

Attempts at Quantifying the Effects of Modelling Assumptions for GMM-based Speaker Recognition

Howard Lei

CSU – East Bay, Hayward, CA

Farnaz Ganjeizadeh

CSU – East Bay, Hayward, CA

Erik Olivar

CSU – East Bay, Hayward, CA

Abstract

Speaker recognition approaches rely heavily on the use of Gaussian Mixture Models (GMMs) for speaker modelling. The models can represent arbitrary distributions of feature vectors extracted from speech waveforms, and are easy to train. However, they make several simplifying assumptions on the distribution of the feature vectors, including Gaussianity and time-independence, which are not accurate given the nature of speech. This work seeks to quantify the effects of these assumptions as they pertain to speaker recognition system performance. Experiments are performed using a traditional GMM-UBM system. Initial results suggest that the Gaussian distribution assumption can negatively impact performance, while further investigation is needed to make definitive conclusions.

1. Introduction

Speaker recognition has been an established area of research for the past 15 years, and involves the application of signal processing, statistical, and machine learning algorithms to the recognition of speaker identities in audio recordings. The technology is applicable to high-tech applications, such as voice-based biometrics [1], and forensics [2][3]. The traditional speaker recognition approach that has been widely popular until around 2007 is one that uses Gaussian mixture models (GMMs) to model the feature vectors extracted from the speech waveforms of speakers [4]. It is referred to as the GMM-UBM approach, which involves the use of a Universal

Background Model (UBM) to represent feature vectors from a large set of speakers. The feature vectors are extracted using acoustic signal processing techniques. While more recent approaches for speaker recognition have relied on advanced techniques such as Joint Factor Analysis (JFA) [5] and i-vectors [6][7], the classical approach involving GMM models is still viable in environments where speaker data is limited. This is because the JFA and i-vector techniques rely on large amounts of development data for modelling purposes. Such data, especially those matching the noise and recording environments of the target data, are not always available,

The GMM models consist of a mixture of multivariate Gaussians probability distributions which are easy to obtain (or train) using the feature vectors. Given limited knowledge of the data, GMMs can model feature vector distributions that are difficult to precisely characterize, such as feature vectors resulting from speech waveforms. GMM models are trained using the Expectation-Maximization (EM) algorithm, an iterative algorithm that finds a maximum-likelihood estimate of the model parameters given the feature vectors [8]. The EM algorithm is similar to the K-means clustering algorithm [9], except that it uses soft clustering assignments. In soft clustering, each feature vector is assigned a likelihood of belonging to each GMM mixture. The mixture means, covariances, and weights are updated based on the likelihoods of its frames.

In every speaker recognition system, UBM is needed to represent the distribution of a general population of speakers [4]. The UBM is a speaker-independent GMM model that is used for score normalization and speaker-dependent GMM training, and is itself trained using the

EM algorithm given feature vectors from a large number of speakers. GMM models are widely used not only for the classical GMM-UBM approach but also for the more advanced JFA and i-vector approaches. The i-vector approach seeks to obtain low-dimensional vectors from speech waveforms representing speaker "voiceprints" [6][7]. A UBM is needed for the statistical algorithms used to extract the vectors.

Limited work has been done to investigate the validity of the assumptions that the model makes for speaker recognition. The model assumptions include Gaussian feature vectors distributions generated from different Gaussian mixtures, and time-independence of feature vector sequences. It's evident, however, that speech signal waveform samples can not be considered to be time-independent, because words and sentences contained in speech would not be acoustically coherent if its signal sample values are scrambled and played back in time.

Lastly, the fact that the speaker-dependent GMM models are trained from the UBM (the UBM is used as the initialization parameters in the EM algorithms) can also lead to the fact that the models may overly depend on the UBM and not fit the data well enough. The overall purpose of this work is to aid in the development of speaker recognition approaches by performing experiments that test and analyze the modelling assumptions of GMMs.

This work consists of the following steps:

1. Implement the classical GMM-UBM system, and obtain a baseline performance.
2. Perform experiments investigating the weaknesses in the GMM modelling assumptions.
3. Comparative and analyze the methods.
4. Obtain conclusions on the GMM's modeling assumptions.

