant has an index. The frequency is determined according to the index, the frequency calculation equation is shown below.

$$frequency(x) = \frac{fs}{framesize} \times index \quad (6)$$

where, $frequency(x)$ is the frequency of peak value No. x , fs is the sampling frequency, $framesize$ is the frame size, $index$ is the index of peak value, for example, $16000 \text{ Hz} / 512 * 14 = 437.5 \text{ Hz}$. Figure 3 shows the example of formants of frequency spectrum.

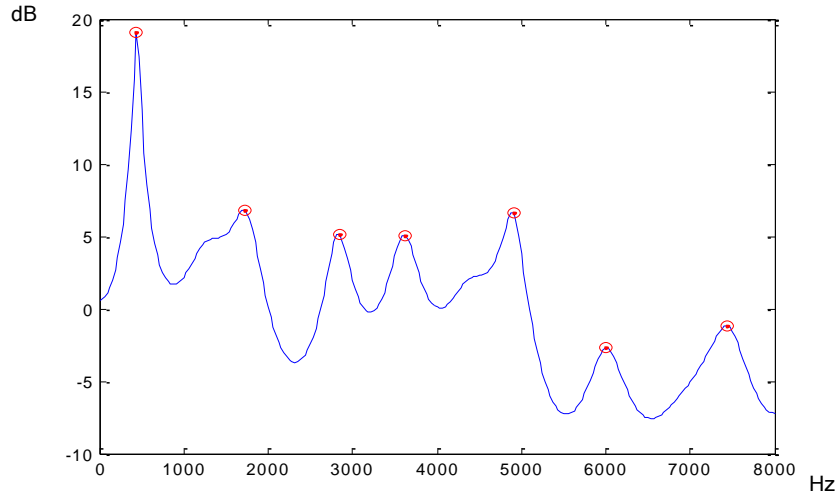


Figure 3. Example of formants of refrain frequency spectrum

Figure 3 shows the vocal sound frequency spectrum of 0~8000 Hz formants, the horizontal axis is frequency, and the vertical axis is dB, the small circles in the frequency spectrum are peak values of formants, named from the first small circle on the left. The first one is F_1 , the second one is F_2 , ..., the seventh one is F_7 . This frequency spectrum has 7 peak values of formants, lower spectral frequency means lower frequency, called low frequency, and higher frequency is called high frequency. F_1 of this spectrum is the frequency lower than 1000 Hz. It is the peak value of maximum sound energy, meaning the frequency of vocal sound is concentrated on F_1 .

The first three formants at frequencies F_1 , F_2 and F_3 were used as the characteristic values of vocal sound. F_1 and F_2 were the frequency characteristic values of normal voice, but F_3 was the characteristic value of peculiar frequency of singer's vocal sound [3]. As the refrain vocal sound has many frequency spectra, there are F_1 , F_2 and F_3 characteristic values. This study obtained the average and standard deviation of characteristic values of F_1 , F_2 and F_3 of all frames as the characteristic values of singer's formants.

3.3 Similarity of singer's vocal sound

Many studies on multimedia information retrieval use Euclidean distance function to measure the distance between data. When the Euclidean distance is used to calculate the music similarity, the dimensions are mostly assumed to have the same importance. However, the characteristic values of singer's voice have different effects on the calculation of

similarity. Considering the complexity of calculation, this study used the weighted squared Euclidean distance instead of square root to calculate the distance between songs.

This study surveyed volunteers interested in singer classification, and obtained proper weight values of dimensions according to the questionnaire answers.

First, the seed refrain q was selected randomly, and then several refrains similar to q were selected as evaluated objects. The user graded the similarity of singer's vocal sound according to the songs selected by the system. The grading range was 0~10 points, the higher the similarity was, the higher the grade was, 0 point means dissimilar at all. The refrains were sequenced according to grades, and compared with the sequence determined by the system using distance equation. The weight values of various dimensions of distance equation were adjusted by using genetic algorithm, so as to minimize the difference between the sequence calculated by the system and the sequence determined by the user. This study did not ask the users to evaluate the distance between singers, because 1) it is difficult for the grader to give an absolute distance value to describe the distance relation between two vocal sounds in a multi-dimensional space; 2) generally, when two vocal sounds are dissimilar to a certain extent, the grader only feels they are dissimilar to each other, it is hard to give a value to express the extent. Therefore, the proposed method hopes that the grader can easily answer the relative relation between vocal sounds in space. The distance between vocal sounds X and Q is calculated as follows.

$$D(X, Q) = \sum_{n=1}^d w_n (x_n - q_n)^2, \sum_{n=1}^d w_n = 1, w_n \geq 0 \quad (7)$$

where, x_n and q_n are the characteristic values of Dimension n of vocal sounds X and Q .

There is no grading data at the beginning, so the dimensions have the same weight value at the beginning, i.e., $w_1=w_2=\dots=w_n=1/n$. Therefore, the weight value of dimensions for the initial questionnaire is set as $1/n$, and the K refrain segments closest to Q are found out by using k NN.

