Australian Journal of Basic and Applied Sciences, 8(21) Special 2014, Pages: 73-79

AENSI Journals

Australian Journal of Basic and Applied Sciences

ISSN:1991-8178

Journal home page: www.ajbasweb.com

Blind Source Computer Device Identification from Recorded Calls

Mehdi Jahanirad, Ainuddin Wahid Abdul Wahab and Nor Badrul Anuar

Faculty of Computer Science and Information Technology University of Malaya50603 Kuala Lumpur, Malaysia
mehdijahanirad@siswa.um.edu.my,

ARTICLE INFO

Article history:
Received 25 April 2014
Received in revised form
8 May 2014
Accepted 20 May 2014
Available online 17 June 2014

Keywords:

ABSTRACT

This study investigates the use of blind source computer device identification for forensic investigation of the recorded VoIP call. It was found that a combination of mel-frequency cepstrum coefficients (MFCCs) and entropy as an intrinsic audio feature captures the specific frequency response due to the tolerance in the nominal values of the electronic components associated to individual computer device. By applying the supervised learning techniques such as naïve Bayesian, linear logistic regression, neural networks (NN), support vector machines (SVM) and sequential minimal optimization (SMO) classifier to the Entropy-MFCC features, state-of-the-art identification accuracy of above 99.8% has been achieved on a set of 5 iMacs and 5 desktop PCs from the same model.

To Cite This Article: Mehdi Jahanirad, Ainuddin Wahid Abdul Wahab and Nor Badrul Anuar., Blind Source Computer Device Identification from Recorded Calls. ***Aust. J. Basic & Appl. Sci.,*** *8(21): 73-79, 2014*

INTRODUCTION

Audio forensics mostly focus on situations that impose trust in authenticity and integrity of audio signals. An example for these scenarios is forensic acquisition, analysis and evaluation of admissible audio recording as crime evidence in court. 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. Authenticity evaluations by using the device-based techniques are attracting widespread interest in fields such as: a) identification of computer-generated audio from the original audio recording file (Keonig, B.E., D.S. Lacey, 2012) identification of the source brand or model of the recording devices, such as telephone handsets, microphones (Garcia-Romero, D., C.Y. Epsy-Wilson, 2010; Panagakis, Y., C. Kotropoulos, 2012) and cell phones (Hanilçi, C., 2012) identification of the speech codecs (Jenner, F., 2011). The most related works to this approach are the recording source forensics.

Kraetzer *et al*. (2007) published the first practical evaluation to identify a device. They adopted the statistical pattern recognition technique to determine the source microphone and its recording environments. 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. With the use of a suitable context model for microphone forensics, Kraetzer *et al*. (2011) extended the works in (Kraetzer, C., 2007) by reevaluating their sample data separately. Kraetzer *et al*. (2012) extended the proposed context model in (Kraetzer, C., 2011) toward better generalization and constructed a new application scenario model for microphone forensic investigations with the aim of detecting playback recordings. In addition to microphone identification, Garcia–Romero and Espy–Wilson (Garcia-Romero, D., C.Y. Epsy-Wilson, 2010) extended device identification by identifying landline telephone handsets. A similar landline telephone handset identification method proposed by Panagakis and Kotropoulos (2012) improved the accuracy of the identification by using sparse representation classifier (SRC). Hanilçi *et al*.(2012) proposed a cell phone identification method that uses SVM by identifying the audio source based on the cell phone brand and model. However, these studies lack sufficient study to eliminate convolution by speech context. Moreover, they only focused on identifying source recording devices.

This paper focuses on the novel idea of recognizing the communicating acquisition devices based on the recorded call. We present a case study that records calls by using stationary Notebook that makes VoIP calls to computer devices of the same model. The motivation for this case study is based on the fact that the

Corresponding Author: Mehdi Jahanirad, Ainuddin, Faculty of Computer Science and Information Technology University of Malaya50603 Kuala Lumpur, Malaysia
E-mail: mehdijahanirad@siswa.um.edu.my

combination of anonymity, ease of access and free offerings of VoIP services provide fertile ground for criminal activity. Thus, identifying the computer devices used in VoIP call can help the forensic investigator to reveal useful information in addition with authenticity of the recorded call. This study identifies source brand and model of the computer devices from recorded calls based on entropy-mel-frequency cepstrum coefficient (MFCC) features. Extracting the entropy of MFCCs adds an advantage to the application of MFCCs in by eliminating the effects of speech contents. The method evaluates the feasibility of entropy-MFCC features and its robustness against speech signals by using classifier benchmarking.

