2016

A Brief Review of Speaker Recognition Technology

Clark D. Shaver
Portland State University

John M. Acken
Oklahoma State University

Citation Details

Shaver, Clark D. and Acken, John M., "A Brief Review of Speaker Recognition Technology" (2016). *Electrical and Computer Engineering Faculty Publications and Presentations*. 350.
http://pdxscholar.library.pdx.edu/ece_fac/350

A Brief Review of Speaker Recognition Technology

Clark D. Shaver[1] and John M. Acken[2]

1. Department of Engineering Technology, Pittsburg State University, Pittsburg, KS 64759 USA e-mail: cshaver@pittstate.edu,
2. Department of Electrical Engineering, Oklahoma State University, Stillwater OK 74078 USA e-mail: john.m.acken@okstate.edu.

ABSTRACT

This paper reviews the development of speaker recognition systems from pre-computing days to current trends. Advances in various sciences which have allowed autonomous speaker recognition systems to become a practical means of identity authentication are also reviewed.

1. INTRODUCTION

A Speaker Recognition (SR) system measures the attributes of a person's voice or speech in order to make an assessment regarding that person's identity. Though the task is common for humans to perform, i.e. recognizing a voice on the phone, autonomous SR tasks are difficult. Autonomous SR systems have had measured successes and setbacks throughout the years. Major advances throughout the last five decades have helped overcome many major challenges for SR systems. Today's systems provide a practical means of verifying user access rights, identifying personnel in a group and even limited use in forensic applications.

The earliest research in SR was in the realm of human abilities. Later, war time research allowed for significant advances in autonomous systems, producing a tool to allow visual inspection of voice. Advances in signal processing and the advent of the computer permitted true autonomous systems to be developed. Despite some limitations, certain applications have made sufficient advances to make commercial systems a reality. This article illustrates how recognition systems have advanced throughout the years and identifies current and future research trends in this field.

2. EARLY BEGINNINGS

The problem of recognizing an individual by their voice is an age old issue. Genesis records Isaac's dilemma in verifying a speaker when Jacob acts as an imposter of his brother Esau. Isaac's confusion was with two contradictory biometrics. "The voice is Jacob's voice, but the hands are the hands of Esau." Jacob trusted tactility over auditory "and he discerned him not." (Gen. 27:22-23) The speaker recognition problem appears in a judicial case as early as 1660 [1]. A couple of centuries later, academic research would begin investigating voice biometrics.

In March of 1932, Charles and Anne Lindbergh's baby boy was abducted and subsequently killed. The investigation led to a clandestine payoff in a cemetery where a Lindbergh operative met with an anonymous male claiming to be the kidnapper. Charles Lindbergh sat in a nearby car. Lindbergh overheard the anonymous man say "Hey Doctor, over here, over here". Two and a half years later at the trial of the accused kidnapper, Bruno Hauptmann, Lindbergh claimed to be able to identify Hauptmann's voice as the same voice heard in the cemetery [1].

The Lindbergh case spurred Frances McGehee to initiate the first documented research on the reliability of earwitnesses [2, 3]. Since McGehee, research into SR has been a consistent topic in forensics and psychology research. The later development of the autonomous SR system has its roots in the work of McGehee.

3. THE FIRST SPEAKER RECOGNITION SYSTEM

Over the last 70 years SR has made major advances (see Figure 1). In 1962 an article was published in Nature by a Bell Laboratories Physicist Lawrence Kersta entitled, "Voiceprint Identification" [4]. Two years previous, Bell Laboratories had been approached by law enforcement agencies about the possibility of identifying callers who had made verbal bomb threats over the telephone [5]. After two years of research Kersta claimed he had developed a method to identify individuals with high success rates. His method utilized earlier work performed by other Bell Laboratories' scientists, Potter, Kopp and Green who were working on voice identification for military applications during World War II [6]. They had developed a visual representation of speech called a spectrogram. A spectrogram displays the frequency and intensity of a speech signal with respect to time. Kersta's method was an aural-visual method. A spectrogram was inspected visually for pattern matching and scored by an interpreter.

