Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ANSWERING INCOMING CALL FOR IMPLICIT AUTHENTICATION USING SMARTPHONE

[1]M.F.A. ABDULLAH, [2]H. K. BASHIER, [3]S. SAYEED, [4]I. YUSOF, [5]A. AZMAN, [6]S. Z. IBRAHIM, [7]LIEW T. H.

[1,2,3,4,5,6,7]Faculty of Information Science and Technology (FIST), Multimedia University, Melaka, Malaysia

E-mail: [1]mfikriazli.abdullah@mmu.edu.my, [2]me.the.fren@gmail.com, [3]shohel.sayeed@mmu.edu.my, [4]ibrahim.yusof@mmu.edu.my, [5]afizan.azman@mmu.edu.my, [6]sitizainab.ibrahim@mmu.edu.my, [7]thliew@mmu.edu.my

ABSTRACT

Smartphones are being used to keep sensitive data and make private transaction other than making calls and receive short messages. Thus, authentication of the smartphones becomes very crucial and important aspect. However, users feel inconvenience and difficult with current authentication methods, from password up to physical biometrics. Implicit authentication system emerged intending to improve the security and convenience of the smartphone users. One of the approaches is considering the way users answer incoming phone calls using their smartphones. We study and evaluate the voice signal from users when answering incoming phone calls. Our study shows that the voice signals capable of authenticating the smartphone users. The experiment conducted shows a very high performance with 98.9% accuracy. These findings will promisingly augment that the novel implicit and transparent authentication system based on voices of answering incoming phone calls is feasible so that authentication of smartphone's users become easier and unobtrusive.

Keywords: *Implicit Authentication, Smartphone Security, Speaker Recognition, Signal Processing*

1. INTRODUCTION

Recently, smartphone users keep sensitive data and make private transaction via smartphone. Thus, authentication for smartphone becomes very crucial aspect and getting many research interests. Authentication methods for smartphone have comprised many methods. When relying on secret knowledge, smartphones use secret knowledge like password either from alphanumeric PIN to graphical click-based or pattern password [1]. However, using secret knowledge, the authentication system imposes burden to users to remember the secret knowledge. On the other hand, some authentication system use idea of authenticating someone based on one's own characters either physical human traits, such as fingerprint and face, or behavioural human traits, such as gait, and phone usage [2]. There are few smartphones incorporating physical biometric features like those reported in [3] and [4].

From user perspective, the authors in [5] reported in a survey that users want increased security authentication that is transparent when authenticating users for their convenience. Another survey [6] also mentions about 60% of respondents wish to have easier form of mobile authentications. That survey sequentially elaborates the inconvenience of common authentication system like password in tiny keyboard of current smartphone among users having 'fat' fingers.

Speaker recognition is the process of automatically recognizing who is speaking by using speaker-specific information included in the voice waves. It can be classified into speaker identification and speaker verification. Identification is the process of determining from which of the registered speakers a given utterance comes. Verification is the process of accepting or rejecting the identity claim of a speaker. Another classification used is in respect of the utterance spoken by the speaker: the recognition can be text-dependent or text-independent. One of the most important advantages of voice biometrics is that the voice signal can be captured, with quality, by almost any device that contains a microphone. This means that the authentication process can be conducted in any place with a smart phone.

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In this paper, we present our study on the approach to implicitly and transparently authenticate smartphone users using voice from the way user responses to incoming phone calls. This approach is classified as text-independent speaker verification. Through this approach, we aim on (1) alleviating burden of users from remembering secret knowledge, (2) being inexpensive in deployment, (3) being safer to if compared to physical biometrics, and lastly (4) being unobtrusive and transparent to users.

The rest of this paper is organized as follows. Section 2 surveys related work in the area of speaker recognition and speaker authentication. Section 3 introduces a novel method for smartphone authentication using a voice from answering incoming phone calls. Section 4 presents an experiment and result of the overall findings. Lastly, section 5 summarizes and concludes our study and presents our future research direction.

2. RELATED WORKS

There are many works related to the speaker recognition using smartphones. Table 1 shows the summarized information from recent research works.

