Global Journal of Enterprise Information System

DOI: 10.18311/gjeis/2017/15973

Automatic Speaker Recognition: Current Approaches and Progress in Last Six Decades

Nilu Singh*, Alka Agrawal and R. A. Khan

SIST-DIT, Babasaheb Bhimrao Ambedkar University (Central University), Lucknow, Uttar Pradesh, India; nilu.chouhan@hotmail.com, alka_csjmu@yahoo.co.in, khanraees@yahoo.com

Abstract

Automatic speaker recognition is the process to recognizing speaker automatically by their speech/voice on the basis of specific characteristics of his/her speech signal. These voice specific characteristics are called speech features. Over the past six decades many recent advances in the area of speaker recognition have been achieved, but still many problems remains to be solved or require better solutions. The main problems in speaker recognition are session variability, channel mismatch and recording conditions of voice. To develop an efficient speaker recognition system it needs to examine stable parameters of voice features parameters over time, unaffected from variation in speaking, background noise, channel distortion and robust against variation of physical problems. This paper overviews recent advances and general ideas of speaker recognition technology.

Keywords: Advancement of Speaker Recognition, Principle of Speaker Recognition, Speaker Recognition, Speech Features

Paper Code: 15793; **Originality Test Ratio:** 8%; **Submission Online:** 01-May-2017; **Manuscript Accepted:** 05-May-2017; **Originality Check:** 08-May-2017; **Peer Reviewers Comment:** 03-June-2017; **Double Blind Reviewers Comment:** 14-June-2017; **Author Revert:** 29-June-2017; **Camera-Ready-Copy:** 09-July-2017

1. Introduction

In today's life security is indeed in great demand. It may be individual, organizational or country base. In this contest, security biometric plays an important role and provides a solution to maintain high security level. Speaker recognition is a biometric recognition technique; it can be decomposed as bio and metric. Bio represents life & metric represents measures. In broad sense metric completely focuses on measuring the property of creature (derived from the greek words). More specifically biometrics is the technology for measuring and analyzing human's behavioral or physiological individuality. These techniques can be used to recognizing a person on the basis of his/her voice, face, iris, DNA, signature, retina scan, fingerprint, hand geometry etc.[1-2,6]. Recent year's biometric authentication has shown a significant technology, progress. Popularity of this technology lies behind the fact that it is less prone to attacks.

Human voice (speech signal) contains different types of information making it a strong candidate for authentication. A speech signal uttered by a person is able to identify person. Figure 1 shows category of recognition through a speech signal. By using a speech signal mainly three kinds of recognition are performed; speech recognition (what is spoken), speaker recognition (who is speaking) and language identification (identifying the spoken language by the speaker). Speaker recognition is again categorized as speaker identification and speaker verification. Speaker verification is one to one (1:1) matching system whereas speaker identification is one- to- n (1: n) matching system. In speaker verification, claimed identity is matched against specific speaker's voice model while in speaker identification system tries to match an unknown speaker against the entire voice database. Speaker recognition is again divided into text-dependent and text-independent. Text-dependent system requires providing the same text/utterance for training and testing while text-independent system does not depend on specific text[3-5].

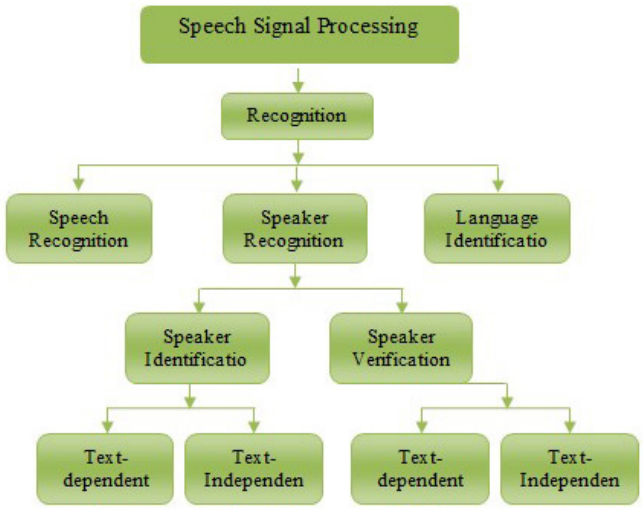

Figure 1. Origin and Categorization of Speaker Recognition.

