Engineering Applications of Artificial Intelligence 36 (2014) 320–331

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Blind source mobile device identification based on recorded call

Mehdi Jahanirad *, Ainuddin Wahid Abdul Wahab, Nor Badrul Anuar, Mohd Yamani Idna Idris, Mohd Nizam Ayub

Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 7 April 2014
Received in revised form
10 July 2014
Accepted 11 August 2014
Available online 16 September 2014

Keywords:
Pattern recognition
Mel-frequency cepstrum coefficient
Entropy
Device-based detection technique

ABSTRACT

Mel-frequency cepstrum coefficients (MFCCs) extracted from speech recordings has been proven to be the most effective feature set to capture the frequency spectra produced by a recording device. This paper claims that audio evidence such as a recorded call contains intrinsic artifacts at both transmitting and receiving ends. These artifacts allow recognition of the source mobile device on the other end through recording the call. However, MFCC features are contextualized by the speech contents, speaker's characteristics and environments. Thus, a device-based technique needs to consider the identification of source transmission devices and improve the robustness of MFCCs. This paper aims to investigate the use of entropy of Mel-cepstrum coefficients to extract intrinsic mobile device features from near-silent segments, where it remains robust to the characteristics of different speakers. The proposed features are compared with five different combinations of statistical moments of MFCCs, including the mean, standard deviation, variance, skewness, and kurtosis of MFCCs. All feature sets are analyzed by using five supervised learning techniques, namely, support vector machine, naïve Bayesian, neural network, linear logistic regression, and rotation forest classifier, as well as two unsupervised learning techniques known as probabilistic-based and nearest-neighbor-based algorithms. The experimental results show that the best performance was achieved with entropy–MFCC features that use the naïve Bayesian classifier, which resulted in an average accuracy of 99.99% among 21 mobile devices.

1. Introduction

Audio forensics has recently received considerable attention because it can be applied in different situations that require audio authenticity and integrity (Kraetzer et al., 2012). Such situations include forensic acquisition, analysis, and evaluation of admissible audio recordings as crime evidence in court cases. Digital audio technology development has facilitated the manipulation, processing, and editing of audio by using advanced software without leaving any visible trace. Thus, basic audio authentication techniques, such as listening tests and spectrum analysis, are easy to cross over. Authenticity of audio evidence is important as part of a civil and criminal law enforcement investigation or as part of an official inquiry into an accident or other civil incidents. In these processes, authenticity analysis determines whether the recorded information is original, contains alterations, or has discontinuities attributed to recorder stops and starts. Current approaches used to define audio recording authenticity are based on artifacts extracted from signals, which consist of (a) frequency spectra introduced by the recording environment (i.e., environment-based techniques), (b) frequency spectra produced by the recording device (i.e., device-based techniques), and (c) frequency spectra generated by the recording device power source (i.e., ENF-based techniques) (Maher, 2009). The performance of environment-based techniques (AlQahtani and Al Mazyad, 2011; Muhammad and Alghathbar, 2013) depends on the presence and the amount of foreground speech, background noise and environmental reverberations. Moreover, advanced audio forgery software counterfeit the environmental effects without leaving any trace in the original file, which is a disadvantage of the environment-based techniques. Although ENF-based techniques (Cooper, 2011; Ode Ojowu et al., 2012) provide high accuracy and novelty, they have limitations because ENF is only sometimes embedded in the recordings. A special case is when the appliance is battery powered and locates outside the coverage of the electromagnetic field that is generated from the electric network. Even if the ENF pattern is detectable on the audio evidence, this method requires the ENF archive that is only available for limited areas.

In terms of device-based techniques, previous works focused on identifying the source of the recording devices. Nevertheless, the potential of the device-based techniques is hardly limited to this scope. Device-based techniques have been studied in three different directions: (a) identification of computer-generated

* Corresponding author.
E-mail addresses: mehdijahanirad@siswa.um.edu.my (M. Jahanirad), ainuddin@um.edu.my (A.W.A. Wahab), badrul@um.edu.my (N.B. Anuar), yamani@um.edu.my (M.Y.I. Idris), nizam_ayub@um.edu.my (M.N. Ayub).

http://dx.doi.org/10.1016/j.engappai.2014.08.008
0952-1976/

audio from the original audio recording file (Keonig and Lacey, 2012); (b) identification of the source brand, model, or individual acquisition devices that were used, such as telephone handsets, microphones (Garcia-Romero and Epsy-Wilson, 2010; Panagakis and Kotropoulos, 2012), and cell phones (Hanilçi et al., 2012); and (c) identification of the call origin to determine the network that was traversed (i.e., cellular, VoIP, and PSTN) and detailing the fingerprint of the call source, such as through a speech coder (Balasubramaniyan et al., 2010; Jenner, 2011; Sharma et al., 2010).

This paper focuses on the second branch of the device-based technique, namely, the identification of the brand, model and individual mobile devices used by using the audio recording. This particular focus makes this paper similar to studies on blind source camera identification by using digital images (Kharrazi et al., 2004; Celiktutan et al., 2008; Swaminathan et al., 2007). For example, Swaminathan, et al. (2007) defined feature-based image source camera identification as a blind method that can identify internal elements of a digital camera without having access to the camera. Although the methodology is similar, the adaptation of image forensic techniques in audio forensics introduces more challenges. These challenges are due to the different audio contents such as the human voices, footsteps and musical instruments. Secondly, the recorded audio may be a live record, playback record or call record. Hence, the audio signal processing chain may contain more than one acquisition devices that leave their intrinsic artifacts on the recorded audio. However, in terms of image forensics, usually one acquisition device (i.e. scanner, digital camera) generates the image.

This paper introduces source transmission device identification to identify the brand/model and individual mobile devices used where their VoIP conversation is received and recorded by a stationary device based on entropy–Mel-frequency cepstrum coefficient (MFCC) features. In addition, the suggested techniques can be extended to other types of communication devices and networks. A call recording signal is a mixture of an audio processed from a mobile device and transmitted to a stationary device with the audio processed from the stationary device and transmitted to the mobile device. This audio signal includes the intrinsic artifacts of its corresponding transmitting and recording ends. The mobile device artifacts are caused by its frequency response multiplied by the spectrum of the original audio signal and delivered through calls traverse cellular, PSTN and VoIP network. The frequency response of the acquisition devices is defined based on the fact that for different electronic components, every realization of an electric circuits produces specific transfer function (Hanilçi et al., 2012). Furthermore, the control environments are applied to the recording setup to perform silent, speechless conversation. This setup eliminates the influences by environments and the speakers. However because MFCC features are well known to model the context of the audio signal (e.g., speech), for silent recording we use the entropy of Mel cepstrum coefficients to intensify the energy of MFCCs. The experimental setup evaluates the feasibility of entropy–MFCC features against other statistical moments of MFCCs through different types of classifier and clustering techniques that are usually used in machine learning applications [e.g., support vector machine (SVM), naïve Bayesian, and logistic regression].

The remainder of this paper is organized as follows: Section 2 discusses related works. An overview of the source mobile device identification scheme is introduced in Section 3 and the overall methodology is described in Section 4. Section 5 details the experimental setup and results. Finally, Section 6 explains further implications of the practical study, its limitations, and future applications.

2. Related works

The process of identifying a device involves developing an efficient device identifier that determines the brand/type of the device used in processing the audio signal before recording. A considerable amount of research has been conducted on identifying the source recording devices, but to our knowledge, no forensic research has been performed to identify the transmission device from a recorded call. Most of the existing works in media forensics use multimedia data mining techniques as tools to identify the source devices. In general, multimedia data mining includes four steps: data preparation, feature extraction, feature analysis and decision making. However, no direct performance comparison is possible amongst existing methods. The reason is, methods and techniques adopted for each step differs with respect to the device type, the nature of the audio material, and the application scenario. Hence, Table 1 presents an overall performance of existing device-based techniques based on different audio features and machine learning techniques. This observation allows to identify the present state of the source acquisition device identification approaches in audio forensics, discover their contribution toward optimization and propose new directions.

As shown in Table 1, microphones, telephone headsets, and cell phones are the three main devices for identification. Previous works have mainly detected audio features in any of the following three categories: time domain, frequency domain, and Mel-

Table 1
Comparison of methods for device identification using audio.