The article is structured as followed: Section 2 discusses related work. Section 3 lists the data collected, and Section 4 describes the baseline GMM-UBM approach. Sections 5, 6, and 7 describe the new methodologies. Section 8 describes the speaker recognition performance measures, and Section 9 describes the results and provides a discussion. Section 10 provides a summary and discussion of future work.

2. Related Work

This work is primarily inspired by the work of Gillick and Wegmann, 2011 [10], who investigated the acoustic model assumptions of Hidden Markov Models (HMMs) for automatic speech recognition. HMMs are similar to GMMs, but also accounts for time-

dependence in the distribution of the data that it models. Their primary conclusion was that there are significant data/model mismatches when using HMMs, especially given the still-existent time-independence assumptions of the model. Gillick and Wegmann found that the Gaussianity assumption of the model is less of an issue. The focus of this work is more on the Gaussianity assumption, because the time-independence assumption of GMMs used in speaker recognition has been a widely known problem with few solutions. In addition to the aforementioned work, the work of Reynolds, 1992 [4] discusses the classical GMM-UBM approach. The work of Taufiq and Hansen, 2011, investigates ways to train better UBMs (including the use of less data) in the GMM-UBM approach. The work of Bar-Yosef and Bistriz, 2009 [11] investigates the use of different UBMs for score-normalization purposes in speaker recognition systems. The conclusion of this last work is that the use of multiple UBMs can be advantageous for score-normalization purposes.

Some speaker recognition work that relates to popular culture includes Sargin's research is based on using speaker recognition to identify celebrities in YouTube videos [12]. Speaker recognition used for recognizing celebrities in TV broadcasts were discussed by Everingham et al. [13]. There are many other works of speaker recognition found in literature, too numerous to list here.

3. Data

The data consists of recorded speech from California State University, East Bay students and faculty. The recorded subjects include four females and 17 males, spanning 9 different accents. Each subject was asked to read two paragraphs carefully selected from a textbook containing numbers and simple wording. The duration of the first paragraph is roughly two minutes, while the second paragraph is roughly one minute. The recordings were taken using a Blue Snowball USB microphone with the omni-directional microphone setting, with a sampling frequency of 44,100 samples per second. The total amount of speech used in all experiments is roughly one hour. We note that the dataset we used using is significantly smaller compared to standard datasets, such as the NIST Speaker Recognition Evaluation Datasets [14]. However, our aim is to understand the GMM modelling assumptions, for which a smaller dataset with fewer variables can be more suitable.

4. Baseline GMM-UBM approach

The baseline for our experiments is the classical GMM-UBM approach, which is based on training GMMs to model the distribution of feature vectors of extracted from speech waveforms. The feature vectors are Mel-Frequency Cepstral Coefficients (MFCCs) C0-C19, a total of 20 dimensions. In addition, the first and second time derivatives of the coefficients of each feature vector dimension are appended to generate vectors of 60 dimensions. The typical feature extraction approach extracts one MFCC feature vector for every 10ms of speech using 25ms windows of speech, such that an entire speech waveform is represented by a sequence of feature vectors. Every minute of speech should hence contain 100 vectors. Each MFCC feature vector dimension is mean and variance normalized across the duration of each waveform. Because our work is focused on the modelling approaches and not on the MFCC feature vectors, we will omit a full description of the feature vector extraction process from the acoustic and signal processing standpoint. For those interested, the work of [15] describes the MFCC features in detail.

The GMM-UBM approach involves first training a UBM via the EM algorithm on a set of speech data across multiple speakers. The UBM represents the speaker-independent model. In our particular implementation of the system, speaker-dependent GMM models are trained using the EM algorithm from each speaker's data, and the UBM is used to initialize the algorithm. The following equation describes the probability density function (pdf) of a GMM model:

$$p(x; \mu, \Sigma, \omega) = \sum_{m=1}^M \omega_m N(x; \mu_m, \Sigma_m) \quad (1)$$

where \mathbf{x} is a vector, $N(x; \mu_m, \Sigma_m)$ is a pdf of a Gaussian distribution with mean μ_m and covariance matrix Σ_m , and ω_m are the mixture weights. M is the number of Gaussian mixtures. The UBM is trained using the first paragraph of speakers 1-10, while the speaker-dependent models are trained using the first paragraphs of each speaker. Hence, the first paragraphs of each speaker comprise the training data.

