The following publication Y. Jiang and F. H. F. Leung, "Mobile phone identification from speech recordings using Weighted Support Vector Machine," IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 2016, pp. 963-968 is available at https://doi.org/10.1109/IECON.2016.7793279.

Mobile Phone Identification From Speech Recordings Using Weighted Support Vector Machine

Yuechi Jiang
Centre for Signal Processing
Dept of EIE, PolyU
Hung Hom, Hong Kong
yuechi.jiang@connect.polyu.hk

Frank H. F. Leung
Centre for Signal Processing
Dept of EIE, PolyU
Hung Hom, Hong Kong
frank-h-f.leung@polyu.edu.hk

Abstract—**In this paper, we propose a mobile phone identifier called Weighted Support Vector Machine with Weighted Majority Voting (WSVM-WMV) for a closed-set mobile phone identification task. The proposed WSVM-WMV can be regarded as a generalization of the traditional SVM identifier. On using Mel-frequency Cepstral Coefficients (MFCC) and Linear-frequency Cepstral Coefficients (LFCC) as the feature vectors, the proposed identifier can improve the identification accuracy from 92.42% to 97.86% and from 90.44% to 98.33% respectively, as compared with the traditional SVM identifier in identifying a set of 21 mobile phones.**

Keywords—audio forensics, mobile phone identification, Weighted Support Vector Machine, Weighted Majority Voting

I. Introduction

Recently audio forensics has drawn much attention, where audio authenticity or integrity needs to be verified. With the fast development of the internet, more and more data are stored in the form of audios. Sometimes it is necessary to validate the authenticity of the audio signal in terms of privacy, recording conditions or ownership. There are many aspects of audio authentication, such as audio recording environment authentication, recording device authentication and recording date authentication [1]. In this paper, we focus on a closed-set mobile phone identification task, which is a kind of recording device authentication. Recording device authentication can help verify or identify the recording location and ownership, and this information can be used as court evidence in some situations [1]. The objective of our method is to identify the underlying recording mobile phones of the speech recordings from a set of candidate mobile phones, based on some training speeches recorded using these candidate mobile phones.

Recording device authentication comprises a feature extraction process and an identification process. In the literature, different feature extraction methods and identification methods have been developed and tested. Generally, the Mel-frequency Cepstral Coefficients (MFCC) is the most widely used frame-level feature vector, and Support Vector Machine (SVM) is widely used as the identifier. For example, in [2], a Gaussian supervector (GSV) was used to store the feature information, and a linear SVM was used as the identifier to identify 8 microphones and 8 handsets. The GSV was obtained by stacking the mean vectors of the Gaussian Mixture Model (GMM) constructed from MFCCs. In [3], MFCC, Linear Prediction Cepstral Coefficients (LPCC) and Perceptual Linear Predictive Coefficients (PLPC) as the features were compared, and GMM was employed as the identifier to identify a set of 16 microphones. The results show that LPCC outperforms other features. In [4], the authors used Power-normalized Cepstral Coefficients (PNCC) and MFCC as the features and employed the Gaussian Mixture Model-Universal Background Model (GMM-UBM) to identify a set of 14 cell phones. The results show that MFCC outperforms PNCC a lot, and MFCC with the energy coefficient performs better than that without the energy term. In [5], the authors used MFCC as the features and compared the performances of using Vector Quantization (VQ) against using SVM with the Generalized Linear Discriminant Sequence (GLDS) kernel as the identifier to identify a set of 14 cell phones. Experimental results show that GLDS-SVM outperforms VQ. In [6], the authors used GLDS-SVM as the identifier and compared the performances of using different features, including MFCC, Linear-frequency Cepstral Coefficients (LFCC), Bark-frequency Cepstral Coefficients (BFCC) and LPCC, in identifying a set of 14 cell phones. The experimental results show that the raw MFCC performs better than the others; however, when using the feature normalization techniques, LPCC can outperform other features.