*Author for correspondence

The rest of the paper is organized as: the next section describes about basic terminology used in speaker recognition. In section 3 discussions are made about principle of speaker recognition system and section 4 presents the advancement in speaker recognition technology. In section 5 discussions is made about the factors which affect the system performance. Section 6, discusses about biometric techniques that how to decide which one is better. Finally paper concludes at section 7.

2. Basic Terminology used in Speaker Recognition

Speaker recognition and speech recognition are speech signal based authentication technology. Lots of functions are responsible during voice generation and voice contains many components which are used as a well-known parameter for voice. These voice parameters is used to measure the voice characteristics. In speaker recognition technology some common terms are used frequently, which are given below.

2.1 Dialect/ Accent

Dialect is a vocabulary or language spoken by specific group of people; it is also known as regional languages. It is a common style of pronunciation in a particular region or country[7].

2.2 Acoustic

Concerns to the sense of hearing[8].

2.3 Formants

Formant of speech signal is compactness of acoustic energy for a specific frequency in the speech signal. The formant is changes as the voice frequency is changes normally it occurs at 1000 Hz intervals[9].

2.4 Syllable

It is a segment of speech or whole word from which a word can be separated generally containing a vowel[10].

2.5 Articulation

It is a way of speaking that is how the association of speech organs such as tongue, lips, and jaw etc. to make speech sounds[11].

2.6 Utterance

Utterance is a smallest unit of speech of spoken language. It is a normal pause (bounded by breaths) in the start and end of continuous speech[12].

2.7 Phoneme/Linguistics

It is a unit of sound by which we can distinguish one word from another in an individual spoken language[13].

2.8 Intonation

Intonation of speech is concern about the variation in pitch/tone, sometimes stress and rhythm also consider[14].

2.9 Tone/Pitch

Ups and downs occur in speech signal. It is the frequency perceived by the human ear. For example we perceive higher pitch if frequency is higher and perceive lower pitch if frequency is lower[15].

2.10 Paralinguistic features of voice

Generally paralinguistic features are those that is not words such as facial expression, tone/pitch, body language and gestures[13].

2.11 Timbre

It is defined as the distinctive property of a complex sound, or also says that distinguish sound (musical) from one to another even they have the same loudness and pitch[16].

2.12 Voice ensity (Vocal/coustic ensity)

It is perceived as the loudness of the sound. Intensity of voice is a measurement of radiated power (energy produced and radiated into the close air, per second measured in watts) per unit area. Intensity depends on the sound source for example intensity decreases as the distance increases from the sound source[17].

2.13 Volume

It is an arbitrary term for the amount of sound which is perceived by an average listener and measured in terms of acoustic power or intensity[17].

2.14 Voice Frequency

Voice frequency is an audio range which is used for the transmission of speech. The frequencies (humans to hear through the air as sound) of the vibrations must occur between 20 to 20,000 Hz[18,19].

2.15 Fundamental Frequency

It is the lowest frequency in a periodic waveform also known as first harmonic frequency[20]. Average Fundamental Frequencies-Children: 500 Hz

Women: 250 Hz
Man: 130 Hz

2.16 Loudness

The loudness is defined as that it is a perceptual quantity which can only be evaluated by a frequencies because loudness fluctuates according to pitch. For example the human ear is perceived pitch in range of 1000-3000 Hz[17].

2.17 Jitter

Cycle to cycle variability in fundamental frequency[20].

2.18 Shimmer

Cycle to cycle variability in amplitude[20]

2.19 Speed of voice

The rate of change of distance with time and the magnitude of velocity[20].