The remainder of this paper is organized as follows: Section 2 discusses an overview of the methodology. Section 3 outlines the recording setup. Section 4 describes the experiments and evaluates the performance of the proposed method. Finally, Section 5 discusses further implications of the practical study, its limitations, and future applications.

1. Source Computer Device Identification Scheme:

In our implementation we followed (Bhatt, C.A., M.S. Kankanhalli, 2011) by using audio mining techniques: a) creates blocks through preprocessing the recorded samples, b) determines intrinsic computer device fingerprint through feature extraction and c) uses supervised learning techniques known as classification.

1.1 Preprocessing:

Preprocessing method uses two different approaches as illustrated by Fig. 1 to create blocks from: a) original speech signal, and b) near-silent segments. This is to justify the robustness of the proposed method against speech signals.

Original Speech Signal:

The pre-processing stage includes sampling, framing, windowing and cleaning the signal. The signals become more distinct when noise is eliminated. Thus, we adopted cleaning to remove the noise generated by environment. The algorithm splits clean signals into overlapping audio frames of length 40 samples. In other words the output is a matrix with 40 columns, when each column represents one block for feature extraction.

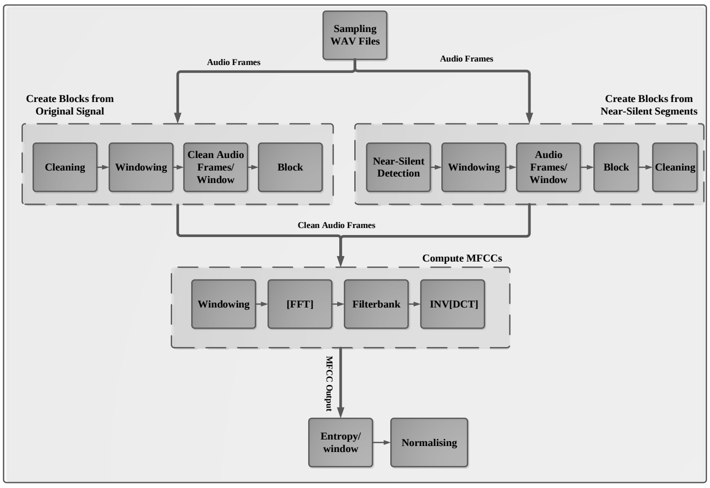

Fig. 1: Flow chart of the proposed feature extraction.

Near-Silent Segments:

This approach uses simple segmentation of the recorded signal in order to extract the near-silent segments as demonstrated in Fig. 2. We implemented the near-silent detection algorithm based on the silent removal approach in (Giannakopoulos, T., 2010), however our objective is to select the silent segments instead of removing them. This method uses two audio features known as signal energy and spectral centroid (Giannakopoulos, T., 2009). This algorithm includes following steps: a) extracts two feature sets from the original recorded signal. b) computes the histogram of the feature series values. c) applies the smoothing filter on histograms. d) estimates two thresholds for each histogram, e) applies a simple thresholding criterion to the feature sets, f) and detects silent segments based on the assigned criterion. Finally, the algorithm assigns equal number of samples from near silent segments into blocks, and then performs cleaning on each block prior to feature extraction.

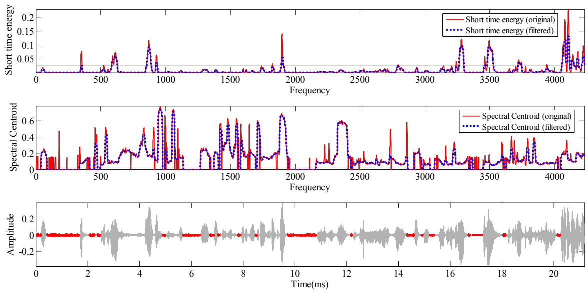

Fig. 2: Visualization of Near-Silent Detection algorithm. Top and middle plots represent the histogram of short time energy and spectral centroids of the signal, respectively. The horizontal line shows the estimated threshold. The bottom plot shows the spectrum of the original signal, when red color determines the near-silent segments.