M is the collection of K refrain segments closest to Q searched out by k NN, $M=\{S_i | S_i \in k$ NN of $Q, i=1,2,\dots,K\}$, the grader grades the similarity between the vocal sound of k songs and the vocal sound of Q . If the grading result of song S is $R(S)$, the sequence $Sort_{user_}M=\{S_{id1}, S_{id2}, \dots, S_{idk}\}$ is obtained from the descending order of grades, $R(S_{idp}) \geq R(S_{idp+1})$. If some vocal sounds have the same grade, they would be arranged in ascending order of their numbers in the database.

At a specific W , the distance equation of system can arrange M in ascending order of distance to generate the following sequence: $Sort_M=\{S_{id1}, S_{id2}, \dots, S_{idk}\}$, $D(S_{idp}, Q) \leq D(S_{idq+1}, Q)$. If some vocal sounds are equally distant to Q , they shall be arranged in ascending order of their numbers in the database.

The variability between the grader's grading and the system's grading was quantified, and the sequence position difference function was defined. If $P(S_{idx}, Seq)$ represents the position of S_{idx} vocal sound in the sorted sequence Seq , for example, the position of vocal sound S_5 in $Sort_{user_}M$ sequence is 8, and it is 3 in $Sort_M$ sequence, then $P(S_5, Sort_{user_}M)=8$, $P(S_5, Sort_M)=3$. Under the conditions of parameter W and seed vocal sound q , the difference is defined as:

$$Dif(S_{idx} | W, Q) = |P(S_{idx}, Sort_{user_}M) - P(S_{idx}, Sort_M)| \quad (8)$$

Following the above example: $Dif(S_5 | W, Q)=5$.

If there are T times of inquiries and answers, $QS=\{Q_1, Q_2, \dots, Q_T\}$ is the collection of random seeds, M_q is the collection of corresponding similar vocal sounds, the problem is defined as follows.

$$W' = \arg \min_W \sum_{Q_i \in QS} \left(\sum_{S_x \in M_{q_i}} Dif(S_x | W, Q_i) \right) \quad (9)$$

There are too many possible combinations of adjustment of weight value W ($O(k^d)$, k is the number of probable values of w , d is the dimension of vocal sound characteristic value). This study used Generate Algorithm (GA) to determine a better weight value setting. The GA is a searching algorithm for solving optimization in computational science. The GA is developed from some phenomena in evolutionary biology. These phenomena include inheritance, mutation, selection and crossover [13, 14].

The value w_n of each dimension of W is expressed in 16 bits, and

$$\sum_{n=1}^d w_n = 1, w_n \geq 0.$$

The connection of values of d dimensions forms a chromosome. We define the fitness function as follows.

$$f(W) = \frac{1}{\sum_{Q_i \in QS} \left(\sum_{S_x \in M_{q_i}} Dif(S_x | W, Q_i) \right) + 1} \quad (10)$$

After evolution of enough number of generations of GA, each dimension will have proper weight value. Due to GA is a probability method, the weight values will be different in different experiments. Therefore, we don't show the value. If the readers want to rebuild the experiment, the reader can turn the number of generations and the other parameters of GA.

4. Experiments and Evaluation

This study developed a simple search system for singers with similar vocal sound. The users can type in the singer's name, and the system uses the refrain segments of this singer and the distance to find k singer names with similar vocal sound. This study collected 300 songs of 50 singers in this system, and calculated the characteristic values of refrains before experiment. All the singers belong to pop music artists. The search efficiency of system is defined as follows.

$$precision = \frac{N_s}{k} \quad (11)$$

where, k is the quantity of similar singer names returned from the system, N_s is the quantity of similar singer names in the returned data regarded by the user.

In the first experiment, we want to know the inference of k value to precision. Moreover, Euclidean distance is adopted to measure the difference of song. The experimental result is shown in Figure 4. Moreover, the experimental result shown in Figure 5 is under the weighted squared Euclidean distance. To get the parameters of weights, the number of similar segments of refrain q is 5 in the questionnaire. Each questionnaire for volunteer (20 undergraduate students) contains 5 questions about distance measurement. It's clear to find

that the precision result under using weighted squared Euclidean distance is better than the one using typical Euclidean distance.

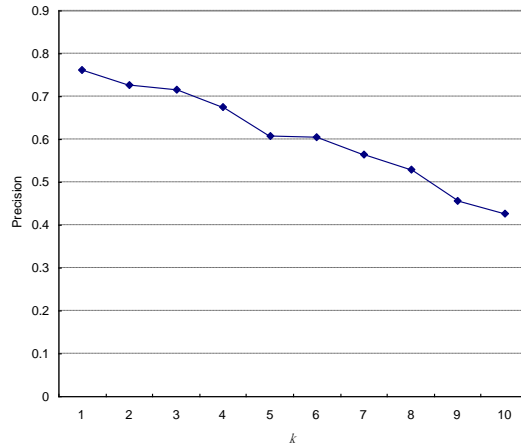


Figure 4. Search efficiency result under Euclidean distance measurement

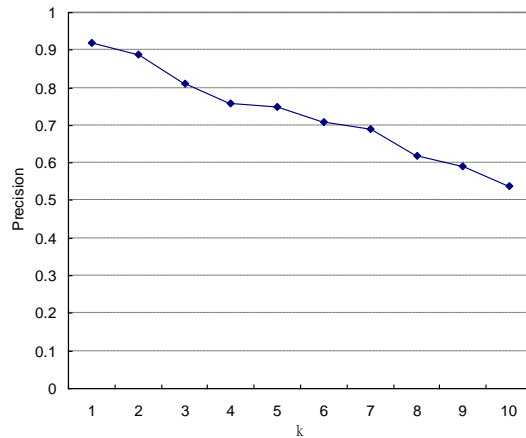


Figure 5. Search efficiency result under weighted squared Euclidean distance measurement