3. Principle of Speaker Recognition

In the modern digital era where insecurity is prevailing everywhere maintaining security is a big challenge. Lots of cases are being reported in daily life related to edit audio clips and wrong claim for identity. Speaker recognition is a technique to automatically recognizing a speaker on the basis of information extracted by his/her speech. It can be divided into two categories; speaker identification and speaker verification. This method provides security in confidential areas. For example, to prove the claimed identity of a person, his/her voice is treated through forensic test[3]. This technique is very useful to authenticate a person's identity. The aim of automatic speaker recognition is to acquire the voice of speakers and to create voice model for each speaker and finally to compares these models with an utterance of the speaker to prove his/her identity. Different individuals have different voice. Even voice of a person may differ time to time. The variation in different people's voices is termed as inter-speaker variability and the variation in the same person's voice is termed as intra- speaker variability[4,5,20]. The speaker recognition relies on the ability of human being to identify other person's voice with the following observations:

Human being is able to recognize the voice of any person (whom he/she knows and communicated to each other frequently).

A person is able to identify other person's voice to whom he/she communicate frequently irrespective of the communication medium or background noise.

A person is able to recognize other person even if communication happens after a long gap (even years).

A human is able to identify the 'state of mind' person is also able to recognize the 'state of mind' (emotions level that is speaker is happy, sad, neutral, cold, some health issue etc.) of the speaker by listening his/her voice.

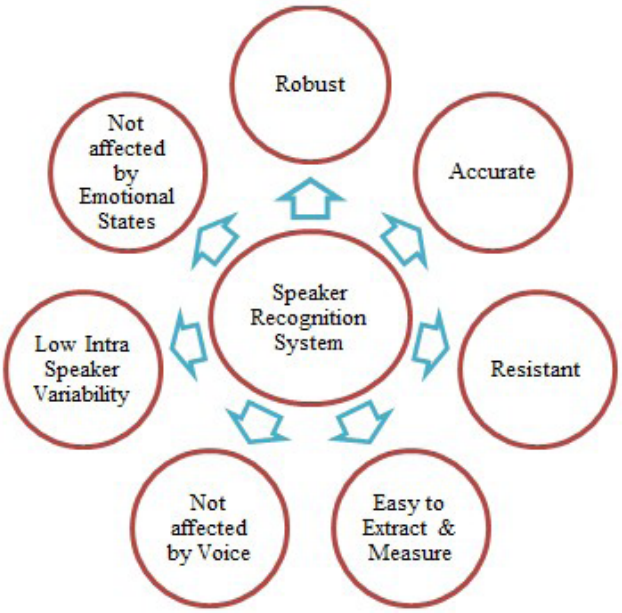

Figure 2. Characteristics of a robust speaker recognition system.

A strong speaker recognition mechanism must focuses on the factors on the basis of which a person recognizes the voice of other persons. If it can be known that how human recognizes the voice of any person (whom he/she communicate frequently) then this will help to make a robust speaker recognition system which is more and more accurate. Figure 2 shows the characteristics of speaker recognition system. A speaker recognition system should be robust, accurate, resistant, easy to extract and measure, not affected by voice, low intra-speaker variability and not affected by emotional states etc.

4. Advancement in Speaker Recognition

Advancement in the area of ASR 1974 to 2016 has been shown on the basis of different parameters. The summary is presented in Table 1, the terms defined in columns in Table 1 are: "Developer/Author/year" refers to who developed and used the particular techniques, "Organization" is the lab or company or institution where the work has been done, "Database (Population)" is the number of speakers in which test has been conducted for speaker

verification or identification, "Features extraction techniques" that is refers to the technique used to measure speech signal features, "Modeling" refers to the method is used for the matching of signal, "Voice type" shows how the voice is acquired such as telephone, lab, noisy place etc., "text-type" means the system is text-dependent or text-independent, "accuracy" sows that how much the system is accurate for recognition. The complete information in the table gives a general overview of speaker recognition research from 1974 to 2016. To be focused here have taken some selected & significant studies.

