SCIENCE CHINA
Information Sciences

• LETTER •

January 2019, Vol. 62 019202:1–019202:3
https://doi.org/10.1007/s11432-018-9513-3

Multidimensional zero-crossing interval points: a low sampling rate acoustic fingerprint recognition method

Xianghu YUE[1], Fang DENG[1,2*] & Yue XU[2]

[1]*School of Mathematics and Statistics, Beijing Institute of Technology, Beijing* 100081*, China;*
[2]*School of Automation, Beijing Institute of Technology, Beijing* 100081*, China*

Received 9 May 2018/Accepted 15 June 2018/Published online 7 November 2018

Citation Yue X H, Deng F, Xu Y. Multidimensional zero-crossing interval points: a low sampling rate acoustic fingerprint recognition method. Sci China Inf Sci, 2019, 62(1): 019202, https://doi.org/10.1007/s11432-018-9513-3

Dear editor,
In recent years, acoustic fingerprint recognition technique has been widely utilized in areas such as battle environments, environmental monitoring, security systems, and personal identification [1], which plays a pivotal part in recent effort to perfect machine audition. Currently, the most commonly used acoustic fingerprint characteristics include linear predictive cepstrum coefficient (LPCC), and Mel frequency cepstrum coefficient (MFCC). LPCC creates the voice model based on an all pole model, which, in essence, is complex cepstrum, i.e., logarithm elimination after signal Z transformation, followed by inverse Z transformation [2]. MFCC creates a sound feature extraction model according to human ear auditory characteristics [3]. Although LPCC and MFCC both have high recognition precision, the sampling rate is always over 8 kHz, resulting in large data amounts and complex computations, imposing high requirements for the hardware [4]. Therefore, the two methods are not suitable for resource constrained equipment, such as wireless sensor network (WSN) nodes, which have limited memory space and computation capability [5–7].

Zero-crossing interval points are commonly used in electronics, mathematics and image processing [8]. Due to its simplicity, zero-crossing features could be used for sound event recognition in WSN at a low sampling rate (e.g., 3 kHz). At the same sampling rate, the amount of calculation is greatly less compared with the two above methods. However, the traditional method of one-dimension zero-crossing interval points has dramatically reduced recognition precision. In the present study, a low sampling rate acoustic fingerprint recognition method based on the multidimensional zero-crossing interval points is presented. By the expansion of characteristic dimensions that the traditional method extracted using zero-crossing interval points, the proposed method improves the recognition rate and noise resistance without increasing the sampling points or sampling rate. Compared with the traditional zero-crossing recognition method, the current algorithm increased the dimensions of recognition, and showed high recognition precision at low sampling rate, e.g., 3 kHz. Thus, it could meet the application requirements of resource constrained equipment for low sampling rate, small data amount and computation.

Methodology. Determination of zero-crossing points. The median value of sound wave data was determined, which was the sampling value X without presence of any sound. The values of sampling points were $X(1), X(2), \ldots, X(k)$, respectively. If $X(i)$ satisfies

$$(X(i) - X)(X(i+1) - X) \leqslant 0 \quad (i \leqslant k-1), \quad (1)$$

then $X(i)$ is called a zero-crossing point. Let $y(n) = i$, where n represents the n-th zero-crossing point. All zero-crossing points are recorded as $y(1), y(2), \ldots, y(\varepsilon)$, where ε is the total number

* Corresponding author (email: dengfang@bit.edu.cn)

of zero-crossing points.

Counting zero-crossing interval points. The number of sampling points between two adjacent zero-crossing points ($y(i+1)$ and $y(i)$) were collected and saved into vector L_1:

$$L_1 = \{\ell_i | \ell_i = y(i+1) - y(i) - 1\} \ (i = 1, 2, \ldots, \varepsilon - 1).$$

The occurrence number of every element in matrix L_1 was collected. $\Im_1(\ell_i)$ represents the probability of the occurrence of element ℓ_i in L_1. The results were stored in vector W_1:

$$W_1 = \{w_{\ell_i} | w_{\ell_i} = \Im_1(\ell_i)\} \quad (0 \leqslant \ell_i).$$

W_1 was taken as the first dimensional vector of the acoustic fingerprint characteristics matrix (The end of ℓ_i collection of interval point depended on the sampling rate and target sound source). In the experiment, the sampling rate was 3 kHz, and target sound sources were tank, airplane, train and human voice. If more than 10 sampling points of ℓ_i were identified between two zero-crossing points, noise was believed to be the reason. So the situation of $\ell_i > 10$ was ignored. The two steps above were traditional feature extraction process of one-dimensional zero-crossing interval points acoustic fingerprint recognition method.