According to the experimental result, when the search quantity is less than 10, the average number of singers with similar vocal sound is higher than 50%. The precision value decreases as the k value increases.

Another experiment assumed that the user forgot singer A, and wanted to find singer A's name according to singer B's name. This study aimed to know whether singer A's rank in the search result can be higher. The experimental result is shown in Table 1.

Table 1. Experimental results of retrieval rank

Rank	1~5	6~10	11~15	16~200
%	41	34	19	6

As seen, A is in top five at 41%, in top ten at 75%, and in top 15 at 94%. In other words, the users can find the singer's name out of top 15, proving that the proposed method has a good effect.

5. Conclusions

The similarity of vocal sound of singers can be identified effectively by using the characteristic values captured from DFT and formants as well as the weighted squared Euclidean distance. This study designed a supporting system for music search, which searches the names of forgotten similar singers according to given singer names. The experiment proved that the proposed method has good efficiency.

The refrain vocal sound capture is the key in this study. At present, it is processed manually, but it will be very inconvenient if there are massive singer data. The future study will design a proper method to find the signal segments of singer's vocal sound automatically, so as to enlarge the application scale of this search system.

Acknowledgements

This work was supported in part by the NSC in Taiwan under the contact numbers NSC101-2221-E-020-025 and NSC102-2218-E-020-002-. Furthermore, I thank Cheng-Yu Chiang and Hsien-Meng Hsieh for their implementation of the experiment.

References

- [1] Y. E. Kim and B. Whitman, "Singer identification in popular music recordings using voice coding features", Proceedings of 3rd International Conference on Music Information Retrieval, (2002), pp. 164-169.
- [2] I. V. Bele, "The speaker's formant", Journal of Voice, vol. 20, no. 4, (2006), pp. 555-578.
- [3] P. R. Cook, "Identification of control parameters in an articulatory vocal tract model with applications to the synthesis of singing", PhD Thesis, Stanford University, USA (1991).
- [4] T. J. Millhouse and F. Clermont, "Perceptual characterization of the singer's formant region: a preliminary study", Proceedings of 11th Australasian International Conference on Speech Science and Technology (2006), pp. 253-258.
- [5] W. H. Tsai and H. M. Wang, "Automatic singer recognition of popular music recordings via estimation and modeling of solo vocal signals", IEEE Transactions on Speech and Audio Processing, vol. 14, no. 1, (2006), pp. 330-341.
- [6] D. Chow and W. H. Abdulla, "Robust speaker identification based on perceptual log area ratio and Gaussian mixture models", Proceedings of 8th International Conference on Spoken Language Processing, (2004), pp. 1761-1764.
- [7] M. F. McKinney and J. Breebaart, "Features for audio and music classification", Proceedings of 4rd International Conference on Music Information Retrieval, (2003), pp. 151-158.
- [8] D. Pye, "Content-based methods for the management of digital music", IEEE International Conference on Acoustics, Speech, and Signal Processing, (2000), pp. 2437-2440.
- [9] K. West and S. Cox, "Features and classifiers for the automatic classification of musical audio signals", Proceedings of 5rd International Conference on Music Information Retrieval, (2004), pp. 1-6.
- [10] C. C. Cheng and C. T. Hsu, "Content-based audio classification with generalized ellipsoid distance", Proceedings of the Third IEEE Pacific Rim Conference on Multimedia, (2002), pp. 328-335.
- [11] M. Negnevitsky, "Artificial intelligence: a guide to intelligent systems", 2nd edition, New York: Addison-Wesley, (2002).
- [12] J. Han and M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann, (2001).
- [13] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm", IEEE Intelligent Systems and their Applications, vol. 13, no. 2, (1998), pp. 44 - 49.
- [14] N. H. Liu, "Comparison of content-based music recommendation using different distance estimation methods", Applied Intelligence, vol. 38, issue 2, (2013), pp. 160-174.

Author



Ning-Han Liu received the BS and MS degree from National Taiwan Normal University in 1992 and 1995, respectively and the PhD degree in computer science from National Tsing-Hua University in 2005. He is an associate professor at Department of Management Information Systems, National Pingtung University of Science and Technology, Taiwan since 2010. His research interests include multimedia databases, audio processing in WSN, and artificial intelligence. Dr. Liu is a member of the IEEE Consumer Electronics Society. He is also a member of TAAI (Taiwanese Association for Artificial Intelligence).