Speaker recognition system can be used in access control, telephone banking, biometric investigation, crime investigation etc. There is a number of commercial/organizational/personal automatic speaker recognition system including T- NETIX, ITT, Lernout & Hauspie, Veritel and voice control system. Studies says that, sprint's voice FONCARD" is the largest scale deployment of any biometric system till date[1,2,7]. It is very difficult to make a meaningful comparison between text-dependent and text-independent speaker recognition system in absence of standard comparison criteria. As there are different techniques dealing with different recognition problems so it is not easy to decide which one is better. For instance Gish's segmental 'Gaussian model' and Reynolds"

'Gaussian Mixture Model' for text-independent approaches are used to deal with unique problems e.g. sounds or articulations present in the test signal, but not in training voice signal[7].

During the literature survey, it was found that the following areas of speaker recognition have been gaining great attention in terms of research:

Accuracy of speaker recognition system[2,14,18].

Development of more robust system for speaker recognition[16]

Development of feature extraction techniques for voice feature extraction[19]

Different speaker modelling techniques for speaker identification and verification etc. [13,15]

Table 1 summarizes few relevant research works in the area. Of course, these works have their own worth. Nobody can deny the importance of these works. But still there exist many questions which are still unanswered:

Is there any standard way available to decide about the number of voice parameters?

Which voice parameters are essential to include during the development of speaker recognition system?

What is the maximum time limit for voice recording to achieve maximum accuracy?

It is possible to make a robust speaker recognition system in real which is not affected by background noise, session variability, recoding environment etc.?

What and how many speech parameters should be included to develop a robust speaker recognition system?

What factors are more responsible for enhancing the system performance as well as degradation?

The main objective of the work is to focus on the recent advances and development in the area of speaker recognition and the problems still remains unanswered.

Table 1. Progress in Speaker Recognition in last six decades (some selected)

Developer/ Author/year	Organization	Database (population)	Features Extraction/ Modeling/ Matching Method	Features	Voice type	Text type/ system type	Accuracy (%)
F. K. Soong, et.al./1985 [8]	AT&T Bell Laboratories	50 male and 50 female)	vector quantization (VQ)	short-time spectral	Telephone	Independent	98%
B. S. Atal/ 1974 [9]	Bell laboratories	10 speakers	LPC	Cepstrum	Lab	Independent	93%
Colombi, et al./199634	AFIT	138	HMM monophone	Cepstrum	office	Dependent	Error: identification 0.22% (10s) verification 0.28% (10s)
Alfredo Maesa/2012[24]	Voxforge. or g	450 speakers	MFCC	spectral subtraction	Audio data- base	Independent/ Identification	>96%
Douglas A. Reynolds/1995[2]	Lincoln Laboratory	49	GMM	Short Utterance	Telephone	Independent/ Identification	96.8%
Rabah W et.al./2004[10]	King Abdulaziz University	20	SVD-based algorithm	LPC/ Cepstral	office	Independent/ Identification	94%

Najim Dehak et.al./2007 35	NIST-2006	NA	GMM-JFA	prosodic features	Lab	language identification	Improvement 8% (all trials) and 12% (English only
Sharada V. Chougule/2015[11]	Finolex Academy of Management & Technology	97	NDSF	Spectral	Lab	Independent/ Identification	~(98-100)%
Yang Shao et.al./2008[12]	Ohio State University	34(18 male 16 female)	GFCCs	auditory features	Telephone	Independent/ Identification	~99.33%
Vincent Dubreucq/1994[13]	Digital speech laboratory, RMA	21	HMM	Pitch	Lab	Independent/ Recognition	VER=7.6% RER=7.7%
Douglas A. Reynolds/2001[18]	TIMIT(168), NTIMIT(168), Switchboard (113)	449	GMM	Unconstrained speech	Lab	Dependent/ Recognition	99.7%, 76.2%, 82.8%
Rabah W et.al./2003[14]	King Abdulaziz University	10	SVD-based algorithm	LPC/Cepstral	office	Independent/ Identification	99.5%
P. Krishnamoorthy/ 2011[15]	TIMIT	100	GMM-UBM	MFCC	Lab	Independent/ Identification	80%
Sriram Ganapathy/2014[16]	SRE database (NIST-2010)	random	AR model	FDLP	Lab	Dependent/ Recognition	relative improvements of up to 25%
Hesham Tolba/2011[23]	Arabic speakers	10	HMM/ GHMM	MFCC	Lab	Dependent/ Independent	100%/80%
Chih-Hung Chou et. Al./2015[25]	ALTERADE2-70,	16	VQ/GMM-PQ	OOS	Lab	Dependent	Recognition Rate 88.3%
Emmanual Perrin et. Al./1994[26]	E-HERRIOT	60	Acoustical Signature	Vocalic Space	Standard Protocol	Dependent	>90%
Ergun Yucesoy et al./2016[27]	E Gender database INTER SPEECH 2010	299 speakers	GMM-SV	prosodic features	Lab	Dependent	90.4%, 54.1% and 53.5% in gender, age, and age & gender categories
Xuanjing Shen et al./2014[28]	TIMIT speech database	38(19 Female, 19 Male)	LFA-SVM Gaussian kernel	12-order MFCC,	Lab	NA	81.52%
Anzar S.M et al./2016[29]	English language data base for adaptive speaker recognition (ELDASR)	50(Male/ Female)	GMM/ MFCC	MFCC super template	Lab with intra-class variations	NA	Improved (% NA)
Isaias Sanchez-Cortina et al./2016[30]	video Lectures. net, poliMedia	NA	logistic regression model	NB model	Online educational lectures	Dependent	Relative improvement between 2% and 7%.

