Online Music Style Recognition via Mobile Computing

Lizhu Yuan, Jilin Normal University, China* Yue Zhang, Jilin Normal University, China

ABSTRACT

Music is a widely used data format in the explosion of internet information. Automatically identifying the style of online music in the internet is an important and hot topic in the field of music information retrieval and music production. Recently, automatic music style recognition has been used in many real-life scenes. Due to the emerging of machine learning, it provides a good foundation for automatic music style recognition. This paper adopts machine learning technology to establish an automatic music style recognition system. First, the online music is processed by waveform analysis to remove the noises. Second, the denoised music signals are represented as sample entropy features by using empirical model decomposition. Lastly, the extracted features are used to learn a relative margin support vector machine model to predict future music style. The experimental results demonstrate the effectiveness of the proposed framework.

KEYWORDS

Empirical Model Decomposition, Mobile Computing, Music Style Recognition, Relative Margin Support Vector Machine, Waveform Analysis

1. INTRODUCTION

With the rapid development of the Internet, people begin to have the opportunity to come into contacting with different music styles around the world and enjoy the pleasure brought by music (Kohut 2018). Although the music has some different expression ways in different countries or nations, the music has always been able to express people's thoughts, convey people's thinking (Stewart 2019). The music fully expresses its value in human life.

Up to now, there are countless kinds and quantities of musical instruments, such as guitar, piano, erhu, and cello (Zhao 2019). At the same time, the storage methods of music files are becoming diversification, including WAV files, MIDI files, MP3 files and WMA files (Fu 2021). Each file format has its own storage characteristics. WAV files are easy to generate and edit, but the compression ratio is not high. MIDI file has small size and storage space and is especially adapted for the needs of long time music. MP3 file is also a very common music format, which has high compression rate and poor sound quality.

The emergence of musical instruments and the diversification of music storage methods promote the formation of music genres (Ojanen 2020). The common music genres include jazz, classical, pop, hip-hop, rock and roll, etc. Recently, music information retrieval has become a new research topic, and a large number of researchers have studied this field (Murthy 2018). Because some users are only interested in one genre of music, music recognition and classification system can divide music

DOI: 10.4018/IJMCMC.293748

*Corresponding Author

This article published as 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 production in any medium, provided the author of the original work and original publication source are properly credited. into different genres. In this way, it is convenient for users to search and manage music in different periods of time, such as sports and rest. When the same song is sung by different people, the songs are different due to the different range, timbre and the performance of various musical instruments. Exactly extracting features of music leads to the low efficiency of music genre classification and recognition. With the deepening of a large number of research works, the music genre classification and generation will gradually develop in a better direction.

With the rapid development of music storage and computer technology, digital audio processing technology has been greatly improved (Gao 2019). Music processing software is also diversified, such as windows recorder, tools when buying a sound disc (Liang 2021). Speech recognition, text to speech conversion, speech compression coding and other technologies have gradually begun to use computer technology instead of manual completion (Lero 2019; Delić 2019). These methods are based on digital audio processing. In general, a piece of music is vectored. The features of the music are represented as a series of numbers to store and transmit. Different formats of music are finally represented in digital form, which makes the analysis of music become very simple and efficient by combining with the music processing software in the computer. At the same time, with the rapid development of computer technology, the accuracy of analyzing and processing music has been greatly improved, and the cost is reduced.

In the field of music style recognition, the most important thing is the extraction of related features and the selection of classifiers (Ghosal 2018). When selecting different music feature vectors for music style recognition, different classification results will be produced. At present, the commonly used features are tone, timbre, loudness etc. (Prince 2019) In order to improve the effect of style recognition, researchers have taken many methods, such as signal processing method (Wang 2019). These methods improve the effect of music classification to a certain extent. However, the feature extraction is still difficult under some special cases. In order to further improve the performance of music style recognition, it is necessary to deeply mine the internal association between music signals.

With the development of machine learning and deep learning in face recognition, speech recognition and image recognition, people gradually try to apply machine learning and deep learning to the field of music generation and recognition. The deep learning is widely used in storing and processing large amounts of data, such as recurrent neural network (Purwins 2019). However, we cannot get the data of the previous moment or earlier, which makes the effect of classification or the acquisition of music features, such as tone, timbre, loudness, rhythm, be inaccurate. This paper adopts a robust support vector machine to avoid the issue that training set needs massive samples in deep learning for online music style recognition. The flowchart of music style recognition is summarized as following figure.