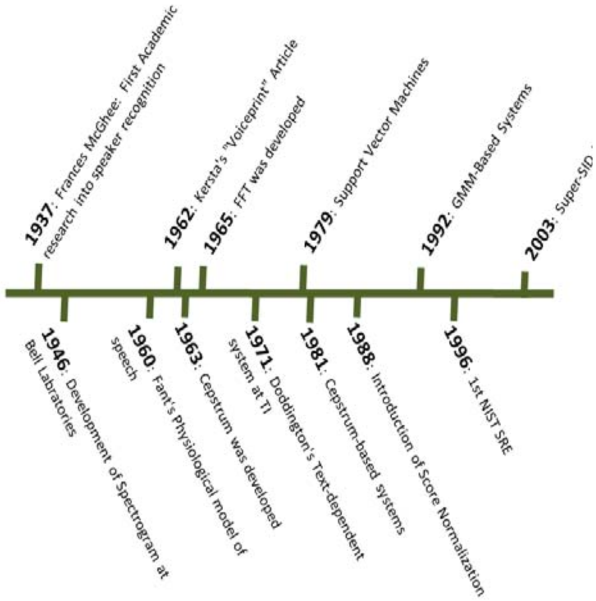

Figure 1: Timeline of major speaker recognition advances

Kersta's research, which produced extremely good results, sparked much research over the next few years. In fact, his article sparked an entire field of research. The first few years following Kersta's publication were intense. There were plenty of researchers with dissenting views. No researcher was able to replicate the incredible results of Kersta's work.

To help settle the matter, a research project was undertaken by Oscar Tosi, a professor at Michigan State who had doubts about Kersta's "voiceprint". His research was done in conjunction with the Michigan State Police and sponsored by the Federal Department of Justice. When his research was finished, Tosi's work yielded promising results for the emerging field [7].

Tosi's research was not without critics of its own. One year after Tosi's research was published his results were refuted by MIT scientist, Richard Bolt. Bolt's team illustrated holes in Tosi's methodology [8, 9]. The primary criticism was that Tosi's research lacked in practical applications. The FBI, being interested in the forensic application of speaker identification, requested another study be performed by the National Academy of Sciences. The results from this study showed that the technical uncertainties in forensic applications were substantial enough to claim the use of voiceprints were unreliable in any legal, forensic application. However, voiceprints are still found useful in certain circumstances. In fact the FBI has utilized a form of Kersta's spectrogram analysis as late as 2002 [5].

Kersta had not developed 'the solution' to speaker recognition. Today, the success rates with the spectrogram inspection method, given an expert interpreter and proper environmental circumstances, can be very high. But, "the good performance reported in Kersta's paper has not been observed in subsequent evaluations simulating real-life conditions" [10].

4. ENABLING SPEAKER RECOGNITION

In the 1960's, the same period of the 'Voiceprint' investigations, several unrelated developments arose which would eventually contribute to autonomous SR. These developments covered a broad range of disciplines. For instance, Gunnar Fant produced a physiological model of human speech production in 1960 [11]. The Fant model became the basis for understanding how to analyze speech for SR. Research into the physiological aspects of voice led future researchers to represent voice as a linear source-filter type model. Understanding voice using such a model allowed for many advances in discovering identifiable characteristics in an individual's voice.

Separate developments were occurring at this time in the field of computers. As computers became more accessible to more scientists, problems of implementation of continuous-domain mathematical solutions in a discrete world arose more often. In 1965 Cooley and Tukey published their method of digital implementation for the Fourier transform: now known as the Cooley-Tukey Fast Fourier Transform (FFT) [12]. The FFT gave scientists a method of frequency analysis in computer based systems. Two years earlier Bogert, Healy and Tukey had published a study on echo detection in seismic signals titled "The Quefrency Alanysis of the Time Series for Echos: Cepstrum, Pseudo-Auto-Covariance, Cross-Cepstrum, and Saphe Cracking" [13]. This oddly titled paper described a method of echo detection by taking the "spectrum" of a log-magnitude spectrum. Inspired by their echo-detecting Cepstrum, Michael Noll explored the use of the Cepstrum for pitch detection of a human voice [14]. Alan Oppenheim's research into homomorphic signal separation, led to the Complex Cepstrum, which is the complex-valued Fourier transform of the log spectrum [15]. Ronald Schafer soon joined Oppenheim research efforts. Oppenheim and Schafer, building on Noll's pitch detection, used cepstral analysis to model speech [16, 17]. The Cepstral speech model has become an important tool for SR systems.