* RMA: Royal Military Academy
* NDSF: Normalized Dynamic Spectral Feature
* VER: Verification error rate
* RER: Rejection error rate
* FDLP: frequency domain linear prediction
* SRE :NIST-2010 speaker recognition evaluation database
* AR Model: Auto Regressive Model
* NB Model : word-dependent naïve Bayes (NB)
* JFA: joint factor analysis

Speaker recognition is one of the emerging research domain for persons authentication and enhances the security in the areas

including access control, voice authentication, banking by telephone and many more[5,17]. It is very difficult to find the fix voice parameters by which a good speaker recognition system with maximum accuracy can be developed. Therefore to design and develop robust speaker recognition system, continuous effort is needed. Speaker recognition technology has many advancement and development till date but technology development and evaluation are two sides of the same coin. So keeping this point in mind it can be concluded that without having a good measure of progress nobody can make valuable progress[5]. Till date various investigations have been proposed for evaluation of speaker recognition but in real a complete tool has not yet been developed.

5. Factors Affecting the Performance of Speaker Recognition Systems

For speaker recognition systems there are many challenges which occur at the time of data acquisition. During data acquisition inside lab, there must be complete quite when users articulate for enrolling on the system and must record his/her voice more than ones. For better performance, it is expected from speakers that they provide their voice recording at an interval of time such as a week or a month. This is called session variability and it is useful to model slight changes in speaker's voice[31,32]. Variations in speaker's voice may be due to the reasons e.g. the speaker may be stressed; speaker possibly may suffer from cold. The main factor which more affects the performance of system is quality and quantity of training data. The other reasons may include environmental factors such as channel variation, type of handset, background noise etc. Many researchers[31-33] have discussed that one of the most severe problems during speaker recognition is Intra-speaker variability of speech features. Performance of speaker recognition system is affected by many factors including:

Quality of speech recorded;
Environmental condition during the speech recoding;
Type of microphone used;
Transmission channel bandwidth (landline & cell phone);
Physical and emotional states of the person;
Session variability;

The above mentioned factors should be taken into account during the design & development of speaker recognition system or while comparing two different system performance. System performance is excellent when speech is recorded in good conditions including high quality microphone, quite environment and training and testing session[5,17].

6. Which Biometric Technique is better

There are many biometric techniques including voice, iris, fingerprint, face recognition, DNA, Signature, retina scan and hand geometry etc. Hence it is very common question asked by many that which one is better. The answer lies in the fact that it is very difficult to compare one technique (biometric) with the other as there are so many factors and on the basis of these factors biometrics is evaluated. These factors include efficiency, accuracy, uses, accessibility, cost etc. Hence there is not a complete approach by which comparison of biometric is possible. Each biometric have its own pros and cons. Hence no one can claim the superiority of any approach. Although sometimes comparison is possible on the basis of specific factors, for example accuracy, ease of access, usability etc. [21,22] However the potential of speaker recognition technology is that it relies on a signal (voice) which is natural and available unobtrusively to acquire without any special equipment or training. The primary use of this technology is for remote system accessibility and forensics. Also it is easy to use and portable and the leading factor is high accuracy[18].