1.2 Intrinsic Computer Device Fingerprints:

Assuming that the computer device is a linear time-invariant system, its influence on the recorded call is modeled by the convolution of its impulse response and the original speech. The convolution means the spectrum of any recorded speech segment is the product of the spectrum of the original speech signal and the device frequency response. Well known MFCC features are selected to capture the device frequency response because of the fact that the convolution in time domain is represented by the summation in cepstrum domain. Thus, MFCCs produce inherent invariance toward linear spectral distortions. Furthermore, the entropy of MFCCs reduces the dimensionality of the feature space. According to information theory, this value increases for silent segments that contain uncertainty and reduces for the speech segments that contain information (Beigi, H., 2011). As a result this paper computes Entropy-MFCC features as the intrinsic computer device fingerprints through three stages: a) computes the MFCCs, b) computes the entropy of MFCCs, and c) normalizes the output.

MFCCs:

The complex cepstrum of the signal is defined as the Fourier transform of the log of the signal spectrum. The signal is transferred to a linear frequency scale by using fast Fourier transform and is converted to the mel frequency scale by using a filter bank. Finally, the MFCCs are determined based on the analysis of the short-time mel frequency log spectrum. This analysis includes computing the inverse discrete cosine transform (DCT) of the log spectrum of the signal. The mel-cepstrum output is consists of N frames and 12 coefficients.

Entropy:

The entropy of the MFCC vectors was computed in two stages. First, the spectrum is normalized into the probability mass function (PMF), where is the mel-cepstrum coefficient d in frame l and is the PMF of the signal. In the second stage, the entropy is computed by using

$$H_d = -\sum_{l=1}^{N} {}_l p_d \log_2 {}_l p_d . \qquad (1)$$

Overall, 12 entropy-MFCC features were extracted by using MATLAB functions.

Normalization:

The last and one of the most important steps in preparing the features is normalization. This step reduces large differences between the maximum and minimum data values.

1.3 Supervised Learning Techniques:

Supervised learning techniques known as classification problems can be implemented through different learning algorithms, such as statistical modeling, linear, non-linear, and ensemble learning models. However, to determine which approach is the most efficient for a particular problem, evaluation metrics are required. We selected five simple classifiers based on the fact that the simplicity-first methodology is the best choice for analyzing practical datasets. Furthermore, these algorithms demonstrated high performances for similar works in the literature. Table 1 details the employed classifiers in the experiment and their specifications.

2. Recording Setup:

This setup enables recording of VoIP communication between computer devices and the single stationary inside faculty building. The stationary user makes Skype call to computer user, then records signals in mono with WAV format by using Pamela for Skype-Version 4.8. The setup collects recordings with respect to computer devices including: (a) five iMacs of identical model located in Multimedia Research Lab, (b) five desktop PCs of identical model located in Micro Lab. At the same time the stationary is located inside the open corridor. All conversations are conducted between the same male and female over the experiment. Table 2 indicates the specifications for computer devices that employed in the setup. All iMacs are 21.5-inch, Late 2012 model built with 3.1GHz Intel Core i7 processor, stereo speakers, and dual microphones. All desktop PCs are Lenovo ThinkCentre M81 7518 model with 3.3 GHz Intel Core i3 processor that attached to the same external microphone and uses 32 bit, Windows 7 professional Service Pack 1 as operating system.

Table 1: Classification Algorithms used in the experiment.

Classification Algorithms	Specifications
Naïve Bayesian	Works based on Bayes' rule of conditional probability that assumes independence.
Linear Logistic Regression	Measures goodness of fit by using the log-likelihood of the model.
Neural Network	Primarily learns the network structure and the connection weights by fixing the network structure to determine the weights.
Support Vector Machine	Uses the LibSVM wrapper that implement a multi-class SVM classifier with a radial basis function (RBF) kernel (Chang, C.C., C.J. Lin, 2011).
Sequential minimal-optimization	Implements the sequential minimal-optimization algorithm for training a SVM classifier using a Polynomial kernel.