Based on Table 1, Mel-frequency Cepstral Coefficients (MFCCs) feature was the most popular features used to extract information from voice signals. There is a work in [12] that explores the usage wavelet transform to replace the discrete cosine transform in MFCC feature calculation. They argued that the resulting parameter for Mel-frequency Discrete Wavelet Cepstral Coefficients (MFDWCs) is more in line with human auditory characteristics. Abdulrahman Alarifi et al. [14] had proposed liner predictive coding (LPC) and perceptual linear prediction (PLP) as features. Meanwhile, Wen-Shiung Chen and Jr-Feng Huang [22] had presented a speaker recognition that combines a non-linear features, named spectral dimension (SD), with MFCC.

Table 1. Speaker Recognition Systems Using Smartphones.

References	Years	Features	Classifiers	Mobile implementation
[7]	2013	MFCC	HMM	No
[8]	2012	MFCC	GMM	No
[9]	2012	MFCC	kNN, VQ, GMM, SVM	Input
[10]	2012	MFCC	PLDA	**Yes**
[11]	2012	MFCC	VQ	No
[12]	2012	MFDWC	SVM, DTW	**Yes**
[13]	2012	MFCC	GMM	**Yes**
[14]	2012	MFCC, LPC, PLP	SVM	Input
[15]	2012	MFCC	VQ	**Yes**
[16]	2012	MFCC	VQ	No
[17]	2012	MFCC	GMM	No
[18]	2011	MFCC	GMM	Input
[19]	2010	MFCC	ANN	No
[20]	2010	MFCC	GMM, SVM	No
[21]	2010	MFCC	GMM	**Yes**
[22]	2009	MFCC, SD	GMM	No
[23]	2004	MFCC	VQ	**Yes**

Gaussian Mixture Models (GMM), Support Vector Machine (SVM), and Vector Quantization are the major classifiers used for speaker recognition. Other algorithms which are proposed are Hidden Markov Model (HMM) [7], k-Nearest Neighbors (kNN) [8], Probabilistic Linear Discriminant Analysis (PLDA) [10], Dynamic Time Warping (DTW) [12] and Artificial Neural Network (ANN) [19]. In term of mobile implementation, some works have done a complete implementation with input and processing using smartphones. However, some works just use the smartphone as the input terminal to get the voice data. Databases are also being used for the experiments instead of collecting the voice data manually. The databases are MIT Device Speaker Verification [8], MOBIO (Mobile Biometry) [10], [20], [21], PDAm data set [11], IITG-Multi-Variability (MV) [13], English Language Speech Database for Speaker Recognition (ELSDSR) [16],

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Indian Institute of Technology, Kharagpur – Simulated Emotion Speech Corpus (IITKGP: SESC) [17], TIMIT [19], and AURORA 2.0 [22].

3. ANSWERING INCOMING CALLS FOR AUTHENTICATION

The basic functionalities of smartphones are making/receiving call, and sending/receiving short messages. Users perform the actions almost every day during their daily activities. This research is interested with the way users answering incoming phone calls. NQ Mobile reported in January 2012 survey [24] that making phone call is the most frequent activity that user does. Different users answer incoming phone calls differently based on their current context. As an example, users in this country answer incoming phone calls with different word such as "Hello", "Assalammualaikum", "Wei", "Yes", etc.

Answering incoming call has become another way of authenticating smartphone users implicitly. A. Fahmi et al. [25] used the arm's flex (AF) for authentication system using smartphone. AF is defined as the way users bending their arm for picking a phone when responding to incoming calls. Their study shows that AF is unique and has discriminant power to distinguish one user from others. Another work by A. Fahmi et al. [26] explored the ear image of the smartphone users for authentication system. The ear image is captured using front smartphone's camera while the user is answering an incoming call. In this paper, we propose a method of authenticating smartphone users based on the way users answer an incoming calls. The voice signals from answering incoming calls have discriminant power between different users. Fig. 1 shows voice signal when user answering an incoming phone call.