5. EARLY RECOGNITION SYSTEMS

During the same decade (1960's) several investigations into automatic SR systems had began. Pruzansky (Bell Laboratories) investigated systems for SR utilizing spectral pattern matching [19, 20]. This system had limited success. However, the first successfully implemented autonomous SR system was developed by a team led by George Doddington at Texas Instruments in 1977 [10, 20, 21]. This system used digital filter banks to do spectral analysis. It was a text-dependent system that prompts the user for the correct verification phrase. A 'Euclidian distance' based algorithm was used to make a verification decision. Over many years this system had a reported false rejection rate and a false acceptance rate of less than 1% [10].

The early successful systems were all text-dependent. Later research has been able to improve on those early text-dependent successes. Investigations into text-independent methods at the time did not have such promising results. Text-independent research differs from the text-dependent research as scientists look for underlying identifying attributes, as opposed to pattern matching or phonetic event measurements. Text-independent research also trends toward speaker identification, as opposed to the simpler task of verification.

Text-independent research made a major advance in 1969 when James Luck proposed that the cepstrum be applied to SR [22-24]. Cepstral analysis would become the predominant method for obtaining measurable traits in a person's voice. However, it took some time before Luck's concept of cepstrum-based SR became widely used. The results of a study published by Atal in 1974 [25] demonstrated an improvement in identification accuracy of the cepstral approach over other approaches. But many researchers during the decade following Luck and Atal's papers overlooked cepstral-based systems. SR of this era focused on text-dependent systems using spectral features of voice. In 1981 Sadaoki Furui published results of another Bell Laboratory study [26]. Furui described the use of cepstral coefficients and their orthogonal polynomial coefficients in a frame-based system. The system was tested extensively and successfully. The success of the project sparked a renewed research effort in the use of the cepstrum. This approach uses the homomorphic deconvolution capabilities of the Cepstrum to separate the vocal tract envelope from the glottal excitation component of speech. It is the ability to analyze the de-convoluted voice signal that makes cepstral analysis a powerful tool which has dominated voice feature selection for the last three decades [23, 24, 27].

The modeling and decision making algorithms used in SR have also made significant improvement from the simple Euclidian distance method found in the TI system. The Hidden Markov Model, developed in the late 1960's, was employed widely in SR systems during the 1980's. Also a method of vector quantization (VQ), compressing a speaker feature vectors down to a small set, was also studied. However, the research of Matsui and Furui showed that the HMM and the VQ was about as effective as the less computationally demanding Gaussian Mixture Model (GMM) [8, 21, 28].

6. ADVANCES IN SPEAKER RECOGNITION

Each aspect of the recognition system, such as feature selection/extraction, feature modeling, feature classification and decision making, has made significant enhancements in the last fifteen years. The advances in each of the various aspects of speaker recognition have helped turn speaker recognition from solely a scholarly activity to a limited commercial reality. The remainder of this section reviews a few of the modern advances in speaker recognition.

6.1 High Level Features

The Cepstrum Coefficients or other variants of low level, short term (10-20ms) voice features has been the preferred feature for most SR tasks. However, the low-level approach ignores other identifiable information in a person's speech. Low-level features measure attributes of a person's voice (example: Pitch). High level features measure attributes of a person's speech (example: length of pauses between words). The idea that high-level features carried useful information in recognition systems was known for many years [25]. Early investigations tried to capitalize on this. Early attempts had limited success.

With the advent of the cepstrum the emphasis in research reverted back to low-level analysis. Serious investigations related to higher level features for autonomous SR began to reappear around the turn of the century [29]. One notable project, sponsored by NSF and the department of defense, (the Super-SID project) gathered prominent scientists in the field to test the idea of using high level features. The Super-SID project demonstrated a marked improvement when utilizing a fusion of both high and low level features [30].