Ref	*Classification algorithm*	*Devices*		*Recording signal*		*No. of features*			*Evaluation*		*ACC (%)*
		&Types	*No.*	*Non-speech*	*Speech*	*Time domain*	*Frequency domain*	*Mel-cepstrum domain*	*Inter-device*	*Intra-device*	
Kraetzer et al. (2007)	*Naïve Bayesian	M	4	✓	✓	7	–	56	✓		75.99
Buchholz et al. (2009)	*Linear Logistic Regression	M	7	✓		–	2048	–	✓		93.5
Kraetzer et al. (2011)	*Classifier – Benchmarking	M	4	✓	✓	9	529	52	✓	✓	82.51
Kraetzer et al. (2012)	*Rotation Forests	M	6		✓	9	529	52	✓	✓	99.85
Garcia-Romero and Epsy-Wilson (2010)	#Linear SVM	TH	8		✓	–	–	23	✓		93.2
		M	8		✓	–	–	23	✓		99
Hanilçi et al. (2012)	$SVM	CP	14		✓	–	–	24	✓	✓	96.42
Panagakis and Kotropoulos (2012)	#SVM	TH	8		✓	–	8	–	✓		97.58

Note: *The classification is performed by using 10-fold cross validation; # The classification is performed by using 2-fold cross validation, $ SVM-based classification by using GLDS kernel, & M: Microphone, TH: telephone handset, CP: cell phone.

cepstrum domain. Time domain features, such as zero-crossing rate and short-time energy ratio, are computationally light features, but they contain irrelevant data for classification (Ghosal et al., 2009). Frequency domain features are useful particularly in speech-processing field, for instance the linear signal variations induced by the transfer function of the speaker's vocal tract (Shen et al., 1996). However, they show deficiency when non-linear signal variations known as convolutions exist in speech signals (i.e. transfer function of the speaker's vocal tract and the microphone device) (Ye, 2004). Convolutions exist when convolved signals in the time domain are represented as a product of the Fourier transform function in the frequency domain. The Mel-cepstrum domain features such as MFCC, are of great concern because convolved signals are represented by a summation in the Mel-cepstrum domain (Beigi, 2011a). In terms of evaluation, all previous studies used an inter-device identification approach. However, intra-device identification requires more discriminant features and was only implemented in Kraetzer et al. (2011, 2012) and Hanilçi et al. (2012). Identification accuracy is widely used as a metric for evaluation of machine learning techniques. The majority of methods evaluate the performance of the classifiers by using identification accuracy (ACC). ACC is determined based on the total number of true positive (TP) and true negative (TN) classified instances to the total number of instances.

Kraetzer et al. (2007) published the first practical evaluation on microphone and environment classification. They used the combination of time-domain and Mel-cepstrum domain features in the classification by using Naïve Bayesian classifier. Buchholz et al. (2009) focused on microphone classification through the histogram of Fourier coefficients and extracted the coefficients from near-silent frames to capture microphone properties. Kraetzer et al. (2011) developed a suitable context model for microphone forensics by using the feature extraction tool in Kraetzer and Dittmann (2010). The model also utilised classifier benchmarking, and the audio files in Kraetzer et al. (2007) and Buchholz et al. (2009). This study developed towards better generalization, when Kraetzer et al. (2012) constructed a new application model for microphone forensic investigations with the aim of detecting playback recordings. In addition to microphone classification, Garcia-Romero and Epsy-Wilson (2010) proposed an automatic acquisition device identification by using landline telephone handsets and microphones. Panagakis and Kotropoulos (2012) proposed random spectral features and labeled spectral features (LSF) from speech recording for landline telephone handset identification. They also argued the robustness of these features over MFCCs and improved the accuracy of the identification by using sparse representation-based classification (SRC). Hanilçi, et al. (2012) proposed a cell phone identification method that uses MFCC features from speech recording in the classification by using SVM classifier with generalized linear discriminant sequence (GLDS) kernel.

The related works sometimes use time domain features in combination with frequency domain and Mel-cepstrum domain features (Kraetzer et al., 2007, 2011, 2012). However, Kraetzer et al. (2011) proved that time domain features contribute little to classification accuracy. In Buchholz et al. (2009)) and Panagakis and Kotropoulos (2012), high classification accuracy was achieved by using only frequency domain features. This finding proves that the performance of these features is almost identical to that of Mel-cepstrum domain features when convolved signals are eliminated. Buchholz et al. (2009) eliminated the convolution produced by speech signals by filtering Fourier coefficients above the near-silent threshold. Panagakis and Kotropoulos (2012) applied unsupervised and supervised feature selection to minimize the interference caused by signal variations attributed to speech signals. Mel-cepstrum features commonly perform best amongst existing methods. Because convolved signals are represented by the summation in Mel-cepstrum domain, they produce inherent invariance toward linear spectral distortions. MFCC features has been proven as the most effective features in cell phone identification (Hanilçi et al., 2012), however, Panagakis and Kotropoulos (2012) proved their lack of robustness in speech recording. This is because MFCC features are very well known to model the context of an audio recording but in the case of a speech recording, these features are contextualized by the speech characteristics.

This paper proposes a new method to eliminate the effects of speech contents by extracting the entropy of MFCC features from near-silent segments. Aside from reducing the dimensionality of MFCCs, we take advantage of the fact that entropy maximally concentrates the energy of MFCCs in silent frames because of its ability to measure the amount of choice or uncertainty (Nilsson, 2006). This approach accounts for the uncertainty over the data induced by the device frequency response. Hence, entropy captures mobile device specific features from Mel-cepstrum spectrum of the signal flow from source to recipient. The recipient is the stationary device that uses the wireless coverage to communicate through VoIP call to the different mobile devices and record the call. Thus, the recorded transmitting signal contains the intrinsic artifacts of the mobile devices. With this motivation, we propose a technique for source mobile device identification by using VoIP communication based on an entropy–MFCC feature set. This technique is extendable to real-time call recording scenarios by using speech removal algorithm prior to feature extraction.

3. Source mobile device identification scheme

The proposed scheme includes four sections. Section 3.1 discusses the data preparation step through precise preprocessing algorithms. Preprocessing increases the quality and quantity of the data instances to achieve high performance (Bhatt and Kankanhalli, 2011). Section 3.2 describes the feature extraction process including the computation steps of MFCC features and the entropy of MFCC features. Sections3.3 and 3.4 detail supervised and unsupervised learning methods that were implemented for classification and clustering, respectively. The complete steps of the proposed source mobile device identification scheme are shown in Fig. 1.

3.1. Data preparation

The audio files were recorded in two-channel stereo. To capture the nature of audio signals from both channels, we used channel's sum. The data preparation algorithm enhances the audio signals through minimum mean-square error (MMSE)-based noise power estimation approach, as proposed in Gerkmann and Hendriks (2012). Spectral enhancement aims to remove the non-stationary noise corruptions produced by the environment. This method was originally proposed for speech enhancement to reduce the additive noise without reducing speech ineligibility. The method uses speech presence probability (SPP) approach with fixed non-adoptive a priori SNR. This value is selected based on the SNR that is typical in speech presence. The key advantage of this enhancement technique is its low overestimation of the spectral noise power and an even lower computational complexity. Fig. 2 compares the power spectrum envelopes of the near-silent signals prior and after enhancement. The signals are recorded calls between the stationary device and two different units of Samsung Galaxy Note. The signal-to-noise (SNR) levels of the recordings prior to enhancement are 0.84 and 16.33 dB, respectively. Thus, the signals become more distinct when noise is eliminated. Afterwards, we split the enhanced signals into 40 overlapping short audio frames of length 2500 samples. The short audio frames are the data instances for feature extraction. The

experimental results in Section 5 will provide justification on the choices made during data preparation.

3.2. Feature extraction

The internal signal processing components of a mobile device, such as filter, sampler, A/D converter, encoder, and channel encoder, produce the overall transfer function of the device. The signal variations due to this transfer function leave their intrinsic artifacts on the audio delivered through VoIP network to a stationary device that records the received signal.