Expansion of characteristic dimensions. The number of sampling points between every other zero-crossing points $y(i+2)$ and $y(i)$ were collected and stored in matrix L_2, $i = 1, 2, \ldots, \varepsilon - 2$. The number of appearance of every element in matrix L_2 was collected. The results were stored in vector W_2, which was taken as the second dimensional vector of the acoustic fingerprint characteristics matrix. Similarly, the number of sampling points between every $2, 3, \ldots, N-1$ zero-crossing points were collected and stored in matrices $W_3, W_4, \ldots, W_N$. The equation of L_N is

$$L_N = \{\ell_i | \ell_i = y(i+N) - y(i) - 1\}$$
$$(i = 1, 2, \ldots, \varepsilon - N),$$

and the equation of W_N is

$$W_N = \{w_{\ell_i} | w_{\ell_i} = \Im_N(\ell_i)\} \quad (0 \leqslant \ell_i).$$

Establishment of multidimensional characteristic matrix. The short characteristic vectors in $W_1, W_2, \ldots, W_N$ were filled with zero, thereby generating a multidimensional characteristic matrix Φ with N characteristic vectors

$$\begin{aligned} W_1' &= [W_1, 0, \ldots, 0], \\ W_2' &= [W_2, 0, \ldots, 0], \\ &\vdots \\ W_N' &= [W_N, 0, \ldots, 0], \\ \Phi &= [W_1', W_2', \ldots, W_N']. \end{aligned}$$

The algorithm presented in this study is suitable for scenarios of low sampling rate and low energy-consuming acoustic fingerprint recognition, so the classification method with complex computation was not suitable. The least distance classifier is a simple but effective method in pattern recognition technology. Under conditions where the amount of characteristics is small, the recognition precision is comparable to other complex classifiers. Therefore, the least distance classifier was adopted in this study.

For an acoustic fingerprint recognition system, sampling frequency and number of sample points are two important factors for equipment power consumption and recognition rate [9]. In this study, considering the frequency range and degree of difference of the target sound type, a sampling frequency of 3 kHz was used to collect 50 sample points so as to verify efficiency of acoustic fingerprint recognition algorithm for a specific sound type under 8 kHz.

Experiments. CC2530 and ADMP401 sound sensor were used to collect sound signal. A sound recorder was used to play different sounds for the recognition experiment. The experimental scenario was shown in Figure 1(a). The sound recorder played the sound of a car, airplane and train, as well as a human voice. The sound sensor captured the sound signal and sent it to the coordinator for matching-up recognition, and the results were then sent to the computer. The characteristic matrix elements are the probability of occurrence of interval points. Thus, the Euclidean distance between the template and sample sounds was small, and for better observation, results in Figures 1 were magnified by 1000 times. Experimental results of different dimensional characteristic matrices of 4 different types of sound at different SNRs are shown in Figure 1(b)–(f). It can be observed that, when the sampling rate and number of sample points were small, the traditional zero-crossing interval points recognition algorithm was not able to recognize the sound signal effectively, whereas the proposed algorithm showed promising recognition rate at different SNRs.

Conclusion. A new acoustic fingerprint recognition method demonstrated high recognition precision with low sampling rate and small number of sample points. The proposed algorithm showed great increase in recognition precision, meanwhile, it has low requirements for hardware, computation complexity and power consumption compared to MFCC and LPCC, so it is suitable for re-

