Close Speakers Model and Comparative Study in Automatic Speaker Verification

Djellali Hayet[1] and Laskri Mohamed Tayeb[2]

[1,2]Department of computer science, Badji Mokhtar University, Annaba, Algeria,
[1]LRS Laboratory, Badji Mokhtar University, [2]LRI Laboratory, Badji Mokhtar University
hayetdjellali@yahoo.fr, laskri@univ-annaba.org

Abstract

The performance of speaker verification system degrades when the test segments are utterances of short duration, therefore, we investigate the use of model representing our target speaker with his close speaker and his own speech data. We propose to create a new Speaker Model who groups close speakers (CS) achieved with two clustering algorithms in Automatic Speaker Verification A.S.V. Intra and Inter speaker's variability are two clustering algorithm used in voice module. We compare the traditional approach which uses one specific customer model (Maximum a Posteriori Adaptation) with the Close Speaker model (Customers Families).Close Speaker Model (CSM) applied only when speaker model is weak achieves 42% of equal error rate. The results demonstrate that the log likelihood of close speakers is greater than the likelihood of client speaker. The false alarm from client and CSM are closest and we are constrained to enhance speaker model.

Keywords*: MAP Adaptation, Close Speaker Models, Vector quantization, Intra Speaker Variability, Gaussian Mixtures Models*

1. Introduction

Speaker recognition (SRE) aims to recognizing persons from their voice. No two individuals sound identical because their voice production organs are different, larynx sizes; vocal tract shapes [1]. Sub domain of SRE, Automatic Speaker Verification (ASV) is a technique where we have to decide (yes or No) if the acoustic signal and the identity proclaimed originate or not from the same person.

Among robust technique, support vector machine (SVM) and artificial neural networks (ANNs) are the discriminative approaches model the boundary between speakers. The generative models such Vector Quantization VQ and the Gaussian Mixture Models (GMMs) estimate the feature distribution within each speaker [2, 3]. Gaussian mixture model (GMM) is an efficient modeling approach and successful method for text-independent speaker recognition [3, 4].

In Training phase, a preprocessing step and feature extraction is realized then, modeling speakers client and impostors (called the world models UBM: Universal Background Models) by Gaussian mixture models (GMM). To discriminate between client and impostors, a GMM based background model is used to represent the impostor's characteristics [3].

During the test phase, the process starts with extracting acoustics vectors from test signal then the score is calculated (based on the client model and the world models), this score is compared to a threshold decision. The final decision is either acceptance or rejection.

In front of incomplete or few speech data, it was demonstrated that the performances of ASV system degrades and the speaker model is weak [5, 6]. The concept of close speakers were used in cohort models as impostor models in normalization technique, the objective was to set a value of threshold knowing that is difficult to fix it without any normalization[7][8].

We investigate the use of close speakers instead of target speaker because we try to capture the most significant common information for speaker population and use it only if the likelihood of close speaker is better than target speaker likelihood.

We consider the Close Speaker Model CSM as covering the space of speaker dependent broad acoustic classes of speech sounds, and then the acoustic classes not observed in target speaker during training stage are present in Close Speaker Model (CSM) contributing in efficient recognition. The clustering algorithm applied to obtain the best CSM model estimated with distance measure between target speakers (minimal distance).

The first idea comes is, this is a big attempt to the security level if we accept any close speaker acceding to the system. However, the weaknesses of the actual speaker speech data conduct to rejection and anyway this leads to increase the false reject error.

The on line customers accesses (for example Web sites), the required security level is not very constraining. Indeed, certain applications of ASV (others than bank accesses) prefer to authorize an impostor to reject a customer, but if our system is in front of a weak acoustic signal, why not use a group of close customers.

We aim to build a text-independent ASV system based on Maximum a Posteriori Adaptation for target speaker and close speakers (Close Speaker Model). The CS model is realized with two algorithms "Intra Speaker Variability" and "Inter Speaker Variability" in voice module. We constitute customers families which have the closet vocal characteristics and compare the traditional approach which uses one specific customer model with the second called Close speaker model CS (customer's families). However, the customer model is kept for comparative study.

The voice module also verifies whether the characterization of the customers in terms of pitch and formant, will therefore provide a better accuracy of belonging the test signal to the formant client area.

This paper checks if the close speaker model offers better accuracy when the speaker dependent model is weak in few data condition. We analyze the impact of clustering algorithms in modeling. We aim to reduce Equal Error Rate EER in condition of small training data of each customer.