Instead of using MFCC directly, there are also some sparse representation-based features, such as Random Spectral Features (RSFs) [7] and Sketches of Spectral Features (SSFs) [8], which are based on MFCC. Experimental results show that the sparse representation-based features outperform MFCC in identifying a set of 8 telephone handsets from the Lincoln-Labs Handset Database (LLHDB). In [9], the authors used the sparse decomposition of Gaussian mean-shift supervector as the features, and employed cosine metric to verify a set of 14 cell phones. The Gaussian mean-shift supervector is also based on MFCC. Besides using the whole audio sample to extract features, in [10]-[12], non-speech segments or estimated noise was also used to extract features.

The above works have two potential limitations. First, only one feature vector is obtained from an audio sample. However, one feature vector may not reveal the device information well. Multiple feature vectors generated from one audio sample could be considered, which may better characterize the

The work described in this paper was substantially supported by a grant from The Hong Kong Polytechnic University (Projects Account Code: RUG7 and G-YN19).

underlying device. Second, although SVM often exhibits good performance, the aforementioned SVM is the traditional one that assumes each data sample is of the same importance. However, as shown in [13], the traditional SVM is affected by outlier data, which results in the so-called outlier sensitivity problem.

To tackle these limitations, we propose an identifier called Weighted Support Vector Machine with Weighted Majority Voting (WSVM-WMV) in this paper. It generates multiple feature vectors from one audio sample, and then assigns different weights to different feature vectors during training and testing. In fact, WSVM-WMV can also be regarded as a generalization of the traditional SVM identifier.

This paper is organized as follows. In Section II, the ways of extracting MFCC and LFCC are briefly described, and the dataset used in our experiment is discussed. In Section III, the proposed WSVM-WMV identifier is detailed, and different weighting functions are explained. In Section IV, experimental results are presented and discussed. A conclusion will be drawn in Section V.

II. FEATURE EXTRACTION AND DATASET DESCRIPTION

A. Frame-level Feature Extraction

In this paper, 24-dimension MFCC, 25-dimension MFCC (which has an extra energy coefficient), 36-dimension LFCC and 37-dimension LFCC (which has an extra energy coefficient) are used as the frame-level feature vectors. MFCC and LFCC have been shown to be competent to characterize the recording mobile phones [5] [6].

Generally, frame-level MFCC are calculated as detailed in [14]. For calculating the 24-dimension MFCC, we first window the time-domain signal into short-time frames using Hamming window with 10ms frame length and 10ms frame shift. Then, calculate the Discrete Fourier Transform (DFT) of the short-time signal and apply Mel-scale filtering using 48 triangular filters on the DFT coefficients. Finally, we take the logarithm and then perform Discrete Cosine Transform (DCT) on the 48-dimension Mel coefficients, and extract the first 24 DCT coefficients (starting from the second coefficient as the first coefficient is a DC component) to form the 24-dimension MFCC vector. The 25-dimension MFCC is obtained by adding an extra log-energy term to the beginning of the 24-dimension MFCC. The 36-dimension LFCC can be obtained in a way similar to MFCC, where we use 72 triangular filters in linear-scale to filter the DFT coefficients and finally extract the first 36 DCT coefficients (starting from the second coefficient) to form the 36-dimension LFCC. The 37-dimension LFCC is obtained by adding an extra energy term. It should be remarked that Mel-scale filters (for MFCC) are specifically designed for speech signal, while linear-scale filters (for LFCC) are more general.

B. Brief Description of Dataset

Our dataset is modified from MOBIPHONE [15], which contains speech samples recorded using 21 mobile phones of different models as shown in Table I. MOBIPHONE comprises speech recordings from 12 male and 12 female speakers randomly chosen from TIMIT corpora [15]. A total of 504 different speeches are then recorded by the 21 mobile phones in the same environment with the same setup. For each mobile phone, each speaker contributes one speech recording lasting for about 30s. Therefore, each mobile phone contains 24 speech recordings. While the original MOBIPHONE contains only 504 samples, we divide each speech sample into 10 pieces so that we have a larger dataset of 5040 speech samples, and each sample is about 1s ~ 6s. Then for each mobile phone, 120 speech samples coming from 12 randomly chosen speakers are used to form parts of the training set, while the other 120 speech samples are used to form parts of the testing set. Hence, both the training set and testing set contain 2520 speech samples.