Figure 1. The architecture of music style recognition

2. PROBLEM DESCRIPTION AND PREVIOUS WORK

The music style recognition refers to classify electronic music according to the associated features. In essence, it is a process of pattern classification, which is a very complex process and involves many disciplines, such as psychology, signal processing, pattern recognition, etc. The format of electronic music in the network is different from that in real life. The electronic music has its own special format. The scholars put forward a content-based electronic music recognition model, which extracts the mean value and autocorrelation coefficient of electronic music as features, and then establishes an electronic music recognition model (Hong 2018).

Different genres have different music styles. Music style or music genre recognition has been widely studied for a long time (Vishnupriya 2018). Since the 1990s, people began to use artificial methods to determine the genre and style of music. The famous "music chromosome project" asks music experts divide music into different types through their understanding and technology of music. However, due to the limited technical conditions, the artificial method is very immature when facing massive music data. Additionally, different experts have slightly bias on understanding of different music genres. This project spends a lot of financial and material resources, but the result is not satisfactory. Driven by this situation, people began to study the automatic classification algorithm of music genre style recognition.

In 1996, American researchers proposed a classification algorithm, which calculates the mean, variance and autocorrelation coefficient from massive music data. These features are used to further analyze the characteristics of music, such as loudness, tone and other features that people can easily feel. Then, music styles are determined by classifiers. Many improvements are based on these methods. For instance, the acoustic features, including timbre, rhythm, and pitch, were introduced to represent the features of music. However, due to high dimensionality of these features, the feature selection algorithm is used to reduce the dimension of features, which can reduce the calculation and remove the irrelevant redundant information.

With the development of computer science and artificial intelligence technologies, machine learning has been applied to music genre recognition, such as k-nearest neighbors, conditional random field, linear discriminant analysis, Gaussian mixture model and Markov model. With the application of wavelet transform theory, the statistical method is introduced to calculate the statistical value of wavelet coefficients which are used to combine with the classification models in machine learning, such as linear discriminant analysis, Gaussian mixture model, Markov model. In order to further improve the essential features of electronic music, the unsupervised dimensionality reduction is introduced into music style recognition which further improves the music representation. Another way is to introduce high order matrix features to represent electronic music which can improve the effect of classification and recognition. However, the dimensions increase significantly, which makes the calculation process very complicated and the training be very time-consuming.

With the development of deep learning, more and more researches pay attentions to applied deep learning in music generation, such as long short term memory network, recurrent neural network. Generally, the long short term memory network is superior to recurrent neural network for music style recognition. The reason is the difference of the structure. Compared with deep learning, Markov model is relative simple. However, when applying Markov model to music generation, it cannot get the temporary structure for a long time. The Markov model can only learn some simple music examples and quality of produced music is not very good. Generative adversarial nets (GAN) is not very good in dealing with the variable content. One of the problems is that the generated music will add some rhythms different from the original music in it. The limited Boltzmann machine is slightly less controllable.

3. FEATURE EXTRACTION AND MUSIC STYLE RECOGNITION

The wavelet analysis can decompose the one-dimensional signal with different resolutions and refine the original signal (Balasubramanian 2018). In this paper, the wavelet analysis is used to remove the noise and improve the signal-to-noise ratio to facilitate the subsequent signal processing. The signal decomposition form can be described as following equation.

$$\begin{cases} x_k^j = \sum_n h_0 \left(n - 2k \right) x_n^{j+1}, & j \ge 0 \\ d_k^j = \sum_n h_1 \left(n - 2k \right) x_n^{j+1}, & j \ge 0 \end{cases}$$
(1)

In Equation (1), x_k^j represents the low frequency part of the original signal, d_k^j represents the high frequency part of the original signal, h_0 is a low pass filter, h_1 is a high pass filter, n is the wavelength of the filter, k is the scale of wavelet decomposition. Then,

$$k_1(n) = \left(-1\right)^{-n} h_0(N-n) \tag{2}$$

When the original signal after multi-layer decomposition, it can get different levels of coefficients. By using a denoising algorithm to process all levels of wavelet coefficients (such as soft threshold algorithm), the new signal can be represented as follows:

$$\hat{x} = T_h(\gamma, t) = \begin{cases} \operatorname{sgn}(\gamma)(|\gamma| - t), & |\gamma| \ge t \\ 0, & |\gamma| < t \end{cases}$$
(3)

In Equation (3), \hat{x} represents the waveform coefficients, t represents the threshold, $t = \sigma \sqrt{2 \log N}$.