We organized paper as follows, modeling and characterization speakers are introduced in Section 2, the architecture proposed in Section 3, the experiments in Section 4 follows by discussion in section 5 and finally the conclusion in 6.

2. Modeling and Speaker Characterization

2.1 Speakers Characterization

Fundamental frequency (F0) is the most important prosodic parameter. It is determined by the vocal cord vibrations. Combining F0-related features with spectral features has been shown to be effective, especially in noisy conditions [9, 10]. Hence an accurate F0 estimate calculated can be used in an algorithm for gender identification [11]. Several works have implemented pitch extraction algorithms based on computing the short time autocorrelation function of the speech signal [12].This parameters, formant F1, F2, F3 and F4 help us to improve the recognition, reason why we use it in voice module.

2.2 Modeling Speakers

GMM-UBM: This approach requires creating two models, the client model based on his data and the impostor's acoustic model (the world UBM) whose acoustic vectors are derived from a large population of speakers other than our customers. Training both GMM models achieved with the EM algorithm (Expectation-Maximization). However, GMM-UBM technique based on the estimation of Maximum Likelihood ML (Maximum Likelihood), suffers of over fitting when the speech duration of target speaker is low) [3, 4, 13].

GMM-MAP: It is difficult to provide sufficient amount of client speech, to resolve this problem, Maximum a posteriori adaptation have been proposed for creating low level acoustic speaker models from a moderate amount of client data [2, 3]. GMM-MAP approach provides superior performance over GMM-UBM system where the speaker model is trained independently of the UBM. This previous technique uses the world model and client training data to estimate the client model [9, 14, 15, 16, 17].

2.3 Clustering Algorithm

Clustering is needed in various applications such as speech and speaker recognition. Many clustering methods have been proposed. Kinnunen et al [18, 19, 20] have presented an extensive comparison of clustering methods and found the choice of the algorithm is critical only if very small model size is used. They recommend the random swap algorithm (RS) because of its simple implementation and robust performance in all test conditions. However, they recommend the SPLITT algorithm when running time is critical. Kmeans algorithm is widely used and effective in clustering and gives good performances [21].

We focus on K-means algorithm as it is one of the most used iterative partitional clustering algorithms and because it may also be used to initialize more expensive clustering algorithms (EM algorithm) [22]. However, it is established that the K-means algorithm suffers from initial starting conditions effects. This algorithm needs three user-specified parameters: Cluster initialization, number of clusters K, and distance metric [23].

Typically, K-means is executed for different values of K and the best partition is chosen. One way to overcome the local minima is to run the K-means algorithm, for a given K, with several different initial partitions and the minimal squared error is the criterion to select the partition. K-means is typically used with the Euclidean metric for computing the distance between points and cluster centers [24].

Despite being used in wide applications, K-means is not exempt of drawbacks listed below:

1. K-means only converges to local minim: Different initializations can lead to different final clustering.
2. The most critical choice is K. While no mathematical criterion exists, only heuristics are available for choosing K.

The main steps of K-means algorithm is as follows:

Kmeans algorithm:

1. *Select an initial partition with K clusters centroid μ1, μ2, μ3,… μk;*
2. *Repeat steps 3 and 4 until cluster membership stabilizes(convergence).*
   3. *Generate a new partition by assigning each pattern to its closest cluster center.*
   4. *Compute new cluster centers μ j. Centroid(i) = argmin || X(i) - μ j ||.*

3. Proposed Automatic Speaker Verification Architecture

We propose an ASV system based on Maximum a Posteriori Adaptation GMM-MAP helped by a voice module. This module contributes to choose the best close speakers of each genuine after applying two algorithms called speaker intra variability and inter variability. Figure 1 indicates the replacement of target speaker(SM) with close speakers only if the calculated score of SM Model is lower than CSM model. We describe different modules (Figure 2) of our ASV architecture which includes:

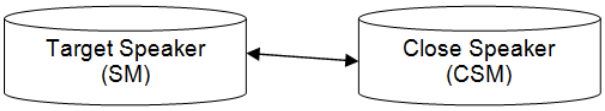

Figure 1. Close Speakers Models

3.1 Training Phase

We first build UBM model, both speaker model(SM) and Close speakers models CSM by Maximum a Posteriori Adaptation. The main steps are described below:

3.1.1 Preprocessing and Features Extraction P.F.E

- **Silence Detection SD**: We remove the frames of silence and noise that decrease the ASV system performance. The energy and ZCR (zero crossing rate criterion is used to select the frames of words (high energy) and remove frames of silence (low energy).
- **Features Extraction FE**: Cepstral analysis is used due to its robust estimation of noisy signal [2]. We extracted 13 cepstral coefficients and their derivatives and second derivative every 10ms calculated on an analysis window of 25ms hamming error. The cepstral mean is applied (Cepstral Mean Subtraction), removing the average distribution of each cepstral parameters.