TABLE I. MOBILE PHONE MODELS

Class	Mobile Phone Model	Class	Mobile Phone Model
1	Apple iPhone5	12	Samsung E2121B
2	HTC Desire C	13	Samsung E2600
3	HTC Sensation XE	14	Samsung Galaxy GT-I9100 S2
4	LG GS290	15	Samsung Galaxy Nexus S
5	LG L3	16	Samsung GT-I8190 Mini
6	LG Optimus L5	17	Samsung GT-N7100 (Galaxy Note2)
7	LG Optimus L9		
8	Nokia 5530	18	Samsung S5830i
9	Nokia C5	19	Sony Ericsson C510i
10	Nokia N70	20	Sony Ericsson C902
11	Samsung E1230	21	Vodafone Joy 845

III. WEIGHTED SUPPORT VECTOR MACHINE WITH WEIGHTED MAJORITY VOTING

In this paper, we propose to use Weighted Support Vector Machine (WSVM) to construct the model that can discriminate different mobile phones. We also employ a Weighted Majority Voting (WMV) strategy to do the mobile phone identification (during testing). In the following, we will introduce WSVM and describe different weighting functions to be used. Finally, we describe the WMV strategy.

A. WSVM and Multiple Feature Generation

As discussed in [13], the traditional SVM aims to minimize the cost function as given by (1) below subject to some constraints, while WSVM aims to minimize the cost function as given by (2) below subject to some constraints,

$$\Phi(\omega) = \frac{1}{2}\omega^T \omega + C\sum_{i=1}^{N} \varepsilon_i \tag{1}$$

$$\Phi(\omega) = \frac{1}{2}\omega^T \omega + C\sum_{i=1}^{N} W_i \varepsilon_i \tag{2}$$

where ω is the coefficient defining the separating hyperplane, C is a constant, N is the number of training data (N=2520 in this paper), ε_i denotes how much the i-th training data violates the constraints, and W_i is the weight assigned to the i-th training data.

The second term in (1) and (2) is a penalty term. It can be seen that on using SVM, C is a constant such that all the training data are treated equally; while on using WSVM, there is an extra weight term W_i. The larger the value of this weight, the more the training data will be penalized, and thus the more important the training data will be. Thus, W_i reflects the importance and contribution of the i-th training data in finding the hyperplane parameter. Since in the proposed WSVM-WMV identifier, multiple feature vectors are generated from one audio sample, (2) is extended as follows.

$$\Phi(\omega)=\frac{1}{2}\omega^T\omega+C\sum_{i=1}^{N}\sum_{j=1}^{K}F(S_{ij})\varepsilon_{ij} \tag{3}$$

where $F(S_{ij})$ is the weight assigned to the feature vector S_{ij} and K is the number of feature vectors generated from one audio sample, as summarised in Table II.

The advantage of using WSVM in training is that different training feature vectors can be assigned with different weights, so that their contributions to build the model are emphasized differently. From Fig. 1, we can see that on using the traditional SVM identifier, the frame-level feature vectors are averaged to form a single feature vector. However, on using the WSVM-WMV identifier, multiple feature vectors are generated by extracting K mean vectors from a K-component Gaussian Mixture Model (GMM). If K=1, WSVM-WMV becomes the traditional SVM. A K-component GMM is obtained by splitting a 1-component GMM step by step, using the mixture splitting technique [16] and the Expectation-Maximization (EM) algorithm [17]. The mixture splitting technique is used to split the Gaussian mixture component with the largest weight in the current GMM into two new mixture components, and the EM algorithm is then applied to retrain the new GMM. In this way, the number of mixture components in the GMM is increased step by step until the desired number of mixture components is achieved and the corresponding weights (W_{ij}) are defined. This technique guarantees that the same GMM parameters are obtained every time for the same number of mixture components, because the parameters are not initialized randomly. Hence, the same set of feature vectors are obtained for the same number of mixture components. In this paper, the results under different values of K will be evaluated. It is remarked that the value of T in Fig. 1 depends on the length of the audio sample, which is not fixed.