6.2 The GMM-UBM

Throughout the years several types of feature modeling have been used. These include the Hidden Markov Model, Vector Quantization, and template matching models. In 1992, a recently graduated PhD student, Douglas Reynolds joined the Information Systems Technology group at Lincoln Laboratories. Reynolds Doctoral work had centered on modeling voice features for SR with Gaussian mixture models. His work led to a new paradigm in SR. [31-33].

The GMM performs similarly or better than other modeling techniques with a significant reduction in computational resources. By itself, the GMM marks a significant improvement in recognition systems. However, the simple multivariate Gaussian mixture models have been improved upon in several respects.

Perhaps the most notable improvement was the addition of the Universal Background Model (UBM) [33]. In addition to modeling a person's voice and testing the likelihood of that person being the authenticated user, it was proposed to use a set of people who were not the authenticated user. This allowed Bayesian theory to be employed and likelihood ratios used. The utterances given from a set of non-authenticated users are used to train a single GMM-UBM. The test utterance provided at time of authentication is tested against the user's trained GMM and against the GMM-UBM. The GMM-UBM is used to represent a speaker-independent distribution of features for that particular system. Therefore, the closer a user's test utterance matches the authenticated training data and the less it matches the UBM, the more likely that user is an authenticated user.

6.3 MAP-Adaptation and Supervectors

A group of scientists, led by Reynolds, employed a form of Bayesian learning called maximum a posteriori (MAP) estimation to perform model adaptation [34, 35]. The basic idea of adaptation is to derive the speaker model using the highly-defined UBM statistics in conjunction with the feature vectors from the speaker's training utterance. Instead of modeling the speaker's voice, adaptation models the speaker's variance from the GMM-UBM. The major advantage of the MAP-adapted GMM is that during authentication of a non-imposter when testing features do not align with the trained model, but do with align with "universal" features, then the negative affect of those features on the likelihood score will be mitigated.

Supervectors in the context of SR are the concatenation of the mean of each element in a multivariate MAP-GMM. The idea of the supervector had been applied to use in HMM's for speech recognition applications during the 1980's [36]. During the 1990's scientists made some attempts to apply similar supervector concepts to SR. It was after the development of the GMM-UBM and MAP adaptation that supervectors became useful for SR. The modern use of supervectors used in conjunction with the MAP-GMM helped commence much innovation with respect to classification techniques, which constitutes a sizable portion of the current research in SR.

6.4 Support Vector Machines

Support Vector Machines (SVM) are used to classify data. In the verification task, the SVM is used to classified data as an authenticated user or an imposter. The advantage of the SVM classifier is that it is able to minimize false reject and false accept error rates by using an optimized non-linear decision boundary (as opposed to a simple threshold).

SVM's were first developed in 1979 by Vladimir Vapnik [37]. In the 1990's SVMs were applied to machine recognized, hand-written digits [38]. The successful use in recognizing hand-writing helped inspire the idea of using SVM in SR. In 1996, Michael Schmidt and Herbert Gish reported on the first attempt at applying SVMs to SR [39].

The first attempt at implementing SVM in SR systems did not demonstrate a real improvement over other methods [31]. However, that first attempt combined with SVM advances in other applications, spurred on further research. Over the decade following Schmidt and Gish's, the SVM method became an important element of SR research [40].

6.5 Score Normalization

One substantial enhancement which has made practical systems a reality is score normalization. Like SVMs, score normalizations are designed to mitigate decision error. The SVM technique attempts to minimize error by altering the decision boundary. Score normalization attempts to minimize error by moving speaker model score vectors away from the decision boundary.

Score normalization research largely began with Li and Porter's proposal in 1988 to normalize the score distribution of the imposter model [8, 41]. This led to many variations of score normalization. Techniques include the Znorm and Tnorm methods. The Znorm normalizes scores during the enrollment period. The Tnorm is similar to the Znorm in purpose. The Tnorm however, is performed during the testing phase [8]. The Hnorm and the HTnorm presents a method to mitigate errors resulting from handset mismatched conditions [42]. Research has trailed off somewhat in relation to score normalization, however, limited score normalization research continues today [43, 44].