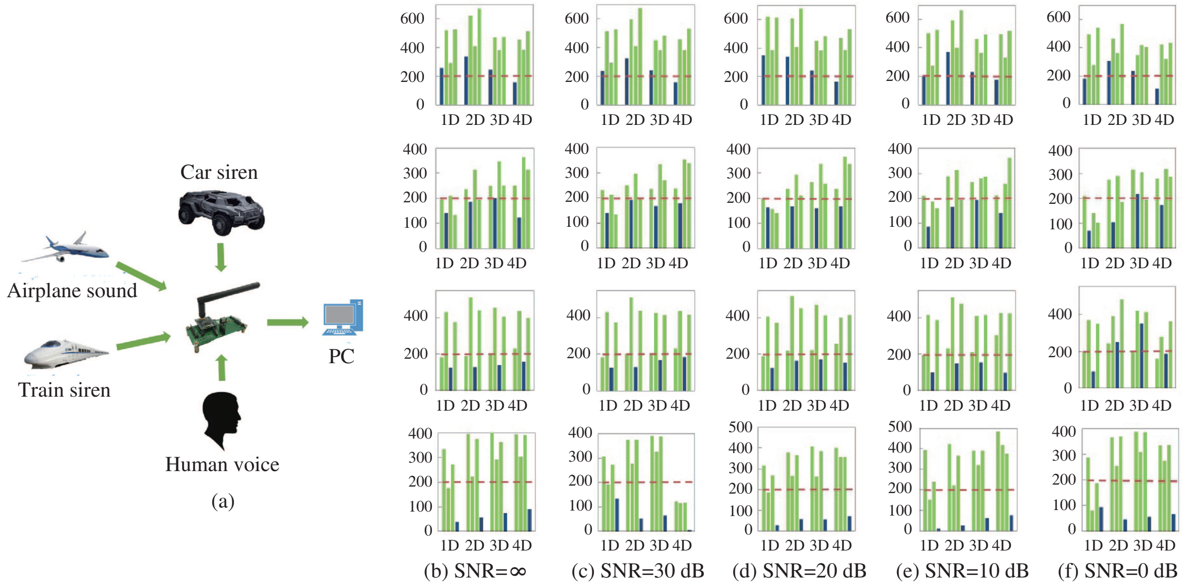

Figure 1 (Color online) (a) Experimental scenario and (b)–(f) Euclidean distances of different dimensional characteristic matrices of four different samples at different SNRs. From top to bottom is car, human voice, airplane and train, respectively. The red dash is the threshold value $D_{\rm th}$ of the Euclidean distance. The blue bar represents the characteristics matrices Euclidean distance between sample sound and corresponding template sound, while the distances between sample sound and the other template sounds are shown in green.

source constrained equipment to conduct acoustic fingerprint recognition of several specific sound types. Results showed that the proposed multi-dimensional zero-crossing interval points acoustic fingerprint method required a sampling rate 62.5% lower than the traditional method. At the same sampling rate, the recognition rate was increased by about 50% compared to the traditional method. Additionally, our method had good noise resistance.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61304254, 61321002) and Beijing NOVA Program (Grant No. xx2016B027).

References

1 Shultz D. When your voice betrays you. Science, 2015, 347: 494
2 Kim H K, Choi S H, Lee H S. On approximating line spectral frequencies to LPC cepstral coefficients. IEEE Trans Speech Audio Process, 2000, 8: 195–199
3 Venturini A, Zao L, Coelho R. On speech features fusion, α-integration Gaussian modeling and multi-style training for noise robust speaker classification. IEEE/ACM Trans Audio Speech Lang Process, 2014, 22: 1951–1964
4 Misra S, Das T, Saha P, et al. Comparison of MFCC and LPCC for a fixed phrase speaker verification system, time complexity and failure analysis. In: Proceedings of International Conference on Circuit, Power and Computing Technologies, Nagercoil, 2015
5 Gungor V C, Hancke G P. Industrial wireless sensor networks: challenges, design principles, and technical approaches. IEEE Trans Ind Electron, 2009, 56: 4258–4265
6 Deng F, Guan S P, Yue X H, et al. Energy-based sound source localization with low power consumption in wireless sensor networks. IEEE Trans Ind Electron, 2017, 64: 4894–4902
7 Zheng J M, Tan Y A, Zhang Q K, et al. Cross-cluster asymmetric group key agreement for wireless sensor networks. Sci China Inf Sci, 2018, 61: 048103
8 Wang C K, Tseng Y H, Wang C K. A wavelet-based echo detector for waveform LiDAR data. IEEE Trans Geosci Remote Sens, 2016, 54: 757–769
9 Taseska M, Habets E A P. Informed spatial filtering for sound extraction using distributed microphone arrays. IEEE/ACM Trans Audio Speech Lang Process, 2014, 22: 1195–1207