TABLE II. NOTATIONS AND PARAMETERS

Notation	Explanation
K	Number of Gaussian mixture components
N	Number of training and testing audio samples
i	Index of audio sample, and $i\in\{1,2\ldots N\}$
j	Index of Gaussian mixture component, and $j\in\{1,2\ldots K\}$
S_{ij}	Feature vector corresponding to the j-th Gaussian mixture component of audio sample i
W_{ij}	Weight of the j-th Gaussian mixture component in the GMM of audio sample i
E_{ij}	Energy coefficient of S_{ij}
$F(S_{ij})$	Weight of S_{ij}
α	A positive scaling factor

B. Weighting Functions used in WSVM and WMV

On using the WSVM-WMV identifier, a crucial point is how to assign different weights to different feature vectors in order to make the identifier work efficiently. A simple choice is to use the weights of different Gaussian mixture components in the GMM. Besides, as inspired by [10]-[12] where low-energy segments were used to extract features, we design several energy-related weighting functions in such a way that the weight decreases as the energy of the feature vector increases. In the following, 8 different weighting functions are described. In some of these weighting functions, α (as defined in Table II) is a positive scaling factor used to satisfy the constraint given by (4) below.

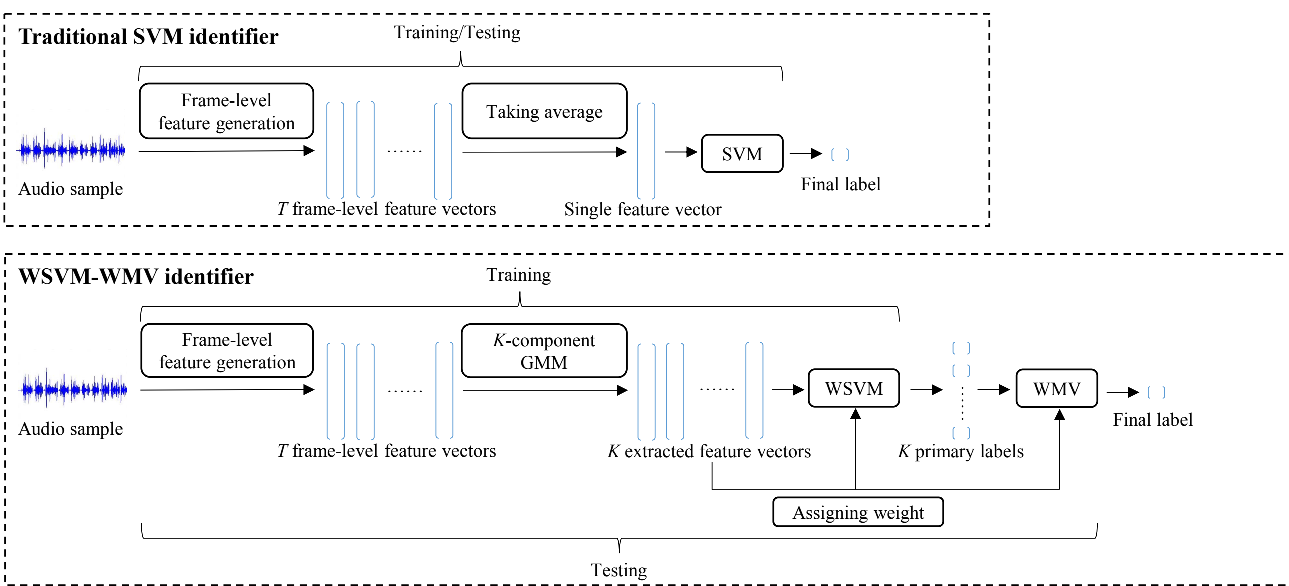

Fig.1. An overview of the traditional SVM identifier and the proposed WSVM-WMV identifier.