3.2.1. Motivation

We have suggested the use of near-silent segments for feature extraction. However, to prove our motivation for this suggestion, let's assume that the audio delivered to the recording stationary contains speech. The spectrum of the speech signal $x(t)$ represents the original input signal in the time domain, where t stands for time, τ is the time period and $h(t)$ denotes the device frequency response, thus the received signal from both sides of the call $y(t)$ is given by

$$y(t) = (x*h)(t) \triangleq \int_{-\infty}^{+\infty} x(\tau)h(t-\tau)d\tau. \quad (1)$$

Eq. (1) shows the impact of devices on the recorded speech is a convolutional distortion that allows for determining the intrinsic artifacts for identifying the devices. The Fourier transform of $x(t)$ and $h(t)$ in terms of the angular frequency ω is defined as $x(\omega)$ and $h(\omega)$ where the convolution of two functions in (1) is the product of their individual Fourier transforms and is represented as

$$Y(\omega) = F\{(x*h)(t)\} = X(\omega)H(\omega). \quad (2)$$

This indicates that each device leaves its intrinsic fingerprints on the overall recorded speech by modifying the spectrum of its corresponding speech signal. Hence, the method transfers the signal to cepstrum domain to eliminate the non-linearity in (2) by computing the logarithm of its Fourier transform. This can be as written as

$$\log(X(\omega)H(\omega)) = \log(X(\omega)) + \log(H(\omega)) = \hat{x}(t) + \hat{h}(t) \quad (3)$$

where the cepstrum of $x(t)$ and $h(t)$ denoted as $\hat{x}(t)$ and $\hat{h}(t)$. The feasibility of MFCCs in speech/speaker identification as well as acquisition device identification is due to this transformation.

Hanilçi, et al. (2012) proved that MFCCs are the most effective features for cell phone identification, while the cell phones recorded the input speech as an ordinary tape recorder to eliminate complexity. Alternatively, our recording setup eliminates the influences of the stationary device, speakers and speech content by transmitting silent sound from different mobile devices one at a time and recording the received signal with the same stationary device. However, in the absence of speech, the context of the signal is reduced as well as the values of MFCCs. Thus, we propose to use the entropy of the Mel-cepstrum output to intensify the energy of MFCCs. This is because the flat distribution of silence induces high entropy values.

The experiments in Hanilçi et al. (2012) achieved higher identification accuracy rates by logarithmic transformation of features in additive form, whereas in our work, the results in Table 5 of Section 5 shows that the identification accuracy rates for additive form compares well with the multiplicative form. It is evident that the absence of speech and the concentration on noise-like signals increase the robustness of the device identification approach. Furthermore, this work proves the feasibility of the entropy of MFCCs over normalized versions of MFCCs based on the

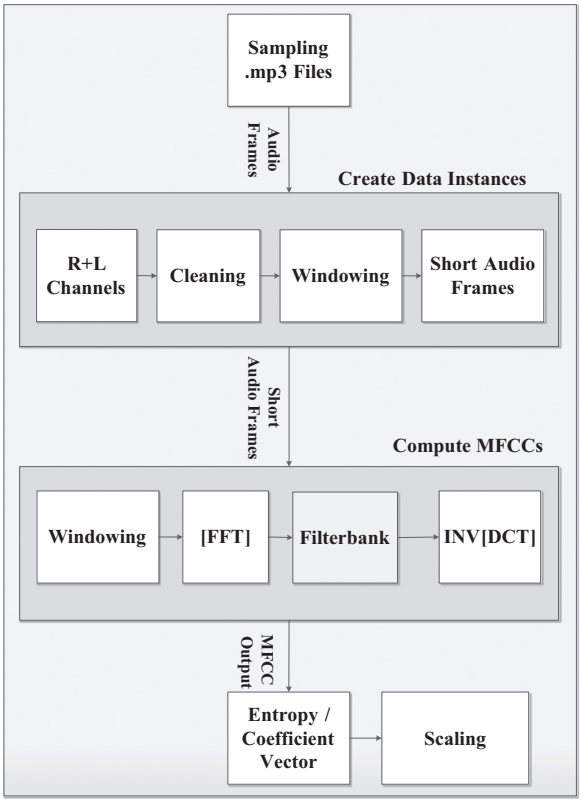

Fig. 1. Flowchart of the proposed data preparation and feature extraction approach.

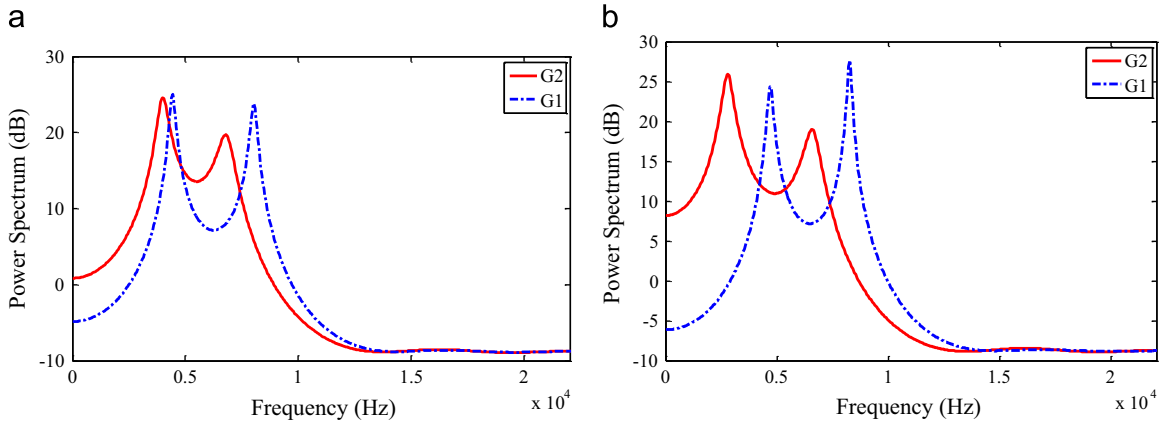

Fig. 2. Power spectrum envelopes of the signal in two mobile devices with the same model (clean vs. noisy signal). (a) Noisy signals. (b) Clean signals.

experiment results given in Table 4 of Section 5. The feature extraction algorithm includes three stages (a) computing the MFCCs, (b) the entropy of MFCCs, and (c) scaling the entropy–MFCC features.

3.2.2. MFCCs

MFCCs are one of the most attractive features of cepstrum domain and convey significant information about the structure of a signal. Thus, these features are widely used for speaker and speech recognition (Beigi, 2011a). The MFCCs are determined by computing the inverse discrete cosine transform of the short-time Mel frequency log spectrum of the signal, as given by

$$ {}_lc_n = \sum_{m=0}^{M-1} a_{ml}\, C_m \cos\left(\frac{\pi(2n+1)m}{2M}\right), \tag{4}$$

where ${}_lc_n$ is the nth MFCC coefficient for the lth frame, M is the total number of triangular filters for$m=\{1,\ldots,M\}$ filter coefficients in the filter bank, ${}_lC_m$ is the log spectrum output for the lth frame of the signal and the mth filter coefficients. In addition, the coefficients a_m are determined as

$$ a_m = \begin{cases} \frac{1}{M} & \text{for } m=0 \\ \frac{2}{M} & \forall\ m>0 \end{cases}. \tag{5}$$

The log energy (average log energy of audio frames), and the first and second derivative of MFCC coefficients could also be included in MFCC feature vector. However, preliminary results show less contribution of the log energy, as well as the first and second order cepstral coefficients in achieving the identification accuracy rates. We can consider these features in the future for larger number of training and testing data sets that are collected in real-time basis. Thus, in this work, we used 12 MFCCs, where the Mel-cepstrum output consists of one frame per row and each frame includes 12 coefficients.

3.2.3. Entropy

Entropy intensifies the energy of MFCC outputs. Fig. 3 illustrates the MFCC output ${}_lc_n$ and its entropy, where N is the total number of frames. The feature extraction approach computes the entropy of MFCC vectors in two stages; first it computes the probability mass function (PMF) of the MFCC coefficients and then, it computes the entropy H by using

$$ H_n = -\sum_{l=1}^{N} {}_lp_n \log_2 {}_lp_n. \tag{6}$$

where ${}_lp_n$ is the PMF of the nth MFCC coefficient in frame l (Beigi, 2011c).

3.2.4. Scaling

The last and important step after feature extraction is scaling. This step is important for increasing computational time and the classifier performance. In this work, we have scaled all data instances in the range [0, 1].

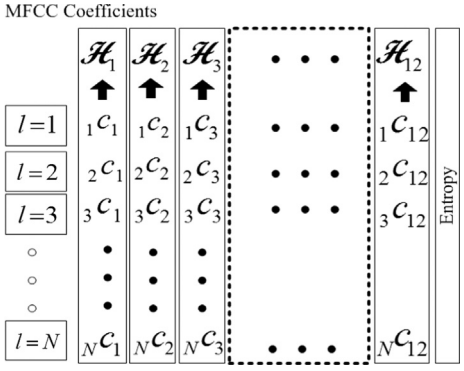

Fig. 3. Entropy–MFCC feature extraction steps.