In our experiments, we use eight mixtures for each GMM ($M=8$), with full covariance matrices. The number of mixtures is small compared to those used in a typical GMM-UBM system, with 512 to 2,048 mixtures. However, the dataset we are using (1 hour of total speech) is also significantly smaller compared to the typical datasets, and hence fewer mixtures are needed. Figure 1 illustrates the GMM training process.

Once the speaker-dependent GMM models are trained, data from the second paragraphs of each speaker (i.e. the test data) is used to generate test MFCC feature vectors for each speaker. The test feature vectors are scored against each speaker-dependent model. Specifically, given a speaker A for which test MFCC feature vectors are generated, and a speaker B for which a speaker-dependent GMM is generated, a log-likelihood ratio (LLR) is computed to generate a speaker-similarity score, as shown in the equation below:

$$score(A, B) = \log \left(\frac{\sum_{i=0}^N p(x_{Ai}; \mu_B, \Sigma_B, \omega_B)}{\sum_{i=0}^N p(x_{Ai}; \mu_{UBM}, \Sigma_{UBM}, \omega_{UBM})} \right) \quad (2)$$

where $p(\bullet)$ is a pdf of a GMM, $score(A, B)$ is the similarity score between speakers A and B, μ_B , Σ_B , and ω_B are the parameters of the GMM trained for speaker B, and μ_{UBM} , Σ_{UBM} , and ω_{UBM} are the parameters of the UBM. x_{Ai} is MFCC feature vector i from the test data (second paragraph) of speaker A, which has a total of N feature vectors. Figure 2 illustrates score computation.

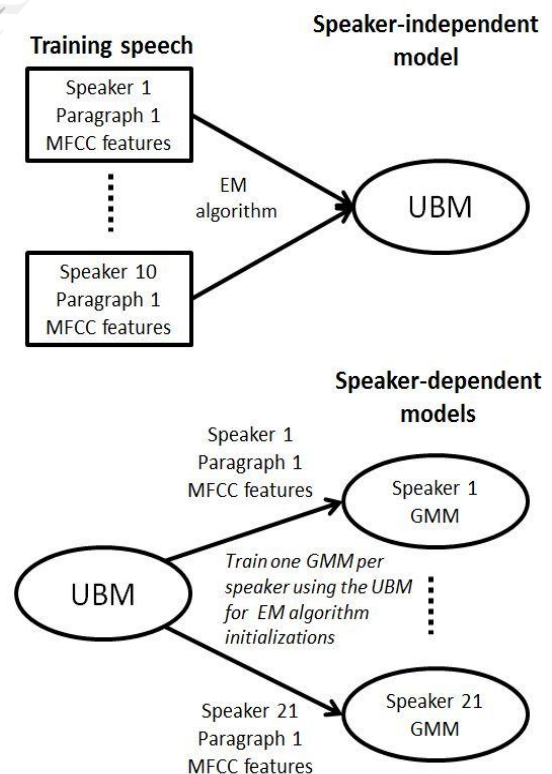


Figure 1: Training of speaker-independent UBM, and speaker-dependent GMM models

Note that our implementation of the GMM-UBM speaker recognition system is completed using publically available MATLAB scripts under the BSD license.

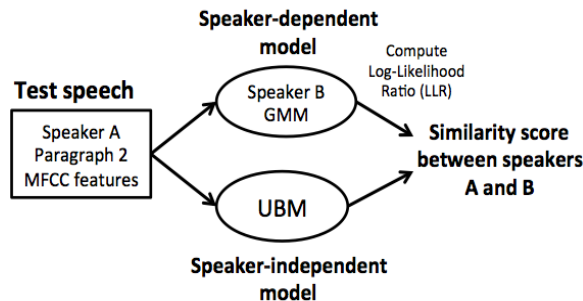


Figure 2: Log-Likelihood Ratio (LLR) speaker similarity score computation

5. Investigation 1: Examining the Gaussianity Assumption of GMMs

The primary focus of this work is to analyse the assumptions of the GMM models, to eventually quantify the effectiveness of the GMM models in modelling MFCC feature vector distributions of speech. The experiments utilize the basic GMM-UBM framework, while varying the method by which the MFCC features of the test data are generated. In the baseline system, the MFCC features are taken for LLR score computation as they are, in the exact sequence as they occur in the test data.