$$\sum_{j=1}^{K} F(S_{ij}) = K \tag{4}$$

a) All-one weighting function:

$$F(S_{ij}) = 1 \tag{5}$$

b) Gaussian weighting function:

$$F(S_{ij}) = KW_{ij} \tag{6}$$

c) Linear weighting function I

$$F(S_{ij}) = \alpha \times (\max\{E_{ij}\}_{j=1}^{K} - E_{ij}) \tag{7}$$

d) Linear weighting function II: Reorder S_{ij} according to E_{ij} for j=1, 2...K in ascending order. Let S'_{ij} be the reordered S_{ij} and E'_{ij} be the corresponding energy coefficient, we then have $E'_{i1} \le E'_{i2} \le \ldots \le E'_{iK}$. Hence, S_{ij} is a permutation of S'_{ij} for j=1, 2…K, and the mapping between S_{ij} and S'_{ij} is recorded. Then

$$F(S'_{ij}) = \alpha \times (K + 1 - j) \tag{8}$$

Finally, $F(S'_{ij})$ is mapped to $F(S_{ij})$ according to the mapping between S'_{ij} and S_{ij} .

e) Linear weighting function III: This weighting function is similar to *d)* except that the weight of S_{ij} whose energy is the highest is dumped.

$$F(S'_{ij}) = \alpha \times (K - j) \tag{9}$$

f) Sigmoid weighting function:

$$F(S_{ij}) = \frac{\alpha}{1 + \exp(E_{ij} - \frac{1}{K}\sum_{j=1}^{K} E_{ij})} \tag{10}$$

g) Exponential weighting function

$$F(S_{ij}) = \alpha \times \exp(-\frac{E_{ij}}{2}) \tag{11}$$

h) Circular weighting function

$$F(S_{ij}) = \alpha \times [R^2 - (E_{ij} - \min\{E_{ij}\}_{j=1}^{K})^2]^{\frac{1}{2}} \tag{12}$$

where

$$R = \max\{E_{ij}\}_{j=1}^{K} - \min\{E_{ij}\}_{j=1}^{K} \tag{13}$$

Function *a)* simply emphasizes different feature vectors equally. Function *b)* aims to set the weights for different S_{ij} to be the same as their weights in the GMM. Functions *c) ~ h)* set the weights according to the energies of different S_{ij} such that the weight decreases as the energy of S_{ij} increases. Functions *c) ~ e)* have the decreasing speed of the weights being steady, as the functions are linear and the gradient keeps unchanged. Functions *f) ~ h)* have varying decreasing speeds, as they are non-linear and the gradient is changing. For function *f)*, the decreasing speed first increases and then decreases, as the energy of S_{ij} increases. For function *g)*, the decreasing speed decreases as the energy of S_{ij} increases. For function *h)*, the decreasing speed increases as the energy of S_{ij} increases. It should be noted that functions *d)* and *e)* map the continuous energy to a discrete set; for example, in function *d)*, the weights can only be some value from the set {α, 2α …$K\alpha$}.

Different weighting functions with different decreasing speeds have been designed so that the relationship between the weights and the energies of different feature vectors can be explored comprehensively. In particular, on using functions *c)*, *e)* and *h)*, the S_{ij} with the highest energy is dumped because high-energy feature vectors may contain more human speech information or speaker information, which may conceal the device information. Since functions *c) ~ h)* contain the energy parameter, they are only applicable to the features that include the energy coefficient.

C. *Weighted Majority Voting Strategy*

In WSVM-WMV, WSVM can be trained by feeding the training feature vector S_{ij} together with the weight $F(S_{ij})$ to the identifier; however, in using WSVM to conduct the identification, there is no "weight" concept. Thus, we propose a WMV strategy to assist the identification such that different testing feature vectors can also be assigned with different weights effectively to indicate their relative importance in making decisions.