After extracting the features, we can visualize the ability of features to differentiate between mobile device classes through histogram. The histogram in Fig. 4 visualizes the differences between mobile devices of different models by using the 12 entropy–MFCC features that were extracted from 1000 data instances from each mobile device. For further investigation, we selected four different pairs of mobile devices and examined the average squared Euclidean distance between their entropy and MFCC feature vectors in Fig. 5. The result of this measurement indicates the considerable distances between feature vectors corresponding to pairs of mobile devices of exact model, therefore justifying the effectiveness of entropy–MFCC features in differentiating individual mobile devices.

3.3. Supervised learning methods

Classification is known as a supervised learning method (Bhatt and Kankanhalli, 2011). Multiclass classification problems can be implemented through different classification techniques. However, to determine which approach is the most efficient for a particular problem, systematic methods are required to evaluate how different methods work and to compare these methods with one another. We selected five different classifiers based on their appearances and high performance in related works compared in Table 1. Because LIBSVM (Chang and Lin, 2011) can perform well whenever applied to pattern recognition approaches, the LIBSVM wrapper is used to evaluate the features by using a multi-class SVM classifier with a *radial basis function* (*RBF*) kernel. We employed other classifiers including Naïve Bayesin, linear logistic regression, Neural Network (Multilayer Perceptron), and Rotation Forest as implemented in data mining tool WEKA (Hall et al., 2009).

In our experiments, the total training and testing data sets for all classifiers were selected by using 10-fold cross-validation. The dataset was divided into 10 parts, where each experiment uses one-tenth of the data for testing and the remaining nine-tenth of the data for training. Both training and testing data consist of an equal number of data instances from each class.

3.4. Unsupervised learning methods

Unsupervised learning method is the general term for clustering algorithms. In this case, no class exists for the prediction, and the data instances are divided into groups based on the relationships between features, such as distance-based similarity measures, as well as hierarchical and incremental relationships (Xu and Wunsch, 2005). We selected density-based spatial clustering of applications with noise (DBSCAN) (Hao et al., 2011) and expectation-maximization (EM)-based clustering (Abbas, 2008) algorithms because they evaluate performance with reliable parameters that enable comparison.

4. Methodology

Audio signals were sampled, quantized, encoded, compressed, and then encapsulated into VoIP packets in a transmission channel. The VoIP packets were transferred to the destination gateway by using peer-to-peer signaling protocols (H.323, SIP), and voice information was recovered through the packets. Fig. 6 illustrates that the DSP performs the first four processes in the transmitting gateway; each step is detailed in Wallace (2011). In this setup, the stationary device records its VoIP calls to mobile devices, while the recording environment is silent and no

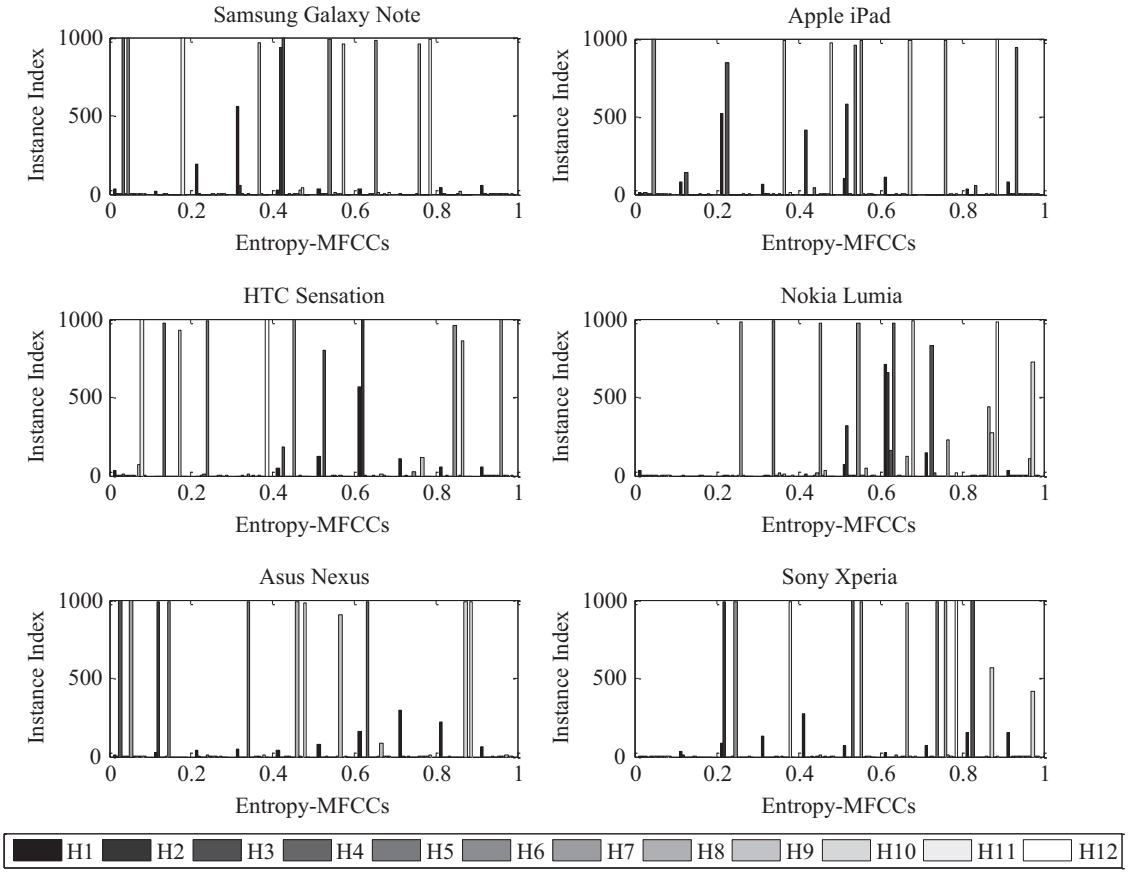

Fig. 4. Histogram of entropy–MFCC features for each mobile device model.

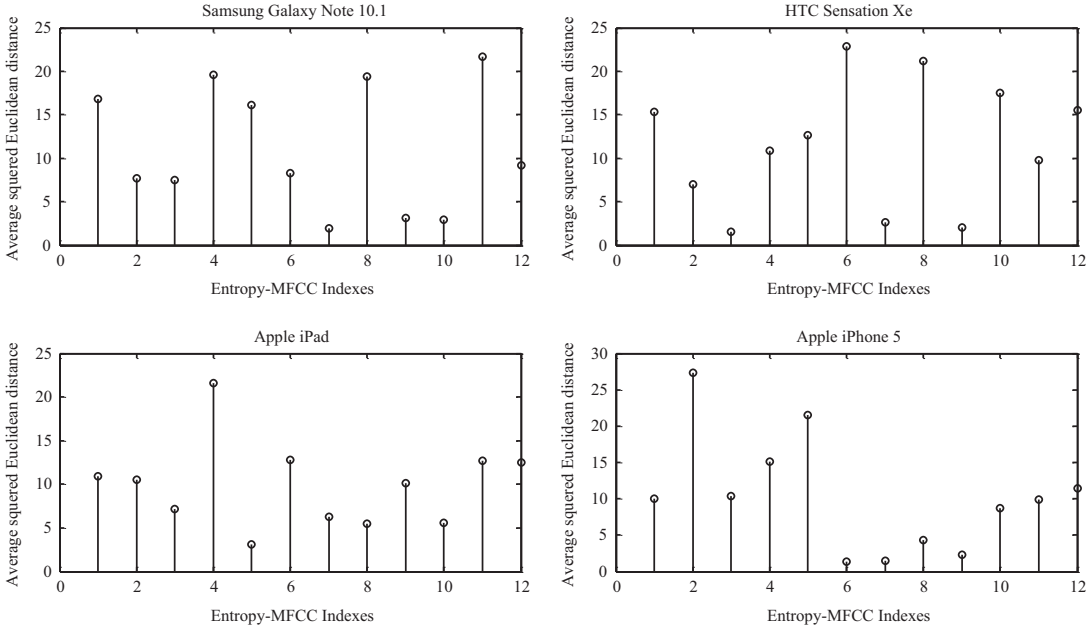

Fig. 5. Average squared Euclidean distances of each entropy–MFCC features on four different mobile devices pairs.

conversation is made between two parties. The stationary device records signals in MP3 format by using Skype Recorder v.3.1 (MP3 Skype Recorder v.3.1). This setup was used for collecting 25 recording files with respect to each mobile device. The mobile devices were of different brands and models, as listed in Table 2. Fig. 1 shows the data preparation approach splits each recording into 40 instances with a length of 0.35–0.42 s, which results in 1000 instances for each mobile device.