Table 2: iMac devices and class names.

iMac			Desktop PC	
Serial No.	Software No.	Class	Serial No.	Class
C02JWXXJDXXX	OS X 10.8.4 (12E55)	i1	XXXXXXXXXXXX3D6B	PC1
C02JWXXDDXXX	OS X 10.8.5 (12F45)	i2	XXXXXXXXXXXX3DMK	PC2
C02JWXXGDXXX	OS X 10.8.2 (12C3103)	i3	XXXXXXXXXXXX3AME	PC3
C02JWXXEDXXX	OS X 10.8.3 (12D78)	i4	XXXXXXXXXXXX3CGM	PC4
C02JWXXKDXXX	OS X 10.8.2 (12C3103)	i5	XXXXXXXXXXXX2ZRB	PC5

3. *Experiments and Results:*

The experiments evaluate the feasibility of the source computer device identification method through features that extracted from original speech signal against features that extracted from near silent segments. Moreover the performance of the proposed scheme is evaluated for inter and intra-model device identification using five classification algorithms that implemented in data mining tool Weka Version 3.6 (Hall, M., 2009).

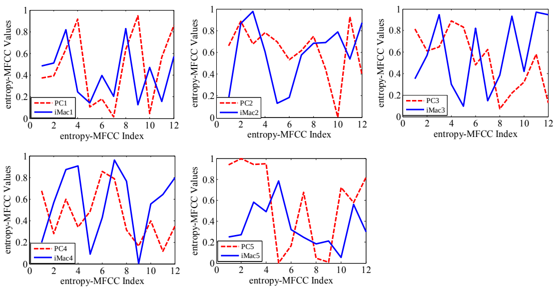

Fig. 3: Histogram of entropy-MFCC features for each iMac devices.

3.1 *Experiment on Original Speech Sample:*

The total of 1320 and 1400 blocks are extracted from original speech samples collected from each iMac and desktop PC devices, respectively. The first part of the experiment visualizes the discriminatory effect of entropy-MFCC features amongst five identical iMacs and desktop PCs by using the histogram of all 12 features, as demonstrated in Fig. 3. For further investigation we selected three different pair of iMacs and examined the squared Euclidean distance between their entropy-MFCC feature vectors in Fig. 4. The result of this measurement indicates the considerable distances between feature vectors corresponding to pair of iMacs, therefore justifying the effectiveness of entropy-MFCC features in differentiating individual iMac devices. The second part of the experiment employs classification benchmarking to evaluate the classification performance among both desktop PCs and iMac devices of identical models.

Table 3 indicates the results of three sets of experiments on original speech signal with all five classifiers, 10-fold cross-validation and default parameters. For intra-model identification based on iMacs, the overall results prove the feasibility of the entropy-MFCC features with high classification accuracies (ACC) (99.82%-100%) and minimal root mean square error (RMSE). Naïve Bayesian classifier performs with higher classification accuracy and computational efficiency. SMO classifier outperforms LibSVM classifier with respect to classification accuracy, but produces lager RMSE. Table 4 shows the confusion matrix for the linear logistic regression classifier, with a total of 12 falsely classified blocks for intra-model device identification amongst five iMacs, the truly and falsely classified blocks (TCB and FCB) are in diagonal and non-diagonal cells, respectively. The performance for intra-model identification based on desktop PCs strongly confirms the previous result, even though the sound was transferred through the same headset microphone for all PCs. Table 5 shows the confusion matrix for the linear logistic regression classifier, with a total of 17 falsely classified blocks for intra-model device identification. Finally the results for source brand/model identification based on all computer devices show that increasing the number of blocks amplified the computation time, however the classification accuracy maintains in the same range. Moreover, lower RMSE values indicates that increasing the number of blocks reduces the average classification error. Table 6 shows the confusion matrix for the linear logistic regression classifier, with a total of 33 falsely classified blocks for source brand/model device identification based on all devices.

Table 3: Performance on intra-model identification using original speech signal.

Intra-model identification Based on iMacs					
Classification Algorithms	*TCB*	*FCB*	*RMSE*	*ACC*	*Elapsed Time (s)*
Naïve Bayesian	6600	0	0	100 %	2.6
Linear Logistic Regression	6588	12	0.0218	99.82 %	35.1
NN	6591	9	0.0202	99.86 %	72.7
LibSVM	6593	7	0.0206	99.89 %	5.5
SMO	6596	4	0.3157	99.94 %	41.5
Intra-model Identification Based on desktop PCs					
Naïve Bayesian	7000	0	0	100 %	0.51
Linear Logistic Regression	6983	17	0.029	99.76%	36.53
NN	6987	13	0.0224	99.81%	140.81
LibSVM	6993	7	0.02	99.90 %	6.0
SMO	6989	11	0.3163	99.84 %	1.18
Source brand/model Identification Based on all computer devices					
Naïve Bayesian	13600	0	0	100 %	1.11
Linear Logistic Regression	13567	33	0.0194	99.76%	146.63
NN	13570	30	0.0183	99.78%	450.22
LibSVM	13579	21	0.0176	99.85 %	15.81
SMO	13575	25	0.2719	99.82 %	3.03

Table 4: Confusion Matrix of linear logistic regression classifier based on iMacs.