Suppose there are totally C classes (C=21 in this paper), and within a group of K testing feature vectors generated from the testing audio sample i, S_{ij} is classified to class P_{ij}, where P_{ij} ∈ {1, 2…C}. Then the final identified class T_i of the testing audio sample i is computed as:

$$T_i = \arg\max_t \sum_{j=1}^{K} F(S_{ij})\gamma(t, P_{ij}) \tag{14}$$

where

$$\gamma(t, P_{ij}) = \begin{cases} 1, & \textit{if } t = P_{ij} \\ 0, & \textit{otherwise} \end{cases} \quad \textit{and} \quad t \in \{1, 2 \ldots C\} \tag{15}$$

In the above, (14) is the WMV strategy proposed to assist the identification. It can be seen that different testing feature vectors have different contributions in making the final decision. The identification accuracy is then the ratio between the number of correctly identified audio samples and the total number of testing audio samples.

IV. Experimental Results and Discussions

The experiment is divided into two parts. In the first part, the performances of the SVMs with and without GMM (i.e. single feature versus K extracted features) are compared. In the second part, the performances of the WSVM-WMV identifier using different weighting functions are evaluated. The SVMs and WSVM-WMV are implemented using LIBSVM [18] under the Radial Basis Function (RBF) kernel, the same parameters and the same stopping criteria. In the following, the feature vectors with the extra energy coefficient are denoted by MFCC_E and LFCC_E.

A. Comparison of SVMs with and without GMM

In this part, the weighting function is the all-one weighting function. Then, when K=1, WSVM-WMV becomes the traditional SVM identifier without the GMM; thus, WSVM-WMV is a generalization of the traditional SVM. Different values of K (i.e. the number of components in the GMM, or the number of feature vectors generated from one audio sample) are evaluated to explore how it can influence the performance of the identifier. The identification accuracy results are shown in Table III and plotted in Fig. 2. From Fig. 2, it can be seen that on increasing the value of K, the performance is improved and the identifier with GMM outperforms the traditional SVM without GMM, no matter which kind of features are used. It is worth noting that features with the energy coefficient tend to perform better than those without the energy coefficient. When K is large enough, the performance tends to saturate.

TABLE III. Comparison of Identifiers With and Without GMM on Identification Accuracy (%)

Identifier	K	Feature			
		MFCC-based		LFCC-based	
		MFCC	MFCC_E	LFCC	LFCC_E
Traditional SVM	1	91.63	92.42	90.32	90.44
SVM with GMM	1	91.63	92.42	90.32	90.44
	10	95.32	95.40	94.96	95.91
	20	**96.15**	**96.39**	96.43	97.18
	30	**96.15**	96.19	**97.02**	**97.46**

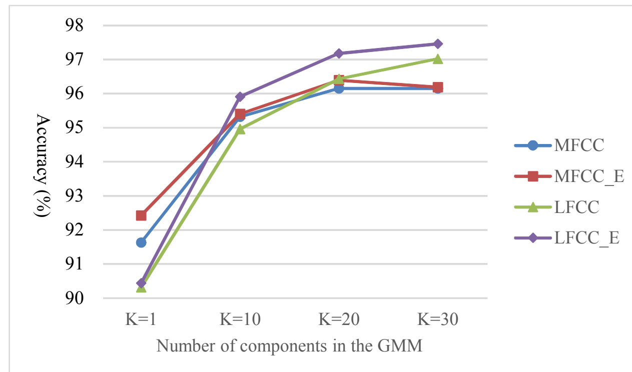

Fig.2. Identifier with GMM under the all-one weighting function.

B. WSVM-WMV with Various Weighting Functions

In this part, different weighting functions used in WSVM-WMV are evaluated. The identification accuracy results using different features and different weighting functions are listed in Table IV, where the highest accuracies for different features

TABLE IV. WSVM-WMV with Different Weighting Functions (%)

K	Feature	Weighting Function							
		a) All-one	b) Gaussian	c) Linear I	d) Linear II	e) Linear III	f) Sigmoid	g) Exponential	h) Circular
10	**MFCC_E**	95.40	94.13	97.10	97.38	**97.54**	97.38	95.56	96.83
	LFCC_E	95.91	93.97	97.14	**97.22**	**97.22**	97.10	95.40	96.75
20	**MFCC_E**	96.39	94.64	97.38	**97.46**	97.38	97.34	95.52	97.10
	LFCC_E	97.18	94.52	97.74	**97.82**	97.70	97.50	96.07	97.62
30	**MFCC_E**	96.19	95.04	97.54	97.66	**97.86**	97.62	95.95	96.75
	LFCC_E	97.46	94.52	**98.33**	98.21	98.17	98.10	96.35	97.86