Mel-frequency Cepstral Coefficient (MFCC) is a feature extraction technique for speech signals in which the extraction of its coefficients is similar to human hearing system. MFCC is a set of linear discrete cosine transforms of the real logarithm of the short-term energy spectrum expressed on a nonlinear Mel-frequency scale. MFCC is a compact and effective representation of the acoustic signals since it takes into consideration the characteristics of the human vocal system [27]. Since MFCC is the most popular and commonly used feature in speaker recognition, Figure 2 shows the MFCCs of voice signal when answering incoming call for two users. The values in the graph can be easily clustered into two groups that represent different user. Thus, the voice signal from answering incoming phone calls could be used as another authentication factor.

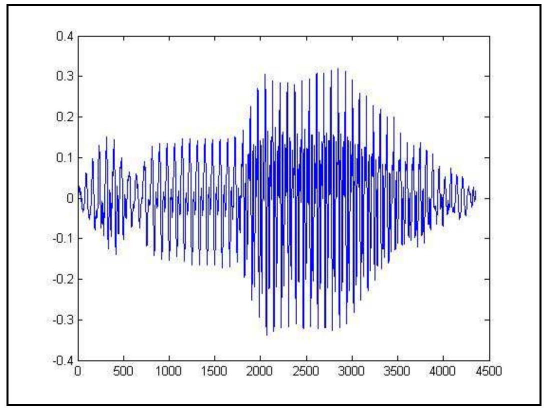

Figure 1. Graph Of Voice Signal During Answering Incoming Call For Two Users.

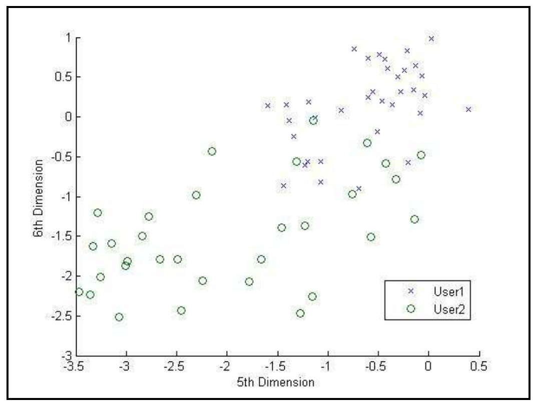

Figure 2. Graph Of Mfccs For Two Users Based On Answering Incoming Call Voice Signals.

4. EXPERIMENT

4.1. Data Set

The experiment is conducted with Samsung Galaxy Wonder which run Android OS (Gingerbread 2.3.3). Several voice data with frequency 16000 Hz were recorded from 9 persons that answering incoming phone call from unknown caller. The recorded voice data contains only the first word spoke by the person in response to the incoming phone call. Each person is recorded for 10 times resulting data in total of 90 voice data. The total voice data is split into 50% as training data and remaining half as testing data. Each user will be tested with a total of 45 voice data, where 5 voice data from himself and other 40 voice data from other users (as imposters with 5 voice data per person respectively).

4.2. Authentication Framework

The whole framework of the smart phone authentication using voice signal from answering

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

incoming call is illustrated in Figure 3. The authentication experiment is conducted in personal computer with Intel Core i3 3.30GHz and 4.00GB of RAM. The application is develop in MATLAB [28] and VOICEBOX [29], an open-source speech processing toolbox.

There are two phases in authentication system which are training phase and testing phase. During training phase, the voice signal is pre-processed with end-point detection modules. End-point detection module is used to determine the start and end points of the first word when user answers incoming call in environment of noise. Another module is to reduce the noise of environment in the voice signal. Then MFCC features are extracted from the voice signal. The MFCC calculation used 12 coefficients and 256 sample frames. For speaker modeling, vector quantization (VQ) with 16 codebook length is used to reduce the dimension of MFCC features and generate a user template. VQ is selected because of the easiness in implementation.

In testing phase, another voice signal is used. The voice signal will go through the same pre-processing and feature extraction processes. The resulted MFCC features will be compared with users' template using Euclidean Distance. The formula used to calculate the Euclidian Distance between two MFCC feature vectors, cb_m and cb_n is given as:

$$D_{mn} = \sqrt{(cb_m - cb_n)^2}$$

The decision of the owner of the voice signal will be is based on the minimum distance.