ACC= 99.82%		*Predicted*				
		i1	*i2*	*i3*	*i4*	*i5*
Actual	*i1*	1318	0	0	1	1
	i2	1	1316	1	2	0
	i3	1	1	1318	0	0
	i4	1	1	1	1317	0
	i5	0	1	0	0	1319

Table 5: Confusion Matrix of linear logistic regression classifier based on PCs.

ACC= 99.76%		*Predicted*				
		PC1	*PC2*	*PC3*	*PC4*	*PC5*
Actual	*PC1*	1397	2	0	1	1
	PC2	1	1396	0	3	0
	PC3	0	0	1400	0	0
	PC4	1	3	0	1395	1
	PC5	2	0	0	3	1395

Table 6: Confusion Matrix of linear logistic regression classifier based on all devices.

ACC= 99.82%		*Predicted*									
		i1	*i2*	*i3*	*i4*	*i5*	*PC1*	*PC2*	*PC3*	*PC4*	*PC5*
Actual	*i1*	1317	0	0	0	0	0	1	0	1	1
	i2	0	1317	0	1	0	1	0	0	0	1
	i3	1	0	1318	0	0	0	1	0	0	0
	i4	1	0	0	1316	0	1	1	0	0	1
	i5	0	1	0	0	1317	0	0	1	1	0
	PC1	0	0	0	1	0	1396	1	0	1	1
	PC2	1	0	2	0	0	1	1396	0	0	0
	PC3	1	0	0	0	0	0	1	1398	0	0
	PC4	3	0	0	0	0	1	0	0	1395	1
	PC5	0	0	0	0	1	0	0	0	2	1397

3.2 Experiment on Near-Silent Segments:

This experiment aims to evaluate the performance of the proposed scheme without the interference of the speech signal. The experiment extracts the near-silent segments according to the algorithm that discussed in Sub-Section 1.1, and creates a total of 1319 and 1400 blocks with respect to each iMac and desktop PC devices. At first the average squared Euclidean distances are calculated as in previous experiment for the same pair of iMacs to allow comparison, as in Fig. 4. The second part of the experiment repeats the evaluation experiments in previous Sub-Section using blocks created from near-silent segments. In overall the results in Table 7 shows slight improvements in classification accuracy, computation time and RMSR with comparison to Table 3. However, for intra-model identification based on iMacs through NN classifier the computation time is noticeably longer with compare to the results for the original speech samples. It is plausible that creating the blocks from near-silent segments is a practical option for eliminating the effects of signal variations by different speakers in real-time scenarios. However, with adaptation of the entropy-MFCC features the performance that obtained for source computer device identification based on original speech signal is in good agreement with near-silent segments. This justifies our hypothesis as discussed in Sub-Section 2.2 for selecting entropy-MFCC features.

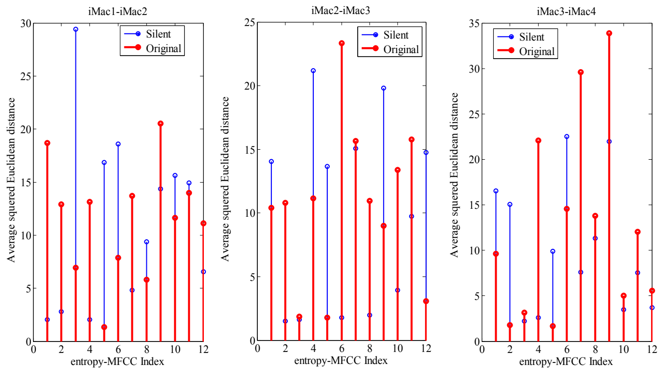

Fig. 4: Average squared Euclidean distances of each entropy-MFCC features on three different iMac pairs.

Table 7: Performance on Proposed Entropy-MFCC Features using Near-Silent Segments.