7. Conclusion

Speaker and speech recognition are two different techniques which use speech signal. Speech recognition is used for matching dictation while speaker recognition is used for speaker authentication. This study has discussed the major contributions during last six decades in the field of automatic speaker recognition system. In this paper authors have not performed comprehensive review rather an overview of some selected advances in the area of speaker recognition technology has been given. In addition, problems which need improvement in future have also been discussed. It cannot be denied that in the last six decades there have been significant achievements in the area but still many issues remains unsolved. These issues require urgent attention. For example the issues to develop such speaker recognition system which is robust against background noise and channel mismatch conditions is still unresolved. At last we have discussed some future trends for research and development in speaker recognition technology.

8. References

1. Zheng TF. Speaker Recognition Systems: Paradigms and Challenges.
2. Reynolds DA. Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. IEEE Transactions

on Speech and Audio Processing. 1995; 3(1):72–83. https://doi.org/10.1109/89.365379
3. Furui S. An Overview of Speaker Recognition Technology. ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, Martigny, Switzerland. 1994 Apr; 1–9.
4. Melin H. Speaker Verification in Telecommunication. Department of Speech,
5. Furui S. Speech and Speaker Recognition Evaluation. Evaluation of Text and Speech Systems. Springer. 2007; 1–27. https://doi.org/10.1007/978-1-4020-5817-2_1
6. Speert D. Brain Facts A Primer on the Brain and Nervous System. Society for Neuroscience, United States of America, Sixth Edition. 2006; 1–80.
7. Campbell JP Jr, Speaker Recognition, Department of Defense Fort Meade, MD, pp 1–26
8. Soong FK et al. A Vector Quantization Approach to Speaker Recognition. AT&T
9. Atal BS. Effectiveness of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification and verification. Journal Acoustical Society of America. 1974 Jun; 55(6):1304–12. https://doi.org/10.1121/1.1914702 PMid:4846727
10. Aldhaheri RW, Al-Saadi FE. Robust Text-independent Speaker Recognition with Short Utterance in Noisy Environment Using SVD as a Matching Measure. J. King Saud Univ, Comp and Info Sci. 2004; 17:23–41.
11. Chougule SV, Chavan MS. Robust Spectral Features for Automatic Speaker Recognition in Mismatch Condition. Procedia Computer Science. 2015; 58:272–9. https://doi.org/10.1016/j.procs.2015.08.021
12. Shao Y, Wang DL. Robust Speaker Identification Using Auditory Features and Computational Auditory Scene Analysis. Department of Computer Science and Engineering, Center for Cognitive Science, The Ohio State University, Columbus, ICASSP, 1-4244-1484-9/08, IEEE. 2008; 1589–92.
13. Bimbot F, Mathan L. Second- Order Statistical Measures for Text-Independent Speaker Identification. ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, Martigny, Switzerland. 1994 Apr; 51–4.
14. Aldhaheri RW, Al-Saadi FE. Text-Independent Speaker Identification in Noisy Environment Using Singular Value Decomposition. ICICS-PCM, Singapore, IEEE. 2003 Dec; 1–5.
15. Krishnamoorthy P et al. Speaker recognition under limited data condition by noise addition. Expert Systems with Applications, Elsevier. 2011; 38:13487– https://doi.org/10.1016/j.eswa.2011.04.069
16. Ganapathy S et al. Robust Feature Extraction using Modulation Filtering of Autoregressive Models, IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2014 Aug; 22(8):1285–95.