6.6 NIST SRE

The ability to quantify performance of any general system can be difficult. A set standard assists in making a comparison of systems. In 1996 the National Institute of Standards and Technology (NIST) began performing system evaluations for text-independent SR systems [45].

In the 1980's speech corpora were developed to standardize SR system testing. In the early 1990's the "Switchboard-1 Corpora" was collected by Texas Instruments. The Speaker Recognition Evaluation (SRE) performed by NIST in 1996 used this Corpus [46]. Additional corpora have been developed to assist in research of specific topics. In 1999 a switchboard corpus utilizing the growing GSM cellular technology was used in the NIST SRE. The following year a different corpus was used with CDMA cellular technology [46]. As the research and testing continues, the Corpuses utilized in evaluations have also changed.

7. CURRENT TRENDS AND FUTURE DIRECTIONS

Commercial text-independent SR systems exist today. Commercial systems perform with low enough error rates to make them practical in many applications. In the 2010 NIST SRE, equal error rates for the best systems were below 2% for core conditions [47].

In the last several years the broader field of pattern recognition techniques has contributed a lot to SR research. Currently joint factor analysis plays a major role in many high performance recognition systems [48-50]. Principal component analysis, linear discriminant analysis, latent factor analysis and many other techniques for dealing with classification in stochastic data have also been applied to SR systems [40, 48, 51-53]. These techniques are offspring of the application of supervectors to SR [40, 54]. Application of pattern classification advances to SR will continue to be a strong field of research.

The fusion of scores from high-level speech features with low-level features was one method that has helped lower error rates. The disadvantage of fused systems is computational cost. The mathematical techniques of pattern recognition applied to SR has reduced error rates a significant amount and reduced the computational cost of the overall systems enough that fused systems using high-level features appears to currently be impractical for real-world systems [48].

One major application requiring improved error rates is identification in forensic applications. Currently caution is required for forensic uses of speaker identification [55]. However, the push toward forensics has led to some interesting research. For instance, performing research to better understand what voice features are common among speakers has recently been undertaken [56, 57]. This research has lead to further research into which vocal features change depending on age, ethnicity, language, emotion, intent, dialect region or other factors. Another topic of research which has been promulgated for forensic purposes reaches back to the beginning of autonomous SR. In 2010 the NIST SRE included a Human Assisted Speaker Recognition (HASR) test [58]. Similar to the idea of the Kersta's voiceprint, HASR attempts to lower error rates by allowing humans (research has been done on both trained and untrained individuals) to supplement the autonomous systems. Early research demonstrates a possibility for further advancement in this field [59].

Of course a major area of research continues to be environmental variability, such as background noise or handset variability [60-62]. Environmental concerns become a major factor in applications where unknown conditions exist. With the advent of the internet and security applications over the internet, security needs in unknown conditions have become more and more prevalent. Therefore, research into environmental concerns will also continue to be a focus of research.

8. SUMMARY

The use of voice as a biometric for identification spans centuries of time. In the early part of the twentieth century, vocal recognition began to be studied as a serious academic venture. Combining the idea of SR with the rise of the computer has led to the autonomous speaker recognition system.

In the very early years of computing, the idea of computer-based voice recognition was proposed. The spectrogram was the first major step toward computer-based SR. It was Kersta's initial usage of the spectrogram in his voiceprint article that really sparked the field of SR research.

The first fully autonomous successful SR system was developed in the early 1970s. The system was a text-dependent authentication system developed and used for access control at Texas Instruments. Advances over the next twenty years have led researchers closer to a successful text-independent system. Over the past 25 years the thrust of research has been toward text-independent systems.

Today, commercial ventures into speaker biometrics have become more common across the globe. With Kersta's initial claims, SR has been long anticipated. After all these years of research, speaker recognition continues to be just on the cusp of full-fledge commercial veracity (2% ERR!).