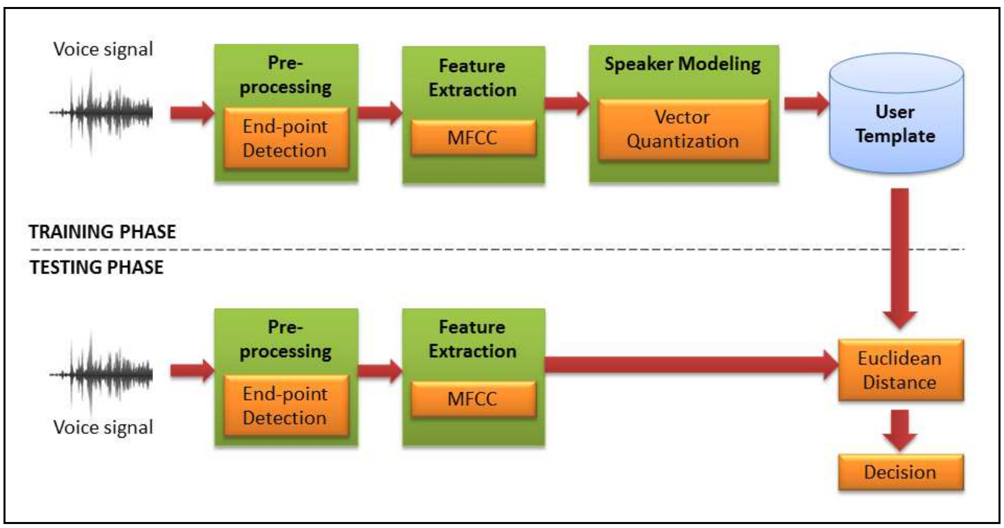

Figure 3. The Framework Of Authentication Based On Answering Incoming Call Voice Signals.

4.3. Result and Evaluation

After training phase, a codebook of each user will be generated and used for testing phase. Table 2 shows the average Euclidean distance of codebook produced by VQ between different persons. The distances between users are very stable and suitable for recognition and authentication system.

The authentication test is repeated for 10 times with different combination of training and testing voice files. The evaluation of the authentication system is based on value of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). Table 3 shows the result of the authentication system for each user. The authentication system has a very good performance with 98.9% accuracy.

Recognition time plays an important role in authentication system since the recognition system should be executed in mobile platform. Thus the experiment measured the recognition time for each user. Figure 5 shows the recognition time in millisecond for each person. The average recognition time is very promising with 12.7 msec.

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2. The Average Euclidean Distance Of Codebook Between Users.

Users	U1	U2	U3	U4	U5	U6	U7	U8	U9
U1	0.00	10.32	11.08	9.99	12.87	12.98	12.97	13.39	13.53
U2	11.66	0.00	13.81	12.86	15.11	14.04	16.33	14.66	15.42
U3	13.74	14.92	0.00	14.50	13.25	15.80	14.24	15.67	15.70
U4	10.49	12.49	13.58	0.00	13.83	11.55	13.71	14.24	14.66
U5	13.71	15.72	13.23	13.69	0.00	14.80	16.06	16.78	15.87
U6	15.13	15.09	16.09	12.64	15.48	0.00	14.76	18.18	16.66
U7	13.13	15.92	12.76	10.72	12.75	11.28	0.00	17.75	16.64
U8	14.87	16.18	16.41	14.40	17.71	18.58	18.82	0.00	16.12
U9	14.09	15.75	17.23	14.44	13.03	16.22	17.43	14.84	0.00

Table 3. The Result Of Authentication System.