17. Singh N, Khan RA, Shree R. Applications of Speaker Recognition, International Conference on Modeling, Optimization and Computing (ICMOC–2012). Procedia Engineering, Elsevier. 2012; 38:3122–6. https://doi.org/10.1016/j.proeng.2012.06.363
18. Reynolds DA. Automatic Speaker Recognition: Current Approaches and Future Trends. MIT Lincoln Laboratory, Lexington, MA USA. 2001; 1–6.
19. Hansen JHL, Hasan T. Speaker Recognition by Machines and Humans. IEEE Signal Processing Magazine. 2015 Nov; 74:74–99. https://doi.org/10.1109/MSP.2015.2462851
20. Singh N, Agrawal A, Khan RA. A Critical Review on Automatic Speaker Recognition. Science Journal of Circuits, Systems and Signal Processing. 2015 Jul; 4(2):14–7.
21. Meuwly D. Forensic Individualisation from Biometric Data. Science and Justic . 2006; 46(4):205–13. https://doi.org/10.1016/S1355-0306(06)71600-8
22. Bishop J. Using the concepts of forensic linguistics, bleasure and motif to enhance multimedia forensic evidence collection. The 2014 International Conference on Security and Management SAM"14, Monte Carlo Resort in Las Vegas, Nevada USA. 2014. p. 21–4.
23. Tolba H. A high-performance text-independent speaker identification of Arabic speakers using a CHMM-based approach. Alexandria Engineering Journal. 2011; 50:43–7 https://doi.org/10.1016/j.aej.2011.01.007
24. Maesa A et al. Text Independent Automatic Speaker Recognition System Using Mel-Frequency Cepstrum Coefficient and Gaussian Mixture Models. Journal of Information Security. 2012; 3:335–40. https://doi.org/10.4236/jis.2012.34041
25. Chou C-H et al. A Statistical Out-of-Speaker Detection Approach for Smart Home Voice-Control Scenario of Protective Warming Care on FPGA, ASE BD & SI, Kaohsiung, Taiwan. 2015 Oct; 1–4.
26. Perrin E et al. Phonatory Signature of the Deaf Child", ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, Martigny, Switzerland. 1994 Apr; 201–4.
27. Yucesoy E et al. A new approach with score-level fusion for the classification of a speaker age and gender. Computers and Electrical Engineering (Elsevier). 2016; 53:29–39 https://doi.org/10.1016/j.compeleceng.2016.06.002
28. Shen X et al. A Speaker Recognition Algorithm Based on Factor Analysis. https://doi.org/10.1109/CISP.2014.7003905
29. Anzar SM et al. Efficient online and offline template update mechanisms for speaker recognition. Computers and Electrical Engineering (Elsevier). 2016; 50:10–25. https://doi.org/10.1016/j.compeleceng.2015.12.003
30. Sanchez-Cortina I et al. Speaker-adapted confidence measures for speech recognition of video lectures. Computer Speech and Language (Elsevier). 2016; 37:11–23. https://doi.org/10.1016/j.csl.2015.10.003
31. Yegna Narayana B, Mahadeva Prasanna SR, Zachariah JM, Gupta Prasanna CS. Combining Evidence from Source, Suprasegmental and Spectral Features for a Fixed-Text Speaker Verification System. IEEE Transactions on Speech and Audio Processing. 2005; 13(4):575–82. https://doi.org/10.1109/TSA.2005.848892
32. Nakasone H. Voice Recognition Capabilities at the FBI. Insttute for Defense and Government Advancement. 2014.
33. Zhang X et al. Rapid Speaker Adaptation in Latent Speaker Space Withnon- Negative Matrix Factorization. Elsevier, Science Direct, Speech Communication. 2013; 55:893–908. https://doi.org/10.1016/j.specom.2013.05.001
34. Colombi J, Ruck D, Rogers S, Oxley M, Anderson T. Cohort Selection and Word Grammer Effects for Speaker Recognition. In International

Conference on Acoustics, Speech, and Signal Processing in Atlanta, IEEE. 1996; 85–8.