REFERENCES

1. Yarmey, A, Yarmey, M, Todd, L., "Frances McGehee; The First Earwitness Researcher", Perceptual and Motor Skills, 106: 387-394, 2008.
2. McGehee, F., "The reliability of the identification of the human voice", Journal of General Psychology. 1937, 17, 249-271.
3. McGehee, F., "An Experimental Study Voice Recognition", Journal of General Psychology, 1944, 31, 53-65.
4. Kersta, L., "Voiceprint Identification", Nature Magazine, December 1962, 196, 1253.
5. Lindh, J., "Handling the Voiceprint Issue", FONETICK Proceedings, 2004.
6. Potter, R., Kopp, G., Green, H., "Technical Aspects of Visual Speech", Bell Labs, New York, 1947.
7. Tosi, O. Oyer, H., Lashbrook, W., Pedrey, C., Nicol, J., Nash, E., "Experiment On Voice Identification", Journal of the Acoustical Society of America, 1972, 51:2030-2043.
8. Bimbot, F., Bonastre, J., Fredouille, C., Gravier, G., Chagnoleau, I., Meignier, S., Merlin, T., Ortega-Garcia, J., Petrovska-Delacretaz, D., Reynolds, D., "A Tutorial on Text-Independent Speaker Verification", EURASIP Journal on Applied Signal Processing vol. 4 431-450, 2004.
9. Bolt, R. Cooper, F., David Jr. E., Denes, P., Pickett, J., and Stevens, K., "Speaker Identification by Speech Spectograms: Some Further Observations", Journal of the Acoustical Society of America, 1973, 54:53-534.
10. Doddington, G., "Speaker Recognition – Identifying People By Their Voice", Proceedings of IEEE, November 1985, 73:1651-1664.
11. Fant, G., <u>Acoustic Theory of Speech Production</u>, Mouton and Co., The Hague, Netherlands, 1970.
12. Cooley, J.W.,. Tukey, J.W., "An algorithm for the machine computation of complex Fourier series" Math Computation, vol. 19, pp.297-301, Apr. 1965.
13. Bogert, Healy, Tukey, "The Quefrency Alanysis of the Time Series for Echos: Cepstrum, Pseudo-Auto-Covariance, Cross-Cepstrum, and Saphe Cracking" in *Time Series Analysis*, ch.15, pp. 209-243, 1963.
14. Noll, A.M., "Cepstrum Pitch Determination", Journal of Acoustical Society of America, vol. 41, pp. 293-309, February 1969.
15. Oppenheim, A., Shafer, R., "From Frequency to Quefrency: A History of the Cepstrum", IEEE Signal Processing, September 2004.
16. Oppenheim, A. V., Schafer, R. W., "Homomorphic Analysis of Speech", IEEE, Trans. on Audio and Electroacousitcs, Vol. 16:2, pp. 221-226, June 1968.
17. Schafer, R. W., Rabiner, L.R., "Digital Representation of Speech", Invited Paper in Proceedings of the IEEE, Vol. 63:4, pp. 662-667, April 1975.
18. Baum, L., Petrie, T., Soules, G., Weiss, N., "A Maximization Technique in the Statistical Analysis of Probabilistic Functions of Markov Chains", Annals of Mathematical Statistics, Vol. 41, No.1, 1970.
19. Pruzansky, S., "Pattern-matching procedure for automatic talker recognition," *J.* Acoust. *SOC.* Amer., vol. 35, pp. 354-358, 1963.
20. Woodard, J., Orlans, N., Higgins, P., <u>Biometrics</u>, McGraw-Hill, 2003.
21. Furui, S., "50 Years of Progress in Speech and Speaker Recognition", Proceedings of SPECOM, Patras, Greece pp.1-9., 2005
22. Luck, J.E., "Automatic Speaker Verification Using Cepstral Measurements", Jrnl of the Acoustical Scty of America, Vol. 46:4B, pp. 1026-1032, 1969.