The following describes the first procedure for testing the Gaussianity assumption. It involves generating artificial MFCC feature vectors for the test data that conform to a GMM distribution. This approach is inspired by [10].

1. Use the EM algorithm to train speaker-dependent models from the test data, using the second paragraphs of each speaker. The UBM is used to initialize the EM algorithms. Hence, each speaker is associated with two speaker-dependent GMMs, one trained using the first paragraph, and one trained using the second paragraph. We refer to the GMM trained using the second paragraphs as GMM-2.
2. Using GMM-2 for each speaker, assign each MFCC feature vector in the test data to the mixture in GMM-2 most likely to have generated the feature vector. This is done by computing the probability of the vector being generated by each mixture in GMM-2, and

choosing the mixture with the highest probability.

3. Replace the original feature vector with a vector sample from the resulting GMM-2 mixture.

Replacing the test data MFCC features with samples from GMM-2 helps determine the effectiveness of the Gaussian distribution assumption, since the new test feature vectors will have been directly sampled from a GMM. Note that because each test GMM is also trained using the EM algorithm, and not simply MAP-adapted, as was done in the typical GMM-UBM approach [4]) our GMM-2 models should more closely match the distributions of the original test feature vector. Figure 3 illustrates this approach:

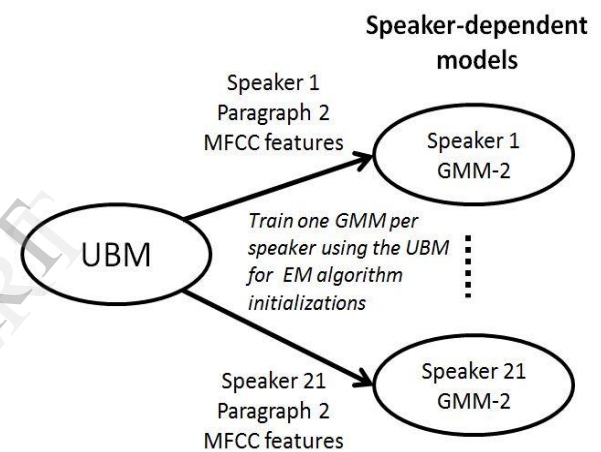


Figure 3: Training speaker-dependent GMMs using the test data (second paragraphs) of each speaker.

6. Investigation 2: Examining the Gaussianity and Time-Independence Assumptions of GMMs

The following describes the procedure for investigating both the Gaussianity and time-independence assumptions of feature vectors. The approach involves re-sampling the test MFCC feature vectors as they are, while imposing the mixture-based assumption of GMMs. This approach is also inspired by [10].

1. Use the EM algorithm to train speaker-dependent models for the test data, using the second paragraphs of each speaker. The GMM-2 models are obtained. The UBM is used to initialize the EM algorithms. This step is the same as the first step of Investigation 1.

2. Using GMM-2 for each speaker, assign each MFCC feature vector in the test data to the mixture in GMM-2 most likely to have generated the feature vector. This step is the same as the second step of Investigation 1.
3. For a given test feature vector, replace it with a random sample from the set of all feature vectors assigned to the same mixture as the given feature vector.

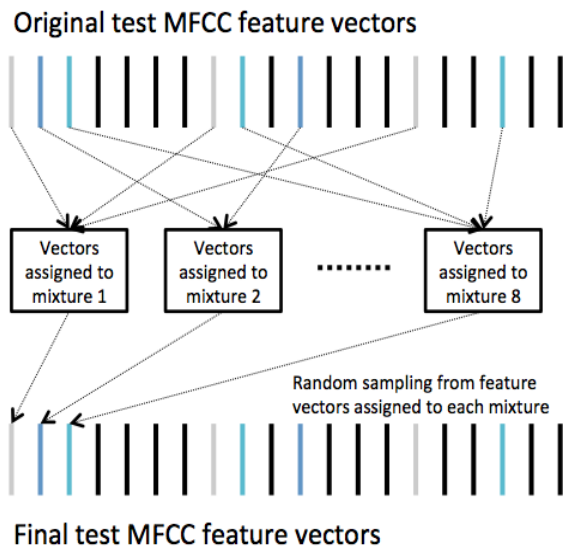


Figure 4: Creating new MFCC feature vectors by sampling of original feature vectors assigned to different GMM mixtures.