35. Dehak N et al. Modeling Prosodic Features With Joint Factor Analysis for Speaker Verification. IEEE Transactions on Audio, Speech, And Language Processing. 2007 Sep; 15(7):2095–103 https://doi.org/10.1109/TASL.2007.902758

Annexure-I

AUTOMATIC SPEAKER RECOGNITION: CURRENT APPROACHES AND PROGRESS IN LAST SIX DECADES

ORIGINALITY REPORT

8%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.scribd.com Internet	42 words — 1%
2	membership.sciencepublishinggroup.com Internet	31 words — 1%
3	P. Krishnamoorthy. "Application of combined temporal and spectral processing methods for speaker recognition under noisy, reverberant or multi-speaker environments", Sadhana, 10/2009 Crossref	24 words — 1%
4	Sadaoki Furui. "Speech and Speaker Recognition Evaluation", Text Speech and Language Technology, 2007 Crossref	20 words — 1%
5	Smits, I.. "A Comparative Study of Acoustic Voice Measurements by Means of Dr. Speech and Computerized Speech Lab", Journal of Voice, 200506 Crossref	15 words — < 1%
6	Computational Models of Speech Pattern Processing, 1999. Crossref	11 words — < 1%
7	www.tcsme.org Internet	11 words — < 1%
8	web.cse.ohio-state.edu Internet	11 words — < 1%
9	www.ee.cuhk.edu.hk Internet	10 words — < 1%
10	Dongdong Li. "Emotional Speech Clustering Based Robust Speaker Recognition System", 2009 2nd International Congress on Image and Signal Processing, 10/2009 Crossref	10 words — < 1%
11	Vijayan, Karthika, Pappagari Raghavendra Reddy, and K. Sri Rama Murty. "Significance of analytic phase of speech signals in speaker verification", Speech Communication, 2016. Crossref	10 words — < 1%
12	Aaron Powers, Sara Kiesler. "The advisor robot", Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction - HRI '06, 2006 Crossref	10 words — < 1%
13	Lecture Notes in Electrical Engineering, 2014. Crossref	8 words — < 1%
14	Kinnunen, T.. "An overview of text-independent speaker recognition: From features to supervectors", Speech Communication, 201001 Crossref	8 words — < 1%
15	www.coursehero.com Internet	8 words — < 1%
16	Dhameliya, Kinnal, and Ninad Bhatt. "Feature extraction and classification techniques for speaker recognition: A review", 2015 International Conference on Electrical Electronics Signals Communication and Optimization (EESCO), 2015. Crossref	8 words — < 1%
17	Chung-hsien Yang. "Robust Speaker Recognition using SNR-Aware Subspace-Based Enhancement and Probabilistic SVMs", 2006 IEEE International Conference on Multimedia and Expo, 07/2006 Crossref	8 words — < 1%
18	www.ksu.edu.sa Internet	8 words — < 1%
19	Hemant A. Patil. "LP spectra vs. Mel spectra for identification of professional mimics in Indian languages", International Journal of Speech Technology, 05/19/2009 Crossref	7 words — < 1%
20	Joseph P. Campbell. "Speaker Recognition", Wiley Encyclopedia of Electrical and Electronics Engineering, 12/27/1999 Crossref	7 words — < 1%
21	Encyclopedia of Biometrics, 2015. Crossref	7 words — < 1%
22	Blitzer, Andrew, Mitchell F. Brin, and Lorraine O. Ramig. "Acoustic Assessment of Vocal Function", Neurologic Disorders of the Larynx, 2009. Crossref	7 words — < 1%
23	Fabio Castaldo. "Compensation of Nuisance Factors for Speaker and Language Recognition", IEEE Transactions on Audio Speech and Language Processing, 9/2007 Crossref	6 words — < 1%
24	"Biometrics", Springer Nature, 1999 Crossref	4 words — < 1%

EXCLUDE QUOTES ON
EXCLUDE BIBLIOGRAPHY ON
EXCLUDE MATCHES OFF

Source: http://www.ithenticate.com/

Citation:

Nilu Singh, Alka Agrawal and R. A. Khan

"**Automatic Speaker Recognition: Current Approaches and Progress in Last Six Decades**",

Global Journal of Enterprise Information System. Volume-9, Issue-3, July-September, 2017. (http://informaticsjournals.com/index.php/gjeis)

DOI: 10.18311/gjeis/2017/15973

Conflict of Interest:

Author of a Paper had no conflict neither financially nor academically.