23. Leeuw, K. and Bergstra, J., The History of Information Security – *A Comprehensive Handbook*, Elvsevier, 2007.
24. Wayman, J., Orlans, N., Hu, Q., Goodman, F., Ulrich, A., Valencia, V., “Technology Assessment for State of the Art Biometrics Excellence Roadmap”, MITRE Technical Report, March 2009, vol 2of3 v1.3.
25. Atal, B.S., “Effectiveness of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification and Verification”, Journal of the Acoustical Society of America, Vol. 55:6, pp. 1304-1312, June 1974.
26. Furui, S., “Cepstral Analysis Technique for Automatic Speaker Verification,” IEEE Trans. Acoust. Speech, Signal Proc., 29, pp. 254-272, 1981.
27. Furui, S., “Selected Topics on 40 Years of Research on Speech and Speaker Recognition”, Keynote Speech in InterSpeech 2009, Brighton, 2009.
28. Matsui, T. and Furui, S., “Comparison of Text-Independent Speaker Recognition using VQ-Distortion and Discrete/Continuous HMMs”, Proceedings of ICSLP, pp. 157-160, 1997.
29. Doddington, G., “Speaker Recognition based on Idiolectal Differences between Speakers,” Eurospeech, Vol. 4, pp. 2517-2520, 2001.
30. D. Reynolds, W. Andrews, J. Campbell, J. Navratil, B. Peskin, A. Adami, Q. Jin, D. Klusacek, J. Abramson, R. Mihaescu, J. Godfrey, D. Jones, and B. Xiang, “The SuperSID Project: Exploiting High-level Information for High-accuracy Speaker Recognition,” Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on, vol.4, no., pp. IV-784-7 vol.4, 6-10 April 2003.
31. Reynolds, D. A., “A Gaussian Mixture Modeling Approach to Text-Independent Speaker Identification”, Ph.D. thesis, Georgia Institute of Technology, September 1992.
32. D.A. Reynolds, R.C. Rose, “Robust text-independent speaker identification using Gaussian mixture speaker models,” Speech and Audio Processing, IEEE Transactions on , vol.3, no.1, pp.72-83, Jan 1995.
33. Reynolds, D. A., “Automatic speaker recognition using Gaussian mixture speaker models”, Lincoln Labratory Journal, 1996, 8:173–192.
34. Reynolds, D, Quatieri, T, and Dunn, R, “Speaker Verification Using Adapted Gaussian Mixture Models Digital” Signal Processing,, 2000, pp. 19-41
35. Reynolds, D. A., “Comparison of Background Normalization Methods for Text-Independent Speaker Verification”, Proceedings of the European Conference on Speech Communication and Technology, Vol. 2, pp. 963-966, September 1997.
36. Gersho, A., Shoham, Y., “Hierarchical Vector Quantization of Speech with Dynamic Codebook Allocation”, ICASSP 1984, pp. 416-419, Mar 1981.
37. V. Vapnik, Estimation of Dependencies Based on Empirical Data, Nauka, Moscow, 1979 .
38. C. Cortes, V. Vapnik, “Support Vector Networks”, Machine Learning, May 1995, 20:273-297.
39. M. Schmidt, “Identifying speakers with support vector networks”, In Proc. of 28th Symposium on the Interface, Sydney, Australia, 1996.
40. Campbell, W., Sturim, D., and Reynolds, D., “Support vector machines using GMM supervectors”, IEEE Signal Processing Letters 13, 5 (May 2006), 308–311.
41. Li, K., Porter, J., “Normalizations and selection of speech segments for speaker recognition scoring”, IEEE Conference on Acoustics, Speech, Signal Processing vol. 1, pp. 595-598
42. Reynolds, D.A., “HTIMIT and LLHDB: Speech Corpora for the Study of Handset Transducer Effect”, ICASSP-97, IEEE, pp. 1535-1538, Apr 1997.