The signal processed by soft threshold algorithm is reconstructed by wavelet analysis as follows:

$$x_{n}^{j+1} = \sum_{k} h_{0} \left(n - 2k \right) x_{k}^{j} + \sum_{k} h_{1} \left(n - 2k \right) d_{k}^{j}$$

$$\tag{4}$$

The original signal is decomposed and reconstructed by wavelet analysis. The noise in the original signal can be effectively eliminated, which can greatly improve the signal-to-noise ratio of the signal. The denoised signal is further processed by empirical model decomposition (Wei 2018).

Empirical mode decomposition can decompose a signal into several intrinsic model function (IMF) components and residual components

$$x(t) = \sum_{i=1}^{c} IMF_i(t) + r(t)$$
(5)

In Equation (5), c represents the number of intrinsic mode function components, IMF_i represents the i^{th} intrinsic mode function component, r(t) represents residual component.

Let $\{x_1, \ldots, x_n\}$ represent a music signal sequence, *m* represent the number of dimensions of music signal, *r* represent similarity tolerance. The sample entropy can be computed by the following procedure.

Step (i). The music signal sequence is used to construct a vector according to the sequence number as follows:

$$X(i) = \left\{ x(i), x(i+1), \dots, x(i+m-1) \right\}$$
(6)

Step (ii). The distance between two sequences X(i) and X(j) as d(X(i), X(j)) as follows

$$d\left(X(i), X\left(j\right)\right) = \max_{k=0 \to m-1} \left\{ X\left(i+k\right) - X\left(j+k\right) \right\}$$

$$\tag{7}$$

Step (iii). Counting the number of signals with d(X(i), Y(j)) < r for i = 1, ..., N - m + 1, j = 1, ..., N - m + 1. The N represents the length of signal. Then, the ratio between n_i^m and the sum of distance N - m is represented as follows:

$$C_i^m\left(r\right) = \frac{n_i^m}{N - m} \tag{8}$$

Step (iv). Let m = m + 1, repeating Step (i) to (iv) to compute $C^{m+1}(r)$.

Step (v). The sample entropy is represented as follows:

$$SE = -\ln \frac{C^{m+1}(r)}{C^{m}(r)}$$
(9)

The parameters \$m\$ and \$r\$ are close to the sample entropy. In general, they can set according to user experience.

The denoised music signals are decomposed by empirical model decomposition to obtain intrinsic mode function components. Then, the sample entropy is computed according to above procedure. The music style features are represented as $[SE_1, \ldots, SE_k]$. Here, SE_k represents the sample entropy of the k^{th} intrinsic mode function component. The whole feature extraction of music style is summarized as following figure.

International Journal of Mobile Computing and Multimedia Communications Volume 13 • Issue 2





The electronic music is represented as sample entropy features. Then, the sample entropy features of all electronic music are used to construct a training set. Then, the training set is used to learn a classifier to predict future music style.

The support vector machine is a widely used classification algorithm. However, it treats every samples in training set equally and is not robust to the noises in the training set. In order to solve this issue, instance-weighted support vector machine (Zhu 2016) and relative margin support vector machine (Zhu 2017) are proposed. In this paper, we adopt relative margin support vector machine as the classification algorithm in music style recognition.

Let $\{X, Y\}$ represent the training set consisting of $\{x_i, y_i\}$ (i = 1, ..., l) in which x_i represents music sample entropy features and y_i represents the music style. When there are only two styles, $y_i \in \{+1, -1\}$. The aim of relative margin support vector machine is to find a hyperplane $f(x) = w^T \phi(x) + b$ via maximizing the minimum relative margin between two classes. The $\phi(x_i)$ represents the mapping of sample x_i in reproducing kernel Hilbert space (RKHS). The mapping function $\phi(x)$ is an implicit function. However, the inner product between mappings $\phi(x_i)$ and $\phi(x_j)$ can be easily computed by a kernel function, $\langle \phi(x_i), \phi(x_j) \rangle = K(x_i, x_j)$. The common used kernel function includes polynomial kernel, Gaussian kernel, hist kernel etc. When the linear kernel is used, the hyperplane is degraded as $f(x) = w^T x + b$.