Hence, the final set of test data feature vectors is a randomized selection the original sequence of test feature vectors. Some of the original feature vectors may not be selected to be a part of the final set. This approach preserves the fact that a GMM is comprised of a set of mixtures, but discards the Gaussianity assumption of the GMMs. Figure 4 illustrates this approach.

7. Investigation 3: Examining the effect of alternate UBMs

In this investigation, a second UBM is trained the using the second paragraphs of 20 speakers, such that the UBM is much more closely matched to the test data. We refer to this UBM as UBM-test. This is in contrast to the original UBM, which is matched more closely to the training data (first paragraphs). First, the approach from investigation 1 (Section 5) is repeated, but with all speaker-dependent GMMs from the training and test data re-trained with UBM-test for EM algorithm initialization. This potentially leads to better

performance, as the speaker-dependent GMMs would be more closely aligned to the test data. We note that in typical speaker recognition experiments, use of test data to train the UBM is not allowed. However, use of test data can help us better understand the GMM modelling assumptions for purposes of this work.

Second, the approach from Investigation 1 is repeated, but only with the GMM-2 models trained with UBM-test for EM initialization. This attempts to quantify the effect of mismatches in the UBM initializations for different GMM speaker models. In the classical GMM-UBM approach, a single UBM is used to train all speaker-dependent models. This simplifies the training but also helps to maintain correspondence of mixtures between the speaker-dependent GMMs and the UBM for LLR score computation. This helps normalize the LLR scores. Given the fact that we are using only eight GMM mixtures, however, maintaining correspondence between the mixtures may be less of an issue.

8. Performance measures

The effectiveness of each speaker recognition approach can be quantified using the following three measures, which are widely used in Speaker Recognition and Speaker Identification research:

- Closed-Set Speaker Identification Accuracy
- Log-Likelihood Cost (CLLR) [16]
- Equal Error Rate (EER)

The closed-set speaker identification accuracy is the percentage that the test data's speaker (test speaker) is correctly identified given the set of all speakers in the dataset, and the knowledge that the test speaker is included among the set of all speakers. A test speaker is correct identified if the LLR score is highest for the speaker-dependent GMM of the same speaker. The higher the accuracy percentage, the better the speaker recognition approach.

The CLLR is computed according to the following equation [16]:

$$C_{llr} = 1 / (2 \log 2) * (\sum \log(1 + 1/s) / N_{TT} + \sum \log(1 + s) / N_{NT}) \quad (3)$$

where the first summation is across all speaker similarity scores with matching training and test speakers (target speaker scores), and the second is across scores with non-matching speakers (non-target speaker scores). N_{TT} and N_{NT} are the total numbers of target and non-target speaker scores. s is a score, in the form of a likelihood ratio [16].

It should be noted that for an ideal speaker recognition system, the first summation in the above

equation should have a higher scores s , while the second summation should have a lower scores s . This implies that the lower the C_{lr} , the better the speaker recognition approach at separating the target and non-target speaker scores. Note also that the C_{lr} examines the set of all scores and gives a measure of how well the system separates the set of *all* target and non-target scores, not just the ones that affect the accuracy.

The third performance measure is the Equal Error Rate (EER). The EER occurs at a scoring threshold where the rate at which non-target speaker scores are misclassified as target speaker scores (false alarms), equals the rate at which target speaker scores are misclassified as non-target speaker scores (misses). Similar to the C_{lr} , The lower the EER, the better the speaker recognition approach at separating the target and non-target speaker scores.