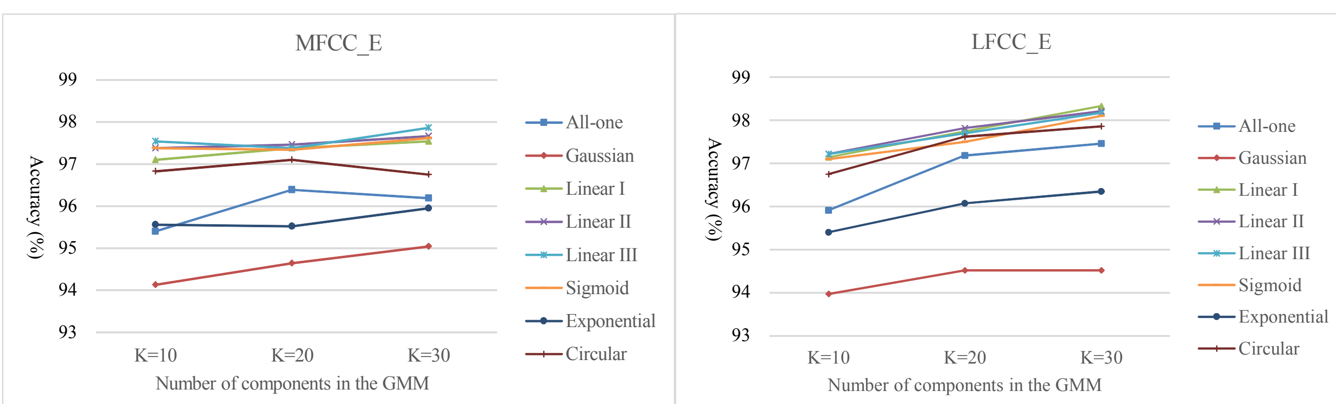

Fig.3. WSVM-WMV with various weighting functions.

under different values of K are emphasized in boldface. Illustrative graphs are plotted in Fig. 3.

From Fig. 3, it can be seen that Gaussian weighting function performs worse than the all-one weighting function, because a larger Gaussian weight merely indicates a higher frequency of occurrence but not a higher importance of the mixture component. The low efficiency of Gaussian weighting function has motivated us to explore the energy-related weighting functions *c) ~ h)*. It can be seen that, except the exponential weighting function, energy-related weighting functions can improve the performance over the all-one weighting function, which indicates the possibility of using energy as the weight assignment criterion.

The exponential weighting function performs worse than other weighting functions. On designing the exponential weighting function, the output is directly related to the absolute energy values rather than the relative energy values, which makes it quite sensitive to small variations of the energy coefficients. Other energy-related weighting functions consider only the relative values of the energy coefficients, resulting in the weights being less sensitive to small variations of the energy values. In particular, functions *d)* and *e)* map the continuous energy to a discrete set of weights. They are even more robust (less sensitive) to energy variations.

V. Conclusion

In this paper, we propose a Weighted Support Vector Machine with Weighted Majority Voting (WSVM-WMV) identifier for a closed-set mobile phone identification task. The important characteristics of this identifier lies in that multiple feature vectors are generated from one audio sample, and different feature vectors can be emphasized differently according to a weighting function. The advantages of using this identifier over the traditional SVM or other identifiers are: 1) in this identifier, more feature vectors are generated and different feature vectors can be of different importance, so that the identifier can be better trained; 2) identification is based on a WMV strategy, which is more robust (less sensitive) to small portions of misidentified testing feature vectors.

The proposed WSVM-WMV can be regarded as a generalization of the traditional SVM identifier, where the latter is just a special case of the former with K=1. To realise different importance for different feature vectors in the WSVM-WMV identifier, we design some energy-related weighting functions, which show better performance over the simple all-one weighting function and Gaussian weighting function.