The hyperplane of relative margin support vector machine can be obtained by solving the following optimal programming.

$$\min_{\mathbf{w},\rho} \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{l} \xi_i
s.t \quad y_i \left(\mathbf{w}^T \varphi \left(\mathbf{x}_i \right) + b \right) \ge \eta_i - \xi_i, \quad i = 1, 2, \dots, l
\quad \xi_i \ge 0, \quad i = 1, 2, \dots, l.$$
(10)

Due to Equation (10) is a convexity problem, the solution can be obtained through the associated dual programming. The η_i represents the relative margin of sample x_i , which is determined by prior

knowledge of the training set. The η_i is defined by relative density degree. The paper adopts prior knowledge of the training set by using nearest neighbors' distribution.

By introducing the Lagrange multipliers α_i for constraints $y_i \left(\mathbf{w}^T \varphi \left(\mathbf{x}_i \right) + b \right) \ge \eta_i - \xi_i$ (i = 1, ..., l) and β_i for constraints $\xi_i \ge 0$ (i = 1, ..., l), the Lagrange function is written as follows:

$$L = \frac{1}{2} \mathbf{w}^{T} \mathbf{w} + C \sum_{i=1}^{l} \xi_{i} - \sum_{i=1}^{l} \alpha_{i} \left(y_{i} \left(\mathbf{w}^{T} \varphi \left(\mathbf{x}_{i} \right) + b \right) - \eta_{i} + \xi_{i} \right) - \sum_{i=1}^{l} \beta_{i} \xi_{i}$$

$$\tag{11}$$

Let the partial derivatives of L for w, ξ_i , b be equal to 0, then the following equations hold

$$\mathbf{w} = \sum_{i=1}^{l} \alpha_i y_i \varphi\left(\mathbf{x}_i\right) \tag{12}$$

$$C - \alpha_i y_i - \beta_i = 0 \tag{13}$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \tag{14}$$

By substituting Equation (12), (13), and (14) into Equation (11), we can obtain the dual form of Equation (10) as follows:

$$\min_{\pm} \quad \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{l} \alpha_i \eta_i$$
s.t.
$$\sum_{i=1}^{l} y_i \alpha_i = 0$$

$$0 \le \alpha_i \le C, \quad i = 1, \dots, l$$

$$(15)$$

The Equation is a quadratic programming which can be solved by many existing algorithms. After obtaining λ parameters w and b in hyperplane f(x) can obtained by the following equations.

$$w = \sum_{x_i \in SVs} \alpha_i y_i \varphi\left(x_i\right) \tag{16}$$

$$b = y_i \eta_i - \sum_{j=1}^l \alpha_j y_j K\left(x_i, x_j\right), \quad 0 < \alpha_i < C$$

$$\tag{17}$$

In Equation (16), the \$SVs\$ represents the support vector set which consists of the samples with nonzero Lagrange multipliers. The final hyperplane is written as follows:

$$f(x) = \operatorname{sgn}\left(\sum_{x_j \in SVs} \alpha_j y_j K(x_i, x_j) + b\right)$$
(18)

The relative margin support vector machine can solve the two-class classification problem. For multi-class classification problem, it still needs to adopts one-to-one strategy or one-to-rest strategy to convert the multi-class classification problem as several two-class classification problems. For multi-class classification problem, the whole procedure of relative margin support vector machine is summarized as following algorithm.

Algorithm (relative margin support vector machine for multi-class classification)						
Step 1: converting the multi-class classification problem via one-to-one strategy or one-to-rest strategy as $c(c-1)/2$ two-class classification problems for one-to-one strategy or c for one-to-rest strategy;						
Step 2: calculating relative margin parameter η_i ; Step 3: solving relative margin support vector machine by Equation (15) for each two-class classification problem; Step 4: calculating values of Equation (16) for all samples in negative class.						