Note that 21 speakers are used for all experiments in this work, and each speaker provides both training and test data. Hence, the speaker recognition approaches all generate 21 target speaker scores, and 420 non-target speaker scores ($21 \times 21 = 441$; $441 - 21 = 420$), where each test data is scored against every speaker-dependent model obtained from the training data.

9. Results and Discussion

Using the dataset described in Section 3, and the performance measures described in Section 8, results are generated for the baseline approach, and the new investigations. The approach described in Section 5 where test MFCC feature feature vectors are obtained by sampling from the GMM-2 models is referred to as *Gaussian Sampling*. The approach described in Section 6 with the random sampling of feature vectors is referred to as *Empirical Sampling*.

The first approach described in Section 7, where UBM-test is used for training all speaker-dependent GMMs, and where sampling from GMM-2 models are used, is referred to as *Gaussian Sampling-UBMtest*. The second approach from Section 7, where different UBMs are used to train the speaker-dependent GMMs, is referred to as *Gaussian Sampling-UBMDiff*.

Lastly, as a sanity-check, the *Gaussian Sampling* approach is repeated, but using the training speaker-dependent GMMs for test MFCC mixture assignment and Gaussian sampling. The resulting sampled feature vectors should be closely matched to the training models, and give a significantly better speaker recognition performance compared to the other approaches. This approach is referred to as *Gaussian Sampling-GMMTrain*. Table 1 summarizes the results.

We caution the reader that, because a very small dataset is used, there are issues of statistical significance in the results. However, this work only

seeks to suggest likely trends resulting from the different investigations, and is not meant to make conclusive statements.

According the Table 1, the results suggest that the *Baseline* and *Empirical Sampling* approaches, with accuracies of 95.2%, outperformed the other approaches. The CLLR and EER for these approaches are not significantly different (3.1 vs. 3.0 for CLLR, and 4.9% vs. 5.7% for EER). The *Gaussian Sampling-UBMtest* approach has the next best accuracy at 85.7%, while the *Gaussian Sampling* approach has the worst accuracy at 76.2%. The *Gaussian Sampling-GMMTrain* approach gives the best accuracy (100%), CLLR (1.0), and EER (0.0%), as expected.

Table 1: Speaker recognition results for all performance measures

Approaches	Accuracy	CLLR	EER
Baseline	95.2%	3.1	4.9%
Gaussian Sampling	76.2%	3.5	9.3%
Empirical Sampling	95.2%	3.0	5.7%
Gaussian Sampling-UBMtest	85.7%	3.7	5.7%
Gaussian Sampling-UBMDiff	66.7%	3.7	9.5%
Gaussian Sampling-GMMTrain	100.0%	1.0	0.0%

The fact that the *Gaussian Sampling* approach has the worst accuracy suggests that using the original test MFCC feature vectors may be preferable to substituting them with Gaussian samples from a GMM model. This seems counter-intuitive, since one might expect that having test MFCC feature vectors that conforms to the speaker-dependent training GMMs would give a closer match between the training model and test features. However, because each set of test feature vectors are scored against the set of all training models, it may be that the test MFCC feature vectors match closely not only to the matching-speaker training model, but to all the non-matching-speaker models as well.

The fact that the *Gaussian Sampling-UBMtest* approach outperforms the *Gaussian Sampling* approach suggests that it is more helpful to use the test data to create UBMs. The *Gaussian Sampling-UBMtest* approach furthers narrows the gap between the speaker-dependent training models and the test MFCC feature

vectors. However, the *Gaussian Sampling-UBMDiff* approached performed slightly worse than *Gaussian Sampling-UBMtest* (66.7% vs 76.2% accuracy), suggesting that using a different UBM for test MFCC feature vector sampling might not produce better test feature vectors for speaker recognition.

Overall, results suggest that for GMM-UBM based speaker recognition experiments, it is better to preserve the test MFCC feature vectors as they are. Altering the feature vectors based on GMM modelling assumptions worsens overall speaker recognition performance (since the *Gaussian Sampling* approach had the worst performance). Results also suggest that there are likely many deficiencies with the GMM modelling assumption. Imposing the Gaussian assumptions on the test MFCC feature vectors may remove many of the characteristics of the feature vectors that are helpful to speaker recognition. Lastly, the fact that the *Baseline* and *Empirical Sampling* approaches perform similarly suggests that the sequential ordering of test MFCC feature vectors is not essential to speaker recognition performance using the GMM-UBM approach. This agrees with the fact that GMMs does not consider time-dependence assumptions of MFCC feature vectors.