However, WSVM-WMV requires more computation time than the traditional SVM. The reason lies in two aspects. First, GMM is used to generate multiple feature vectors from an audio sample, and constructing a GMM is time-consuming owing to the EM algorithm. Second, more feature vectors are fed into the identifier, which makes the identifier take more time to do the identification.

The efficiency of WSVM-WMV indicates that on doing recording device identification, it is better to generate multiple feature vectors to comprehensively describe an audio sample, and let the feature vectors to cooperate with each other during the identification.

References

[1] S. Gupta, S. Cho, and C. C. J. Kuo, "Current developments and future trends in audio authentication," *IEEE Multimedia*, vol. 19, pp. 50-59, Jan. 2012.

[2] D. Garcia-Romero and C. Y. Espy-Wilson, "Automatic acquisition device identification from speech recordings," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'10)*, Dallas, TX, USA, 2010, pp. 1806-1809.

[3] O. Eskidere, "Source microphone identification from speech recordings based on a Gaussian mixture model," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 22, pp. 754-767, 2014.

[4] L. Zou, J. Yang, and T. Huang, "Automatic cell phone recognition from speech recordings," in *China Summit & International Conference on Signal and Information Processing (ChinaSIP),* Xi'an, China, 2014, pp. 621-625.

[5] C. Hanilci *et al*., "Recognition of brand and models of cell-phones from recorded speech signals," *IEEE Trans. Information Forensics and Security*, vol. 7, no. 2, pp. 625-634, 2012.

[6] C. Hanilci and F. Ertas, "Optimizing acoustic features for source cell-phone recognition using speech signals," in *Proc. First ACM Workshop on Information Hiding and Multimedia Security*, 2013, pp. 141-148.

[7] Y. Panagakis and C. Kotropoulos, "Automatic telephone handset identification by sparse representation of random spectral features," in *Proc. Multimedia and Security*, 2012, pp. 91-96.

[8] C. Kotropoulos, "Telephone handset identification using sparse representations of spectral feature sketches," in *Proc. Int. Workshop on Biometrics and Forensics (IWBF)*, Lisbon, Portugal, 2013, pp. 1-4.

[9] L. Zou, Q. He, and X. Feng, "Cell phone verification from speech recordings using sparse representation," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'15),* South Brisbane, QLD, Australia, 2015, pp. 1787-1791.

[10] M. Jahanirad *et al.*, "Blind source mobile device identification based on recorded call," *Engineering Applications of Artificial Intelligence*, vol. 36, pp. 320-331, 2014.

[11] C. Hanilci and T. Kinnunen, "Source cell-phone recognition from recorded speech using non-speech segments," *Digital Signal Processing*, vol. 35, pp. 75-85, 2014.

[12] R. Aggarwal *et al.*, "Cellphone identification using noise estimates from recorded audio," in *Proc. IEEE Int. Conf. on Communications and Signal Processing (ICCSP)*, Melmaruvathur, India, 2014, pp. 1218-1222.

[13] X. Yang, Q. Song, and Y. Wang. "A weighted support vector machine for data classification." *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 21, no. 5, pp. 961-976, 2007.

[14] X. Huang, A. Acero and H.-W. Hon, "Speech Signal Representations," in *Spoken Language Processing: A Guide to Theory, Algorithm and System Development.* Upper Saddle River, NJ: Prentice Hall PTR, 2001, ch 6, pp. 273-333.

[15] C. Kotropoulos and S. Samaras, "Mobile phone identification using recorded speech signals," in *Proc. Int. Conf. on Digital Signal Processing (DSP)*, Hong Kong, 2014, pp. 586-591.

[16] S. Young *et al*., *The HTK book (v3.4),* Cambridge: Cambridge University Press, 2006, pp. 156-157.

[17] D. Reynolds, "Gaussian mixture models," in *Encyclopedia of Biometrics*, Springer US, 2009, pp. 659-663.

[18] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1-27, 2011.