Users	TP	TN	FP	FN
U1	100.0%	100.0%	0.0%	0.0%
U2	96.7%	100.0%	0.0%	3.3%
U3	96.7%	100.0%	0.0%	3.3%
U4	100.0%	99.2%	0.8%	0.0%
U5	100.0%	99.6%	0.4%	0.0%
U6	100.0%	100.0%	0.0%	0.0%
U7	100.0%	100.0%	0.0%	0.0%
U8	96.7%	100.0%	0.0%	3.3%
U9	100.0%	100.0%	0.0%	0.0%
Average	98.9%	99.9%	0.1%	1.1%

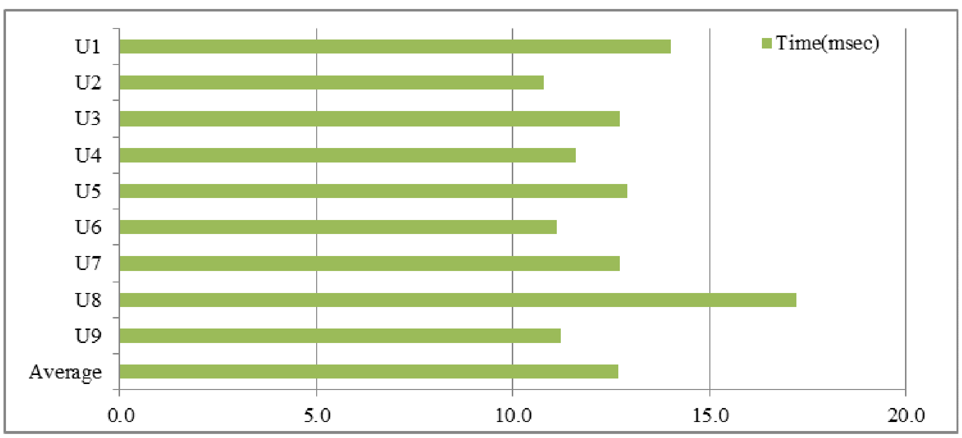

Figure 5. Recognition Time (In Msec) For Each Person And The Average.

5. CONCLUSIONS

We have presented a novel approach for implicit authentication using voices from answering incoming phone calls. The authentication system extracted the MFCC features from the voice signals. VQ is used to classify the users' voices with the templates. Although the results from the experiment show very good accuracy and performance, we still feel the necessary to improve the authentication system with larger data sets. A fusion of multiple modal of biometric, such as voice and face, could improve the accuracy of the authentication system. Thus, the future work would be to implement a multimodal authentication system using smartphone with larger data set.

REFRENCES:

[1] S. Chiasson, R. Biddle, and P. C. van Oorschot, "A second look at the usability of click-based graphical passwords," *Proceedings of the 3rd symposium on Usable*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

privacy and security - SOUPS '07, 2007, pp. 1.

[2] E. Shi, Y. Niu, M. Jakobsson, and R. Chow, "Implicit Authentication through Learning User Behavior," *Lecture Notes in Computer Science*, Vol. 6531, 2011, pp. 99–113.

[3] Android Developers, "Enhancing Security with Device Management Policies." [Online]. Available: http://developer.android.com/training/enterprise/device-management-policy.html. [Accessed: 01-Sep-2013].

[4] Sieger, Hanul, Niklas Kirschnick, and Sebastian Möller, "Poster: User preferences for biometric authentication methods and graded security on mobile phones," *Symposium on Usability, Privacy, and Security (SOUPS)*. 2010.

[5] Furnell, Steven, Nathan Clarke, and Sevasti Karatzouni, "Beyond the pin: Enhancing user authentication for mobile devices," *Computer fraud & security*, Vol. 2008, Issue 8, 2008, 12-17.

[6] Confident Technologies, "Survey Shows Smartphone Users Choose Convenience over Security." [Online]. Available: http://confidenttechnologies.com/news_events/survey-shows-smartphone-users-choose-convenience-over-security. [Accessed: 01-Sep-2013].

[7] Agrawal, Shyam S., et al., "A Hidden Markov Model (HMM) based speaker identification system using mobile phone database of North Atlantic Treaty Organization (NATO) words," *Proceedings of Meetings on Acoustics*, Vol. 19, 2013.

[8] Ren, Tsang Ing, et al., "A hybrid GMM speaker verification system for mobile devices in variable environments," *Intelligent Data Engineering and Automated Learning-IDEAL 2012*, Lecture Notes in Computer Science, Vol. 7435, 2012, pp. 451-458.

[9] Kinnunen, Tomi, et al., "Audio context recognition in variable mobile environments from short segments using speaker and language recognizers," *Odyssey 2012-The Speaker and Language Recognition Workshop*. 2012.