The control conditions were as follows: (a) all signals were recorded by the same stationary recorder, (b) the stationary and mobile devices were in the same isolated room, and (c) a call was recorded without any conversation between two parties. We

enforced these conditions to eliminate the influences by different stationary devices, environments and speakers and to capture the signal variations due to different mobile devices.

5. Experimental steps and results

The experiments evaluate the feasibility of the source mobile device identification method through classification accuracy, robustness, and computational efficiency. This process justifies the choices made to handle data sets, features, training and testing instances, classification, and evaluation technique. The first experiment focused on the data preparation approach. The second experiment employed the most common combinations of statistical moments of MFCCs, such as mean, standard deviation, variance, skewness, and kurtosis (Beigi, 2011b). By modifying the feature extraction algorithm in Fig. 1. "Mean-MFCC," "Stdev-MFCC," "Var-MFCC," "Skew-MFCC," and "Kurt-MFCC" were employed. These feature sets are popular among works on musical instrument classification (Senan et al., 2011), as well as speaker verification (Kinnunen et al., 2012), (Alam et al., 2011) and identification (Molla and Hirose, 2004), and were thus adopted for comparison with entropy–MFCC features. The remaining experiments determined the classification performance for individual mobile device models and brands, respectively.

In all experiments, the performance was evaluated at the data instances basis by using classification and clustering methods. The classification was evaluated based on the following metrics and parameters, as detailed in Witten et al. (2011):

(1) *Identification accuracy* (*ACC*)=*Correct classified instances*/*All classified instances*
(2) *Receiver operating characteristic (ROC)*: determines the cost of misclassification error for each individual class by plotting the true positive rate (TPR) on the vertical axis against the true negative rate (TNR) on the horizontal axis.

The performance of the numeric predictions are measured based on the testing data as in (7–10). For all measures, p_i is the numeric value of prediction and a_i is the actual value for the ith instance, where$i = 1, 2, 3, \cdots N$and N is the total number of test instances.

(3) *Root mean squared error* (*RMSE*): computes the square root to determine the same dimensions as the predicted value itself.

$$\sqrt{\frac{(p_1-a_1)^2+\cdots+(p_N-a_N)^2}{N}} \tag{7}$$

(4) *Mean absolute error* (*MAE*): treats all sizes of error evenly according to their magnitude.

$$\frac{|p_1-a_1|+\cdots+|p_N-a_N|}{N} \tag{8}$$

(5) *Root relative squared error* (*RRSE*): computes the total squared error and normalizes it through dividing by the total squared error based on a simple predictor. The predictor is the average of the actual values from the training data that is represented by $\bar{a}$.

$$\sqrt{\frac{(p_1-a_1)^2+\cdots+(p_N-a_N)^2}{(a_1-\bar{a})^2+\cdots+(a_N-\bar{a})^2}} \tag{9}$$

(6) *Relative absolute error* (*RAE*): considers the total absolute error, with the same normalization approach, as in (10).

$$\frac{|p_1-a_1|+\cdots+|p_N-a_N|}{|a_1-\bar{a}|+\cdots+|a_N-\bar{a}|} \tag{10}$$

Selected metrics and parameters appear in abbreviated form in Tables 3–5 and 7. Alternatively, the performance was measured by using clustering algorithms. However, obtaining the following metrics and parameters is sometimes difficult, as detailed in Witten et al. (2011).

1) *Incorrectly clustered instances*: number of instances assigned incorrectly to the clusters
2) *Unclustered instances*: number of instances that are not assigned to any cluster
3) *Log likelihood* (*LL*): measures the goodness of fit; a larger value indicates that model fits the data better.
4) *Minimum description length* (*MDL*) *metric*: determines the MDL score for k parameters and Ninstances as detailed in Hao et al. (2011),

$$MDL\ Score = -LL + k/2 \log\ N. \tag{11}$$

For n independent features, there are $2n$ parameters corresponding to their mean and standard deviation. The MDL score is smaller for strong clustering techniques. Nevertheless, this value increases for less strong clustering techniques.

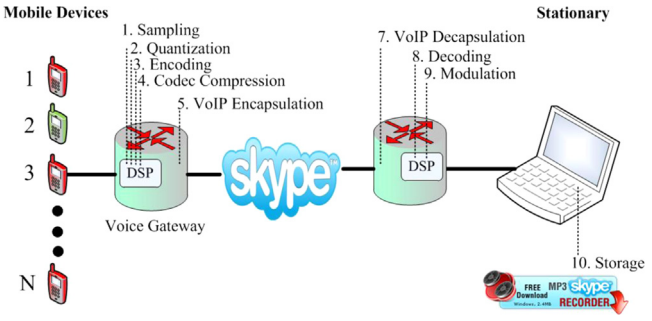

Fig. 6. VoIP network setup for call recording.

Table 2
Mobile devices, models and class names used in experiments.

Class name	Mobile devices	Models	Operating system	Class name	Mobile devices	Models	Operating system
G1	Galaxy Note 10.1-A	GT-N8000	Android 4.1.2	Ip1	Apple iPad	MC775ZP	Apple iOS 5.1.1
G2	Galaxy Note 10.1-B	GT-N8000	Android 4.1.2	Ip2	Apple iPad New	MD366ZP	Apple iOS 5.1.1
G3	Galaxy Note	GT-N7000	Android 2.3.6	I1	Apple iPhone 3	MB489B	Apple iOS 4.1
G4	Galaxy Note II-A	GT-N7100	Android 4.1.2	I2	Apple iPhone 4		
G5	Galaxy Note II-B	GT-N7100	Android 4.1.2	I3	Apple iPhone 4S	MD242B/A	Apple iOS 6.1.2
G6	Galaxy Note II-C	GT-N7100	Android 4.1.2	I4	Apple iPhone 5-A	MD297MY/A	Apple iOS 6.1
GT	Galaxy Tab 10.1	GT-P7500	Android 3.1	I5	Apple iPhone 5-B	MD297MY/A	Apple iOS 6.1
GM	Galaxy Minitab SIII	GT-I8190N	Android 4.1.2	A	Asus	Nexus 7	Android 4.2.2
GS	Galaxy SII			S	Sony Xperia Tipo	ST21i	Android 4.0.4
H1	HTC Sensation XE-A	-	Android 4.0.3	N	Nokia Lumia 710	-	Windows Phone 7.5
H2	HTC Sensation XE-B	-	Android 4.0.3				

Table 3
Performance compression of entropy–MFCC features from enhanced and original audio signals.

Classifiers	*Entropy–MFCC*					*Noisy entropy–MFCC*				
	MAE	*RMSE*	*RAE*	*RRSE*	*ACC*	*MAE*	*RMSE*	*RAE*	*RRSE*	*ACC*
SVM	0.0003	0.0186	0.38%	8.72%	99.64%	0.0005	0.0221	0.54%	10.39%	99.49%
Neural Network	0.0008	0.0179	0.86%	8.42%	99.60%	0.0008	0.019	0.84%	8.91%	99.55%
Naïve Bayesian	0.0001	0.0093	0.10%	4.37%	99.91%	0.0002	0.0139	0.22%	6.54%	99.80%
Rotation Forest	0.0009	0.0154	1.03%	7.25%	99.80%	0.0011	0.0168	1.17%	7.91%	99.73%
Linear Logistic Regression	0.0005	0.0223	0.56%	10.47%	99.47%	0.0007	0.0258	0.77%	12.10%	99.28%

Table 4
Performance of statistical moments of MFCCs.

Classifiers	*Statistical moments of MFCCs*						
	Mean	*Stdev*	*VAR*	*Skew*	*Kurt*	*Combined set*	*Combined with best-first*
SVM	32.83	28.18	30.08	31.35	21.54	69.02	69.62
Neural Network	39.43	28.75	30.78	31.33	21.99	88.42	73.35
Naïve Bayesian	29.29	26.96	27.0	32.90	18.60	60.12	61.18
Rotation Forest	95.10	40.96	42.79	30.57	17.83	89.48	84.36
Linear Logistic Regression	28.27	26.63	33.48	30.79	22.30	84.45	68.75

Table 5
Performance compression of entropy–MFCC features and entropy–[DCT of MFBE] based on model.