In Step 3 of reduced support vector machine, l_p represents the number of samples in positive class. In Step 6 of reduced support vector machine, l_n represents the number of samples in negative class. The parameter δ represents the size of retained subset of the training set. The larger δ is, the more samples are retained in reduced support vector machine.

4. EXPERIMENTS AND SIMULATIONS

This section first adopts GTZAN dataset (Sturm 2013) to evaluate proposed music style recognition method. Then, the proposed method is used to recognize the styles of online music. The GTZAN is a western music genre library, which is widely used to evaluate music style recognition methods. The GTZAN music genre library contains 10 music genres, including blues, classical, country, disco, hip hop, jazz, metal, pop, reggae and rock. Each music genre contains 100 audio tracks which last 30 seconds. Each track is with 22,050 Hz Mono 16-bit and stored in the form .wav format. Each track is split as three parts. Each part lasts 10 seconds. Then, there are 3,000 music clips. The clips are divided as two parts as the training set and test set, respectively. One part contains 2,500 clips (250 per music genre) as training set, while the other part contains 500 clips (50 per music genre) as test set.

In order to ensure the relative margin support vector machine can solve the multi-class classification problem, the one versus rest strategy is adopted in the relative margin support vector machine. The radial basis function is used as the kernel function. The recognition result is reported in the form of confusion matrix in the following table.

	blues	classical	country	disco	hip hop	jazz	metal	pop	reggae	rock	Acc. (%)
blues	41	4	0	0	0	4	1	0	3	0	77
classical	3	38	1	0	0	5	0	0	0	0	81
country	0	3	44	1	0	1	1	0	0	0	88
disco	2	2	2	39	2	0	0	4	4	7	63
hip hop	0	0	0	1	38	0	0	0	1	0	95
jazz	3	3	0	1	0	37	1	0	1	0	80
metal	1	0	3	0	1	0	43	0	1	0	88
pop	0	0	0	7	8	3	0	41	0	5	64
reggae	0	0	0	0	0	0	0	1	40	3	91
rock	0	0	0	1	1	0	4	4	0	35	78
Acc. (%)	82	76	88	78	76	74	86	82	80	70	79

Table 1. The confusion matrix of music style recognition for GTZAN dataset

From the results in Table, it can be found that the proposed method can identify 79% of the music genre in the test set. In Table 2, we compare the proposed method with some previous methods, including decision tree, k nearest neighbors, support vector machine, and linear discriminant analysis. In order to eliminate the randomness when spitting training set and test set, the experiments are repeated 20 times and the results in Table 2 are the mean of the twenty times. The experimental results are reported in terms of accuracy, F1-measure and precision. The best results in each column are highlighted by bold font.

Methods	Accuracy (%)	F1-mesuare	Precision (%)	
Decision tree	75.83	0.7746	78.91	
K nearest neighbors	74.73	0.7639	77.65	
Linear discriminant analysis	76.69	0.7929	80.59	
Support vector machine	77.89	0.8017	81.07	
Ours	79.67	0.8179	82.39	

Table 2. The performance comparison of music style recognition with different methods

From the results in Table 2, it can be found that the proposed music style recognition method archives 79.67%, 0.8179, 82.39% for accuracy, F1-measure, and precision, respectively. The proposed music style recognition method is superior to decision tree, k nearest neighbors, linear discriminant analysis, and support vector machine.

Furthermore, we also used the proposed music style recognition method to identify the style of the music in Internet. First, the music is crawled from Internet. Then, the crawled music is processed by wavelet transform and EMD to convert as extracted features. Lastly, the extracted features are input into the learnt music style recognition model to determine the associated music style. The predicted style is sent to music expert to determine whether the music is correctly identified. The experimental results show that 76.35% music from Internet can be correctly recognized.

International Journal of Mobile Computing and Multimedia Communications Volume 13 • Issue 2

5. CONCLUSION

In this paper, we propose a music style recognition framework based on machine learning and mobile computing technologies. First, a music genre library is constructed by using collected music files. Second, the collected music signals are extracted from these files and denoised by using waveform transform. Third, the denoised signals are processed by using empirical model decomposition to extract sample entropy features. Fourth, the sample entropy features are used to learn a classification model. In this paper, we adopt relative margin support vector machine. Since there are more than two music genres, the one versus rest strategy is adopted in relative margin support vector machine. The music in Internet is represented as associated sample entropy features to input into the learnt classification model. The experimental results demonstrate that the proposed music style recognition framework can identify more than 76% music.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or notfor-profit sectors.