[10] McCool, Christopher, et al., "Bi-modal person recognition on a mobile phone: using mobile phone data," *2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2012.

[11] Baloul, Mossab, Estelle Cherrier, and Christophe Rosenberger, "Challenge-based speaker recognition for mobile authentication," *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2012.

[12] C. Liu, W. Liming, H. Wei, L. Wenjie, and L. Qiudi, "Research of an improved algorithm in Speaker Recognition," *2012 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, May 2012.

[13] G. Pradhan, B. C. Haris, S. R. M. Prasanna, and R. Sinha, "Speaker verification in sensor and acoustic environment mismatch conditions," *International Journal of Speech Technology*, Vol. 15, No. 3, 2012, pp. 381–392.

[14] A. Alarifi, I. Alkurtass, and A. S. Alsalman, "SVM based Arabic speaker verification system for mobile devices," *2012 International Conference on Information Technology and e-Services*, 2012, pp. 1–6.

[15] Mayrhofer, Rene, and Thomas Kaiser, "Towards usable authentication on mobile phones: An evaluation of speaker and face recognition on off-the-shelf handsets," *Fourth International Workshop on Security and Privacy in Spontaneous Interaction and Mobile Phone Use (IWSSI/SPMU)*,2012.

[16] Martinez, Jorge, et al., "Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques," *22nd International Conference on Electrical Communications and Computers (CONIELECOMP)*, 2012.

[17] S. G. Koolagudi, S. E. Fatima, and K. S. Rao, "Speaker recognition in the case of emotional environment using transformation of speech features," *Proceedings of the CUBE International Information Technology Conference on - CUBE '12*, 2012, pp. 118.

[18] Lu, Hong, et al., "SpeakerSense: energy efficient unobtrusive speaker identification on mobile phones," *Pervasive Computing*, 2011, pp. 188-205.

[19] K. S. Rao, A. K. Vuppala, S. Chakrabarti, and L. Dutta, "Robust speaker recognition on mobile devices," *2010 International Conference on Signal Processing and Communications (SPCOM)*, 2010, pp. 1–5.

[20] Marcel, Sébastien, et al., "On the results of the first mobile biometry (MOBIO) face and speaker verification evaluation," *Recognizing*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Patterns in Signals, Speech, Images and Videos, 2010, pp. 210-225.

[21] Tresadern, P. A., et al., "Mobile biometrics (mobio): Joint face and voice verification for a mobile platform," *IEEE Pervasive Computing*, 2012.

[22] W.-S. Chen and J.-F. Huang, "Speaker Recognition Using Spectral Dimension Features," *2009 Fourth International Multi-Conference on Computing in the Global Information Technology*, 2009, pp. 132–137.

[23] Saastamoinen, Juhani, et al., "Automatic speaker recognition for series 60 mobile devices," *9th Conference Speech and Computer*, 2004.

[24] NQ Mobile and NCSA, "Report on Consumer Behaviors and Perceptions of Mobile Security," 2012.

[25] A. Fahmi, P. Negara, E. Kodirov, M. Fikri, A. Abdullah, D. Choi, G. Lee, and S. Sayeed, "Arm ' s Flex when Responding Call for Implicit User Authentication in Smartphone," *International Journal of Security and Its Applications*, Vol. 6, No. 3, 2012, pp. 55–64.

[26] P. N. Ali Fahmi, E. Kodirov, D.-J. Choi, G.-S. Lee, a Mohd Fikri Azli, and S. Sayeed, "Implicit authentication based on ear shape biometrics using smartphone camera during a call," *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2012, pp. 2272–2276.

[27] P. Davis, S., Mermelstein, "Comparison of Parametric Representations for," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, 1980, pp. 357–366.

[28] MathWorks, "MATLAB." [Online]. Available: http://www.mathworks.com/products/matlab/. [Accessed: 01-Sep-2013].

[29] M. Brookes, "VOICEBOX: Speech Processing Toolbox for MATLAB." [Online]. Available: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html. [Accessed: 01-Sep-2013].