Classifiers	*Entropy–MFCC*					*Entropy–DCT of MFBE*				
	MAE	*RMSE*	*RAE*	*RRSE*	*ACC*	*MAE*	*RMSE*	*RAE*	*RRSE*	*ACC*
SVM	0.0002	0.0158	0.28%	7.42%	99.74%	0.0003	0.0183	0.37%	8.60%	99.65%
Neural Network	0.0006	0.0153	0.72%	7.16%	99.68%	0.0007	0.0164	0.74%	7.68%	99.65%
Naïve Bayesian	0	0.0030	0.01%	1.41%	99.99%	0	0.0037	0.02%	1.73%	99.99%
Rotation Forest	0.0004	0.0181	0.39%	8.50%	99.80%	0.0009	0.0154	0.99%	7.21%	99.84%
Linear Logistic Regression	0.0005	0.0223	0.56%	10.47%	99.63%	0.0006	0.0232	0.62%	10.89%	99.42%

The aforementioned parameters appear in Table 6 in abbreviated form.

5.1. Experiment on data preparation approaches

This experiment used the data instances prepared with the enhancement technique as described in Section 3.1 against data instances prepared from the original audio signals to justify the choices made against its alternatives. All 21 mobile devices were employed with 1000 data instances from each to evaluate the performance of the five classification algorithms through 10-fold cross-validation; the results are listed in Table 3. The clean data instances obtained the best performance with 99.91% identification accuracy and root relative squared error of 4.37% by using the naïve Bayesian classifier. The result shows that the environmental noise distortions in the original data instances slightly reduce classification accuracy to 99.80% with respect to the clean data instances. Although the effect of de-noising on classification accuracy is minimal, it increases the computational time particularly for the linear logistic regression model. This finding also suggests the robustness of the entropy–MFCC features against environmental noise distortions.

5.2. Experiment on entropy–MFCC features

This experiment was conducted to indicate the contribution of entropy and MFCCs in identification performance as discussed in Section 3.2. To justify our choices on entropy, we compared the performance of entropy–MFCC features for source mobile device identification with other statistical moments of MFCCs, as adopted in Senan et al. (2011), Kinnunen et al. (2012) and Molla and Hirose (2004). As a result, five feature sets of "Mean-MFCC", "Stdev-MFCC," "Var-MFCC," "Skew-MFCC," and "Kurt-MFCC" were computed. The various moments of MFCCs were concatenated to a single feature vector. This feature vector contains 60 features that were reduced to 48 by using best-first search method. The best-first method traverses the feature space to find the best subset by evaluating each one through the SVM classifier. This search method uses Greedy hill climbing with backtracking algorithms (Goldberg, 1989). The experiment evaluates the performance of all feature sets by using five classification and two clustering algorithms via 10-fold cross-validation. In the second part of the experiment, we eliminated the logarithmic transformation of MFCCs from (5) to compute the DCT of MFBE (discrete cosine transform of Mel-filter bank energies). The

Table 6
Clustering performance based on model.

EM algorithm			
Feature sets	***ICI***	***LL***	***MDL***
Entropy–MFCC	4841	24.10	27.77
DBSCAN Algorithm			
Feature Sets	***ICI***	***UCB***	***GC****
Entropy–MFCC	1	880	21

* All instances=21,000. GC=Generated Clusters, ICI: incorrectly classified instances

identification performance based on DCT of MFBE was obtained and compared against entropy–MFCCs. This comparison was to study the effect of frequency domain features with multiplicative components on identification performance.

5.2.1. Performance comparison in classifying mobile devices based on entropy–MFCCs and statistical moments of MFCCs

Table 4 reveals the classification results for different statistical moments of MFCCs, the combined feature set and its best-first selected features. The highest accuracy rate was always determined when the feature set was used in rotation forest classifier. This classifier achieved an accuracy of 95.10% for "Mean-MFCC" feature set. Meanwhile, for most classifiers, the highest accuracy rate was obtained with combined feature set. However, feature selection produces small improvement in classification accuracy. This result was compared against the performance of the entropy–MFCC features as appeared in Table 5. The entropy–MFCC feature set always outperforms statistical moments of MFCCs with higher accuracy rates. This outcome agrees with the comparison of ROC curves that were obtained from these feature sets. Fig. 7 compares the overall ROC curves of the Rotation Forest classifier among all feature sets and label class. The ROC area for the entropy–MFCC features was close to one, but the value was smaller for other feature sets. This finding indicates that for entropy–MFCC features, the false positive rate is close to zero, and the true positive rate is close to one. Moreover, the ROC area for the "AllSet" features that were produced with the combined statistical moments of MFCCs was significantly smaller than the entropy–MFCCs. Overall, because in near-silent segments, fewer contents exist to be modeled by MFCCs, their value was not large enough to represent the strong discrimination among mobile devices. Meanwhile, entropy intensifies the value of MFCCs in near-silent segments and increases the classification accuracy. Fig. 8 illustrates the classifier benchmarking for Entropy–MFCC feature set, where vulnerability is the performance reduction due to replacing "Entropy–MFCC" with "Stdev-MFCC". As can be seen, the Rotation Forest and SVM classifier exhibited the lowest increase in error rates. This observation suggests the robustness of both classifiers against loss of accuracy rates. Naïve Bayesian classifier generally achieved high identification accuracy at the shortest computation time but with the lowest robustness. Rotation Forest achieved the second-best identification accuracy and the best robustness, but the computation time was considerably slower. Moreover, the performance of the SVM classifier was comparable with the Rotation Forest classifier.

5.2.2. Performance comparison in classifying mobile devices based on entropy–MFCCs and entropy–[DCT of MFBE]

Table 5 shows the performance of the entropy–MFCC feature set against entropy–[DCT of MFBE]. The result shows both feature sets performed comparably. This is because entropy intensifies the energy of the silent segments and moreover, by extracting the features from near-silent segments, convolution due to speech segments is eliminated. The result proves the contribution of entropy in the improvement of the performances of both MFCCs and DCT of MFBE features that was proposed in Hanilçi et al. (2012) for cell-phone identification. In this work, Hanilçi et al. determined that DCT of MFBE features reduces the accuracy rates for identification.

5.2.3. Performance in clustering mobile devices based on entropy–MFCCs

This experiment re-evaluates the performance of all feature sets with probabilistic-based (EM) and nearest-neighbor-based (DBSCAN) algorithms. However, only the entropy–MFCC feature set can diverge to assessable results. Table 6 summarizes the results by using DBSCAN clustering based on the entropy–MFCC feature set with a minimum neighbor distance of $\varepsilon = 0.4$ and minimum cluster size of 200. This algorithm identified 21 clusters with respect to the total mobile devices with only one incorrectly clustered instance. However, 880 instances out of 21,000 instances were unclustered. The EM algorithm inserts the number of clusters beforehand and then determines incorrectly clustered instances, LL, and MDL metrics. Thus, 4841 instances out of 21,000 instances were incorrectly clustered. Smaller MDL indicates strong clustering techniques. DBSCAN assigned more instances to its correct cluster, which makes it the better choice.

5.3. Intra-mobile device identification by using SVM

Source mobile device identification in previous experiments performed well with the SVM classifier in terms of identification accuracy, robustness, and computational efficiency. This experiment analyzes the identification accuracy of individual mobile devices based on the entropy–MFCC feature set and SVM classifier. The confusion matrix in Table 7 shows the correct and incorrect classified instances in diagonal and non-diagonal cells, respectively. Moreover, the proposed method can distinguish among mobile devices of the same model, such as Galaxy Note 10.1 (A,B), Galaxy Note II (A–C), and iPhone 5 (A,B). Minimal misclassifications

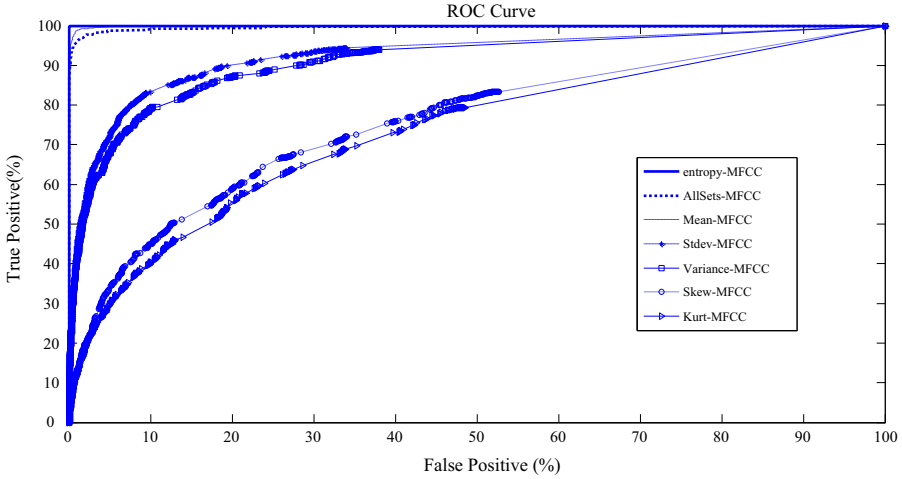

Fig. 7. Overall ROC curves of Rotation Forest classifier using different feature sets on the class of labels.