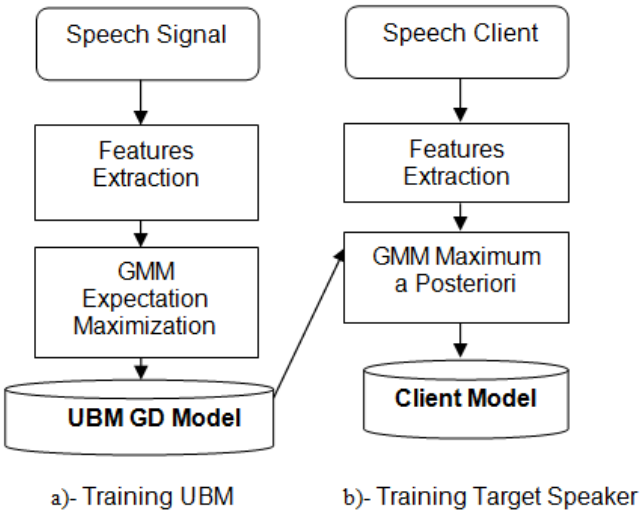

Figure 2. Training UBM and Target Speaker in ASV

3.1.2 Modeling:

- ***GMM-UBM-Maximum***

Likelihood Modeling: The traditional approach Gaussian mixtures models GMM is used for impostor's population modeling called UBM and trained with expectation maximization algorithm (EM). This previous algorithm provides a local maximum with three parameters of Gaussians (mean, covariance and weights). Two gender dependent UBM (male, female) model are trained with Expectation Maximization Algorithm (EM). The model's parameters (mean, covariance and weight of the Gaussian) are evaluated after few iteration of Expectation-Maximization algorithm.

- **GMM-MAP Speaker Adaptation**:

The client model (Speaker Model) is derived from the world model UBM by adapting the GMM parameters (mean, covariance, weights) and his speech. However, experimentally, only the averages of GMM are adapted [3, 15].

3.1.3 Voice Module

The voice module consists of two main phases and provides a subset of close speakers:

- **Algorithm 1: Speaker Intra- Variability**

This algorithm computes an optimal codebook representing each target speaker using the fundamental frequency and formant F1, F2, F3, F4. The k-means algorithm is called and the distance measure is the Euclidean distance.

- **Algorithm 2: Speaker Inter-Variability**

Speaker inter variability is a criteria allow us to discriminate between speaker with efficient manner, to achieve this goal, we implement an algorithm calling k-means algorithm and compute distance between each pair of speaker centroid. After first iteration, we obtain the best close speaker centroid and prune out this speaker from list of speakers, then, repeat the process until empty list. .

For each client, we extract acoustic features from customers' signal, F0 and formant parameters: F1, F2, F3, F4, then, comparing these parameters extracted from the test phase, we eliminate those whose gender is different from the test signal gender. We calculate the average pitch AvgF0 under matlab software. The pitch is extracted with autocorrelation method [10].

We use Kmeans [24] a popular clustering algorithm to classify target speaker and his close speakers, the pseudo code is:

- **Speakers Clustering Algorithm**

Assume the data lives in a Euclidean space, we want k classes for speaker intra variability. We applied speaker inter variability algorithm and use a form of Biclustering (subset of 2 speakers).