REFERENCES

Balasubramanian, G., Kanagasabai, A., Mohan, J., & Seshadri, N. G. (2018). Music induced emotion using wavelet packet decomposition—An EEG study. *Biomedical Signal Processing and Control*, 42, 115–128.

Delić, V., Perić, Z., Sečujski, M., Jakovljević, N., Nikolić, J., Mišković, D., & Delić, T. et al. (2019). Speech technology progress based on new machine learning paradigm. *Computational Intelligence and Neuroscience*.

Fu, Q. (2021, March). Research on the Use of Computer Music in Modern Musical Composition. *Journal of Physics: Conference Series*, 1820(1), 012153.

Gao, D. (2019). Research on Network Transmission and Exchange Technology of Digital Audio. *International Journal of Information and Communication Sciences*, 4(4), 68.

Ghosal, D., & Kolekar, M. H. (2018, September). Music Genre Recognition Using Deep Neural Networks and Transfer Learning. In Interspeech (pp. 2087-2091). Academic Press.

Hong, S., Im, W., & Yang, H. S. (2018, June). Cbvmr: content-based video-music retrieval using soft intra-modal structure constraint. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval* (pp. 353-361). ACM.

Kohut, H., & Levarie, S. (2018). On the enjoyment of listening to music. In *The search for the self* (pp. 135–158). Routledge. doi:10.4324/9780429483097-4

Lero, R. D., Exton, C., & Le Gear, A. (2019, October). Communications using a speech-to-text-to-speech pipeline. In 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) (pp. 1-6). IEEE.

Liang, L., & Liu, J. (2021, April). An exploration of the application of computer music production software in music composition. In 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC) (pp. 794-796). IEEE.

Murthy, Y. S., & Koolagudi, S. G. (2018). Content-based music information retrieval (cb-mir) and its applications toward the music industry: A review. ACM Computing Surveys, 51(3), 1–46.

Ojanen, M. (2020). User Stories of Erkki Kurenniemi's Electronic Musical Instruments, 1961–1978. Academic Press.

Prince, J. B., & Sopp, M. (2019). Temporal expectancies affect accuracy in standard-comparison judgments of duration, but neither pitch height, nor timbre, nor loudness. *Journal of Experimental Psychology. Human Perception and Performance*, 45(5), 585.

Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S. Y., & Sainath, T. (2019). Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, *13*(2), 206–219.

Stewart, J., Garrido, S., Hense, C., & McFerran, K. (2019). Music use for mood regulation: Self-awareness and conscious listening choices in young people with tendencies to depression. *Frontiers in Psychology*, *10*, 1199. doi:10.3389/fpsyg.2019.01199 PMID:31178806

Sturm, B. L. (2013). The GTZAN dataset: Its contents, its faults, their effects on evaluation, and its future use. arXiv preprint arXiv:1306.1461.

Vishnupriya, S., & Meenakshi, K. (2018, January). Automatic music genre classification using convolution neural network. In 2018 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-4). IEEE.

Wang, C., Benetos, E., Lostanlen, V., & Chew, E. (2019, November). Adaptive time-frequency scattering for periodic modulation recognition in music signals. *International Society for Music Information Retrieval Conference*.

Wei, Z., Xiaolong, L., Jin, Z., Xueyun, W., & Hongxing, L. (2018). Foetal heart rate estimation by empirical mode decomposition and MUSIC spectrum. *Biomedical Signal Processing and Control*, 42, 287–296.

Zhao, H., Gan, C., Ma, W. C., & Torralba, A. (2019). The sound of motions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1735-1744). IEEE.

Zhu, F., Yang, J., Gao, J., & Xu, C. (2016). Extended nearest neighbor chain induced instance-weights for SVMs. *Pattern Recognition*, 60, 863–874.

Zhu, F., Yang, J., Xu, S., Gao, C., Ye, N., & Yin, T. (2017). Incorporating neighbors' distribution knowledge into support vector machines. *Soft Computing*, 21(21), 6407–6420.