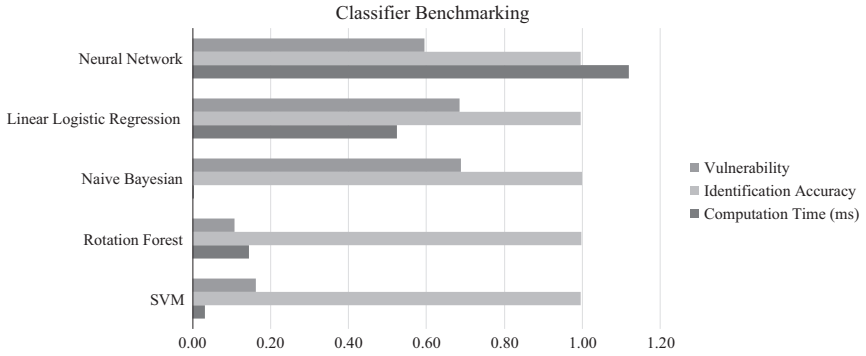

Fig. 8. Classifier benchmarking based on vulnerability, identification accuracy and computation time.

Table 7
Confusion matrix of SVM based on intra-mobile devices identification.

ACC=99.74%		***Predicted***																				
		G1	***G2***	***G3***	***G4***	***G5***	***G6***	***GT***	***GM***	***GS***	***H1***	***H2***	***Ip1***	***Ip2***	***I1***	***I2***	***I3***	***I4***	***I5***	***A***	***S***	***N***
Actual	***G1***	997	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
	G2	0	998	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G3	0	0	996	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0
	G4	0	1	0	998	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	G5	1	0	0	0	996	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	G6	0	0	0	0	0	1000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	GT	0	0	0	0	0	0	997	0	0	0	0	0	0	0	0	0	0	0	2	1	0
	GM	0	2	0	0	0	0	0	997	0	0	0	0	0	0	0	0	0	1	0	0	0
	GS	0	0	0	0	0	0	0	0	997	0	0	0	0	1	0	0	0	0	1	1	0
	H1	0	0	0	2	0	0	0	0	0	997	0	0	0	0	0	0	0	0	0	0	1
	H2	0	0	1	0	0	0	0	0	0	0	998	0	0	0	0	0	0	0	0	1	0
	Ip1	1	0	0	1	0	0	0	0	0	0	0	998	0	0	0	0	0	0	0	0	0
	Ip2	1	0	0	0	0	0	0	0	0	0	0	0	997	1	0	0	0	0	0	1	0
	I1	0	0	0	0	0	0	0	0	0	0	1	0	1	996	0	1	1	0	0	0	0
	I2	0	0	0	0	0	0	0	0	1	0	0	0	0	1	998	0	0	0	0	0	0
	I3	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	997	1	0	0	0	0
	I4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	1	996	0	0	0	0
	I5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	998	0	0	1
	A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	999	1	0
	S	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	997	0
	N	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	998

occurred among mobile devices of different models and brands, which may be a result of signal loss during Skype communication. Overall, the performance was satisfactory for ideal environments, which indicates a promising result when employing entropy–MFCC features in real-world scenarios.

5.4. Inter-mobile devices identification

This experiment represents the mobile devices of the same brand in one class. Five classification algorithms were used among six classes with 10-fold cross-validation for the evaluation of the entropy–MFCC features. Table 8 shows that the Rotation Forest classifier performed better than all the other classifiers for inter-mobile device identification. Furthermore, for most classifiers, the overall performance slightly improved with respect to the classification results based on models. However, in terms of Naïve Bayesian classifier, the classification accuracy was reduced from 99.99 to 99.69% and the error rates were increased. Meanwhile, SVM classifier achieved comparably close accuracy rates with Rotation Forest, but its computation time was faster. Thus, the experiment revisits the performance of the SVM classifier for each particular brand. Table 9 shows the confusion matrix that resulted from 10-fold cross validation by using a six-class SVM classifier. The last row of the confusion matrix is the number of predicted instances and the last column is the total number of instances with

Table 8
Performance of entropy–MFCC features for inter-mobile devices identification.

Classifiers	***Entropy–MFCC***				
	MAE	***RMSE***	***RAE***	***RRSE***	***ACC***
SVM	0.0008	0.029	0.37%	8.56%	99.75%
Neural Network	0.0015	0.0291	0.66%	8.58%	99.69%
Naïve Bayesian	0.0235	0.0953	10.21%	28.12%	97.16%
Rotation Forest	0.0015	0.018	0.64%	5.32%	99.93%
Linear Logistic Regression	0.0024	0.0362	1.02%	10.67%	99.57%

Table 9
Confusion matrix of SVM based inter-mobile devices identification.

ACC=99.75%		***Predicted***						
		Galaxy	*HTC*	*Apple*	*Asus*	*Sony*	*Nokia*	***Total***
Actual	*Galaxy*	8984	3	10	1	2	0	9000
	HTC	6	1990	2	0	0	2	2000
	Apple	14	3	6981	0	1	1	7000
	Asus	4	0	0	996	0	0	1000
	Sony	0	0	2	0	998	0	1000
	Nokia	0	0	2	0	0	998	1000
	Total	9008	1996	6997	997	1001	1001	

respect to each class. The larger misclassifications exist among Samsung and Apple with 9000 and 7000 data instances in each class respectively. An average identification accuracy of 99.75% was achieved for inter-mobile device identification, which is approximately similar to the results for intra-mobile device identification by using the same classifier.

6. Discussion

Prior work has focused on source recording device identification from both speech and non-speech recording. Hanilçi et al. (2012), for example, used cell phone devices as an ordinary tape recorder to collect speech recordings. Although these studies proved that the MFCCs extracted from the speech recording is the most effective feature set to capture device specific features, the results lack evaluation on the robustness of MFCCs. This is because MFCC features are contextualized by the speech contents, speaker's characteristics and environment. In this study, we improved the robustness of MFCCs by computing the entropy of MFCCs from near-silent segments for the prototype of source mobile device identification.

We found that by using all selected classifiers, entropy–MFCC feature set exhibits high performance against statistical moments of MFCCs. Meanwhile, by using near-silent segments, even the combination of entropy with frequency domain features performed well for source mobile device identification. These findings proved the significance of eliminating convolution due to speech signals. Furthermore, in terms of classifiers, Rotation Forest and SVM classifier achieved the best performance with respect to the classification accuracy, robustness and computational efficiency. Some aspects of the proposed method compares well with existing research on acquisition device identification (Kraetzer et al., 2007, 2011, 2012; Buchholz et al., 2009; Garcia-Romero and Epsy-Wilson, 2010; Hanilçi et al., 2012; Panagakis and Kotropoulos, 2012). However, our method adds an advantage to the previous approaches in the following ways: (a) entropy–MFCC features are extracted from near-silent frames, (b) entropy of Mel-cepstrum output intensifies the energy of MFCCs for near-silent frames, and (c) blind identification of mobile devices over the call. This study therefore indicates that entropy–MFCC features identify the distinguishing pattern in mobile devices of even the same model.

Most notably, this study is the first to identify traces of the source transmitting devices by detecting the near-silent segments in a recorded conversation. We found evidence to suggest that our prototype can identify different source brand/model and individual mobile devices in a more practical experimental setup by using communication through any type of service provider, such as cellular, VoIP, PSTN, and their combinations and subsets. However, some limitations are worth noting. Although we found promising results based on silent recording, the proposed method was not reassessed based on speech recording. Future work includes a follow-up work designed to evaluate the accuracy when speech is recorded by the mobile device without transmission and when the experimental setup in Fig. 6 is implemented. This way it would be possible to determine the effects of speech contents and speaker characteristics on identification accuracy. In addition we can perform quantitative comparison with current state-of-the-art approaches.