Intra-model identification Based on iMacs					
Classification Algorithms	*TCB*	*FCB*	*RMSE*	*ACC*	*Elapsed Time (s)*
Naïve Bayesian	6595	0	0	100%	0.54
Linear Logistic Regression	6583	12	0.0218	99.82%	36.28
NN	6587	8	0.0202	99.88%	116.8
LibSVM	6591	4	0.0206	99.94%	5.6
SMO	6593	2	0.3157	99.97%	0.87
Intra-model Identification Based on desktop PCs					
Naïve Bayesian	7000	0	0	100 %	0.51
Linear Logistic Regression	6994	6	0.0164	99.91%	33.15
NN	6995	5	0.0158	99.93%	139.33
LibSVM	6996	4	0.0151	99.94 %	5.30
SMO	6995	5	0.3163	99.93 %	1.09
Source brand/model Identification Based on all computer devices					
Naïve Bayesian	13595	0	0	100 %	1.17
Linear Logistic Regression	13567	19	0.0152	99.86%	144.80
NN	13574	21	0.0183	99.85%	450.60
LibSVM	13578	17	0.0158	99.88 %	14.10
SMO	13578	17	0.272	99.88 %	2.10

3. Conclusions:

A prototype source computer device identification method is developed based on the entropy-MFCC feature set and computer devices that used for the VoIP call. MFCC and entropy features identify the distinguishing pattern amongst individual computer of the same model and computer devices of different brands. This feature exhibited high performance to capture characteristics of the transfer function of the computer devices even with the existence of speech signal's transfer function. The naïve Bayesian classifier always achieved the highest classification accuracy of 100% in blind source computer device identification. However, it is clear that this classification accuracy is too ideal and further studies are required to implement this approach on more real case scenarios and large number of devices, where more advanced classifiers could be required.

ACKNOWLEDGMENTS

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

REFERENCES

Keonig, B.E., D.S. Lacey, 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.

Garcia-Romero, D., C.Y. Epsy-Wilson, 2010. Automatic acquisition device identification from speech recordings. Proc. ICASSP, 1806-1809 Dallas, Texas.

Panagakis, Y., C. Kotropoulos, 2012. Telephone handset identification by feature selection and sparse representations. Proc. WIFS, 73-78, Tenerife.

Hanilçi, C., F. Ertaş, T. Ertaş, Ö. Eskidere, 2012. Recognition of brand and models of cell-phones from recorded speech signals. IEEE Trans. Forensics Security, 7: 635-634.

Jenner, F., 2011. Non-intrusive identification of speech codecs in digital audio signals. ProQuest.

Kraetzer, C., A. Oermann, J. Dittmann, A. Lang, 2007. Acm: Digital Audio Forensics: A First Practical Evaluation on Microphone and Environment Classification. Mm&Sec'07: Proceedings of the Multimedia & Security Workshop, 63-73.

Buchholz, R., C. Kraetzer, J. Dittman, 2009. Microphone classification using Fourier coefficients. IH, LNCS, vol. 5806, pp. 235-246. Springer Berlin Heidelberg, Darmstadt, Germany.

Kraetzer, C., K. Qian, M. Schott, J. Dittmann, 2011. A context model for microphone forensics and its application in evaluations. Proc. SPIE-IS&T, San Francisco. CA.

Kraetzer, C., K. Qian, J. Dittmann, 2012. Extending a context model for microphone forensics. Proc. SPIE, 8303: 83030S-83012, Burlingame, CA.

Bhatt, C.A., M.S. Kankanhalli, 2011. Multimedia data mining: state of the art and challenges. Multimed Tools Appl 51: 35-76.

Giannakopoulos, T., 2010. A method for silence removal and segmentation of speech signals, implemented in Matlab, 1-3. University of Athens, Greece.

Giannakopoulos, T., 2009. Study and application of acoustic information for the detection of harmful content, and fusion with visual information. Dpt of Informatics and Telecommunications, vol. PhD. University of Athens, Greece.

Beigi, H., 2011. Fundamentals of Speaker Recognition. Springer, New York, NY, USA (2011)

Witten, I.H., E. Frank, M.A. Hall, 2011. Data Mining: Practical Machine Learning Tools and Techniques (Third Edition). Morgan Kaufmann, Boston.

Chang, C.C., C.J. Lin, 2011. LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2: 1-27.

Pamela for Skype — Professional Edition 4.8, http://www.pamela.biz/en/

Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, 2009. The WEKA Data Mining Software: An Update. SIGKDD Explorations, 11: 10-18.