10. Summary and Future Work

This work attempts to quantify the effects of a couple of assumptions involving the use of GMMs for speaker recognition, and established a set of results showing the effects of the Gaussianity and time-independent assumptions on feature vectors, albeit on a small dataset. It is the start of a series of investigations on problems with GMM modelling assumptions for speaker recognition, in an attempt to improve its modelling approaches. Future work will involve performing more detailed analysis of the results. Future work could also include the use of a larger dataset with recordings of more voice samples to generate greater statistical significance in the results. The use of large-scale speaker recognition datasets from the NIST Speaker Recognition Evaluations [14] may also be considered. Future work could also extend investigations to other speaker recognition modelling techniques, such as the i-vector approach that's more appropriate for large-scale datasets containing thousands of hours of speech data.

11. References

- [1] 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", in *8th European Conference on Speech Communication and Technology*, Geneva, Switzerland, 2003.
- [2] P. Rose, *Forensic Speaker Identification*. London: Taylor & Francis, 2002.
- [3] J.P. Campbell, W. Shen, W.M. Campbell, R. Schwartz, J.F. Bonastre, and D. Matrouf, "Forensic Speaker Recognition", in *IEEE Signal Processing Magazine*, 2009, pp. 95-103.
- [4] D.A. Reynolds, T.F. Quatieri, and R. Dunn, "Speaker Verification using Adapted Gaussian Mixture Models," in *Digital Signal Processing*, Vol. 10 No. 3, 2000, pp. 19-41.
- [5] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, "Joint Factor Analysis Versus Eigenchannels in Speaker Recognition", in *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15(4), 2007, pp. 1435-1447.
- [6] N. Dehak, R. Dehak, P. Kenny, N. Brummer, P. Ouellet, and P. Dumouchel, "Support Vector Machines versus Fast Scoring in the Low-Dimensional Total Variability Space for Speaker Verification", in *Proceedings of Interspeech*, Brighton, UK, 2009, pp. 1559-1562.
- [7] L. Burget, P. Oldrich, C. Sandro, G. Oldrej, M. Pavel, and N. Brummer, "Discriminantly Trained Probabilistic Linear Discriminant Analysis for Speaker Verification", in *Proceedings of ICASSP*, Brno, Czech Republic, 2011.
- [8] A.P. Dempster, N.M. Laird, D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society, Series B*, Vol. 39(1), 1977, pp. 1-38.
- [9] J.B. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations", *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1*. University of California Press. 1967, pp. 281-297.
- [10] D. Gillick, L. Gillick, and S. Wegmann, "Don't Multiply Lightly: Quantifying Problems with the Acoustic Model Assumptions in Speech Recognition", in *Proceedings of ASRU*, 2011, pp. 71-65.
- [11] Y. Bar-Yosef and Y. Bistriz, "Adaptive Individual Background Model for speaker verification," in *Proceedings of Interspeech*, Brighton, UK, 2009.
- [12] Sargin, M.E., Aradhye, H., Moreno, P.J., Ming Zhao, "Audiovisual celebrity recognition in unconstrained web videos", in *ICASSP*, pp. 1977 - 1980 (2009)
- [13] M. Everingham, J. Sivic, A. Zisserman, "Hello! my name is... Buffy -- automatic naming of characters in TV video", *Proceedings of the British Machine Vision Conference*, Vol. 2, 2006
- [14] "The NIST Year 2012 Speaker Recognition Evaluation Plan", <http://www.nist.gov>, 2012.
- [15] S. Davis and P. Mermelstein, "Comparison of Parametric Representations of Monosyllabic Word Recognition in Continuously Spoken Sentences", in *Proceedings of ICASSP*, 1980.
- [16] N. Brummer and J. du Preez, "Application-Independent Evaluation of Speaker Detection", *Computer, Speech and Language*, Vol. 20, No. 2-3, April-July 2006, pp. 230-275.