Acknowledgment

This work is fully funded by the Ministry of Education, Malaysia under the University of Malaya High Impact Research Grant UM.C/625/1/HIR/MoE/FCSIT/17.

References

Abbas, O.A., 2008. Comparison between data clustering algorithms. Int. Arab J. Inf. Technol. 5, 320–325.
AlQahtani, M.O., Al Mazyad, A.S., 2011. Environment Sound recognition for digital audio forensics using linear predictive coding features. In: Snasel, V., Platos, J., ElQawasmeh, E. (Eds.), Digital Information Processing and Communications Pt. 2., vol. 189, pp. 301–309.
Alam, M.J., Ouellet, P., Kenny, P., O'Shaughnessy, D., 2011. Comparative evaluation of feature normalization techniques for speaker verification. In: Proceedings of the NOLISP, pp. 246–253.
Balasubramaniyan, V.A., Aamir, P., Ahamad, M., Hunter, M.T., Trayno, P., 2010. PinDr0p: using single-ended audio features to determine call provenance. Presented at the Proceedings of the CCS, New York, NY, USA.
Beigi, H., 2011a. Signal processing of speech and feature extraction, Fundamentals of Speaker Recognition. Springer, New York, NY, USA, pp. 143–199.
Beigi, H., 2011b. Probability theory and statistics, Fundamentals of Speaker Recognition. Springer, New York, NY, USA, pp. 239–247.
Beigi, H., 2011c. Information theory, Fundamentals of Speaker Recognition. Springer, New York, NY, USA, pp. 265–299.
Bhatt, C.A., Kankanhalli, M.S., 2011. Multimedia data mining: state of the art and challenges. Multimed. Tools Appl. 51, 35–76.
Buchholz, R., Kraetzer, C., Dittman, J., 2009. Microphone classification using Fourier coefficients. In: Katzenbeisser, S., Sadeghi, A.-R. (Eds.), IH, LNCS 5806. Springer Berlin Heidelberg; Darmstadt, Germany, pp. 235–246.
Celiktutan, O., Sankur, H., Memon, N., 2008. Blind identification of source cell-phone model,. IEEE Trans. Inf. Forensics Secur. 3, 553–556.
Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 1–27.
Cooper, A.J., 2011. Further considerations for the analysis of ENF data for forensic audio and video applications. Int. J. Speech Lang. Law 18 (2011), 99–120.
Garcia-Romero, D. Epsy-Wilson, C.Y., 2010. Automatic acquisition device identification from speech recordings. Presented at the Proceedings of the ICASSP, Dallas, Texas.
Gerkmann, T., Hendriks, R.C., 2012. Unbiased MMSE-based noise power estimation with low complexity and low tracking delay. IEEE Audio, Speech, Lang. Process. 20, 1383–1393.
Ghosal, A., Chakraborty, R., Chakraborty, R., Haty, S., Dhara, B.C., Saha, S.K., 2009. Speech/music classification using occurrence pattern of ZCR and STE. In: Proceedings of the IITA, pp. 435–438.
Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Longman Publishing Co. Inc, Boston, MA, USA.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. SIGKDD Explor. 11, 10–18.
Hanilçi, C., Ertaş, F., Ertaş, T., Eskidere, Ö., 2012. Recognition of brand and models of cell-phones from recorded speech signals, IEEE Trans. Forensics Secur. 7, 635–638.
Hao, L., Lewin, P.L., Hunter, J.A., Swaffield, D.J., Contin, A., Walton, C., et al., 2011. Discrimination of multiple PD sources using wavelet decomposition and principal component analysis. IEEE Trans. Dielectr. Electr. Insul. 18, 1702–1711.
Jenner, F., 2011. Non-intrusive identification of speech codecs in digital audio signals. ProQuest (M.S. thesis). Computer Engineering Department, KGCOE, RIT University, Rochester, New York, NY, USA.
Keonig, B.E., Lacey, D.S., 2012. Forensic authenticity analysis of the header data in re-encoded WMA files from small Olympus audio recorders,. J. Audio Eng. Soc. 60, 255–265.
Kharrazi, M., Sencar, H., Memon, N., 2004. Blind source camera identification. In: Proceedings of the ICIP, Singapore, pp. 709–712.
Kinnunen, T., Saeidi, R., Sedlák, F., Lee, K.A., Sandberg, J., Hansson-Sandsten, M., et al., 2012. Low-Variance multitaper MFCC features: a case study in robust speaker verification,. IEEE Audio, Speech, Lang. Process. 20, 1990–2001.
Kraetzer, C., Dittmann, J., 2010. Improvement of information fusion-based audio steganalysis. Presented at the Proceedings of the IS&T/SPIE, 7542.
Kraetzer, C., Oermann, A., Dittmann, J., Lang, A., 2007. Digital audio forensics: a first practical evaluation on microphone and environment classification. In: Proceedings of the MM&Sec, Dallas, Texas, pp. 63–74.
Kraetzer, C., Qian, K., Schott, M., Dittmann, J., 2011. A context model for microphone forensics and its application in evaluations. In: Presented at the Proceedings of the SPIE-IS&T, San Francisco, CA.
Kraetzer, C., Qian, K., Dittmann, J., 2012. Extending a context model for microphone forensics. In: Presented at the Proceedings of the SPIE 8303, Burlingame, CA.
MP3 Skype Recorder v.3.1. Available: ⟨http://voipcallrecording.com⟩.
Maher, R., 2009. Audio forensic examination. IEEE Signal Process. Mag. 26, 84–94.
Molla, M.K.I. Hirose, K., 2004. On the effectiveness of MFCCs and their statistical distribution properties in speaker identification. In: Proceedings of the VECIMS, Boston, MA, pp. 12–14.
Muhammad, G., Alghathbar, K., 2013. Environment recognition for digital audio forensics using MPEG-7 and mel cepstral features. Int. Arab J. Inf. Technol. 10, 43–50.
Nilsson, M., 2006. Entropy and Speech (Ph.D. thesis), EE Department, SIP, KTH University, Stockholm, Sweden.
Ode Ojowu, J., Karlsson, J.J., Liu, Y., 2012. ENF extraction from digital recordings using adaptive techniques and frequency tracking. IEEE Trans. Inf. Forensics Secur. 7, 1330–1338.

Panagakis, Y., C. Kotropoulos, 2012. Telephone handset identification by feature selection and sparse representations. Presented at the Proceedings of the WIFS, Tenerife.

Senan, N., Ibrahim, R., Nawi, N.M., Yanto, I.T.R., Herawan, T., 2011. Rough set approach for attributes selection of traditional Malay musical instruments sounds classification. Int. J. Database Theory Appl. 4, 59–76.

Sharma, D., Hilkhuysen, G., Hilkhuysen, N., Naylor, P., Brookes, M., Huckvale, M., 2010. Data driven method for non-intrusive speech intelligibility estimation.In: Presented at the Proceedings of the EUSIPCO, Aarlborg, Denmark.

Shen, J.-L., Hwang, W.-L., Lee, 1996. L.-S., Robust speech recognition features based on temporal trajectory filtering of frequency band spectrum. In: Proceedings of the ICSLP, vol. 2.

Swaminathan, A., Wu, M., Liu, K., 2007. Nonintrusive components forensics of visual sensors using output images,. IEEE Trans. Inf. Forensics Secur. 2, 91–106.

Wallace, K., 2011. Configuring basic voice over IP, Implementing Cisco Unified Communications Voice over I.P. and QoS (CVOICE), 4th ed. Cisco Systems, Indianapolis, IN, pp. 165–294.

Witten, I.H., Frank, E., Hall, M.A., 2011. Chapter 5 – credibility: evaluating what's been learned. In: Witten, I.H., Frank, E., Hall, M.A. (Eds.), Data Mining: Practical Machine Learning Tools and Techniques, third ed. Morgan Kaufmann, Boston, pp. 147–187.

Xu, R., Wunsch II, D., 2005. Survey of clustering algorithms. IEEE Trans. Neural Netw. 16, 645–678.

Ye, J., 2004. Speech recognition using time domain features from phase space reconstructions.