43. Yin, C., Rose, R., Kenny, P., "Adaptive score normalization for progressive model adaptation in text independent speaker verification", IEEE, Proc. of ICASSP08, 4857-4860, Las Vegas, NV, 2008.
44. Apsingekar, V. R., De Leon, P. L. Speaker Verification Score Normalization Using Speaker Model Clusters”, Speech Communications, Vol. 1, pp. 110-118, Jan. 2011.
45. Martin, A.F., Przybocki, M.A.: The NIST Speaker Recognition Evaluations: 1996- 2001. In: Proceedings of the the Odyssey Speaker Recognition Workshop, Chania, Crete, Greece, pp. 39-43.
46. Martin, A., “Speaker Databases and Evaluation”, Encyclopedia of Biometrics, NIST, 2009.
47. T. Kohler, “The 2010 NIST Speaker Recognition Evaluation”, SLTC Newsletter, July 2010
48. J. Gonzalez, I. Lopez-Moreno, J. Franco-Pedroso, D. Ramos, D. Toledano and J. Gonzalez-Rodriguez, “ATVS-UAM NIST SRE 2010 System”, in Proc. of FALA 2010, 2010 pp. 415-418.
49. N. Scheffer and R. Vogt, “On The Use of Speaker Superfactors For Speaker Recognition”, In Proc. of ICASSP2010, 2010, pp.4410-4413.
50. S. Kajarekar, N. Scheffer, M. Graciarena, E. Shriberg, A. Stolcke, L. Ferrer, and T. Bocklet, “The SRI NIST 2008 speaker recognition evaluation system”, In Proc. of the 2009 IEEE ICASSP2010, 2010, Washington, DC, USA, 4205-4208.
51. D. Sturim, W. Campbell, Z. Karam, D. Reynolds, F. Richardson, “The MIT Lincoln Laboratory 2008 Speaker Recognition System”, Interspeech 2009, Brighton, UK, Sept. 6, 2009.
52. W. Zhang, Y. Yang and Z. Wu, “Exploiting PCA Classifiers to Speaker Recognition”, Proc. of the International Joint Conference on Neural Networks, Vol. 1, pp. 820 - 823, 20-24 July 2003.
53. Q. Wu, L. Zhang, “Nonnegative Tensor PCA and Application to Speaker Recognition in Noise Enviroments”, IEEE, Fourth International Conference on Natural Computation, 2008, pp. 187-191.
54. R. Kuhn, P. Nguyen, J.-C. Junqua, et al. “Eigenvoices for Speaker Adaptation”. ICSLP-98, V. 5, pp. 1771-1774, Sydney, Australia, Nov. 30 - Dec. 4, 1998.
55. J.F. Bonastre, F. Bimbot, L.J. Boe, J.P. Campbell, D.A. Reynolds, and I. Magrin-Chagnolleau, “Person Authentication by Voice: A Need for Caution,” *Proc. of Eurospeech*, ISCA, Geneva, Switzerland, pp. 33-36, 1-4 September 2003.
56. Schwartz, R., Shen, W., Campbell, J., Granville, R., Measuring Typicality of Speech Features in American English Dialects: Towards Likelihood Ratios in Speaker Recognition Casework, 5th European Academy of Forensics Science, Glasgow, Scotland, Sept. 8, 2009.
57. N. Chen, W. Shen, J. Campbell, P. Torres-Carrasquillo, "Informative Dialect Recognition Using Context-Dependent Pronunciation Modeling", ICASSP 2011, Prague Czech Republic, May 2011.
58. “The NIST Year 2010 Speaker Recognition Evaluation Plan” NIST at http://www.itl.nist.gov/iad/mig//tests/sre/2010/NIST_SRE10_evalplan.r6.pdf
59. R. Schwartz, J Campbell, W. Shen, D. Sturim, W . Campbell, F. Richardson, R. Dunn and R. Granville, “USSS-MITLL Human Assisted Speaker Recognition”, IEEE, ICASSP2011, Prague Czech Rpublic, May 2011.
60. C. Shaver and J. Acken, “Effects of Equipment Variation on Speaker Recognition Error Rates”, IEEE, ICASSP2010, Dallas Texas, March 2010.
61. J. Ming, T. Hazen, J. Glass, D. Reynolds, “Robust Speaker Recognition in Noisy Conditions”, IEEE Transactions on Audio, Speech, And Language Processing, VOL. 15, NO. 5, JULY 2007.
62. K. Kumar, Q. Wu, Y. Wang, M. Savvides, “Noise Robust Speaker Identification Using Bhattacharyya Distance in Adapted Guassian Models Space”, EUSIPCO-2008.