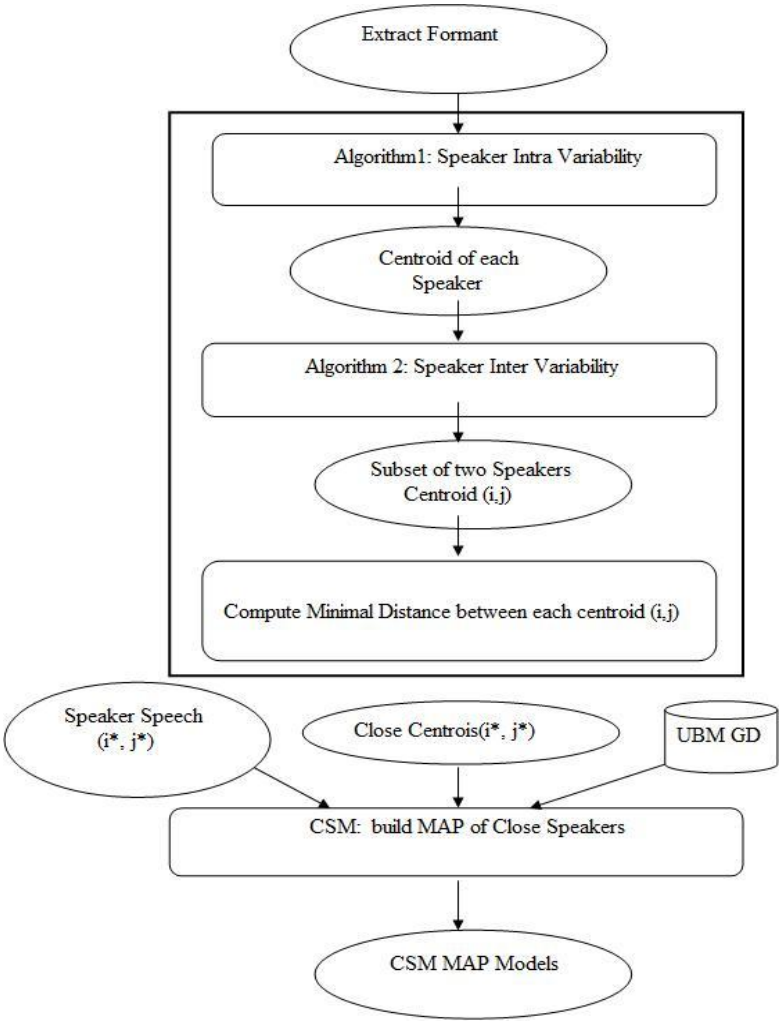

Figure 3. Voice Module Phases

Algorithm 1: Speaker Intra variability

We define k=4, 8, 16, 32, 64, 128 clusters;
For every speaker:
•Assume we start with randomly located speaker cluster centers.
The algorithm alternates between two steps:
Step 1: Assignment step: Assign each f0 & Formant (f0, F1, F2, F3, F4) to the closest cluster.
Step 2: Refitting step: Move each cluster center to the center of gravity of the data assigned to it.
Endfor;

We obtain speaker Centroids Dimension= cd=k*5=10 values. The kmeans algorithm give us cd*n vectors (n is the number of speakers).

Output: CD = k vectors;

Algorithm 2: Speaker Inter Variability

We then apply kmeans again between two different speakers:
Spk1(f0,f1,f2,f3,F4; f0',f1',f2',f3',F4');
Train Phase:
1. ***For T = 1 to n-1 do***
Begin
2. ***For j =T+1 to n do***
Begin
3. *Apply k-means between speaker T and j with k= 4,8,16,32,64,128 clusters;*
4. *Store centroid;*
End;
5. *We select speaker centroid constraint to: minimal distance between speaker T and all others speakers j,*
6. *Speaker j* = min (distance (centroid T, centroid j))*

End
Output: close speakers I and j (centroid I, centroid j*);*

We select this subset (2 speakers) and create their models with GMM MAP.

3.2 Test Phase

3.2.1 Parameterization

The acoustics test vectors are extracted from speaker speech after removing silence frame. Each enrollment is between 10 to 15 seconds. The MFCC are 13, their derivatives and second derivative are also calculated.

3.2.2 Decision

The log likelihood ratio were applied to decide the acceptance or rejection, the score will be calculated as follows:

Log(p(X | λ client) : Client Model Score proclaimed
Log (p(X | λ_{CSM}) : Close Model Score calculated by the voice module(close speakers).
Log (p(X | λ_{UBMF})): Score from the female world model.
Log (p(X | λ_{UBMM})): Score from the male world model:

- ***First case*** **:**

If (LLR(p(X| λclient) > LLR(p(X | λ_{CSM}))) then $\Lambda(X)$ is computed like this :

$\Lambda(X) = \Lambda1(X)$ if Voice Module determins a men
$\Lambda(X) = \Lambda2(X)$ Else women

Knowing that $\Lambda1(X)$ et $\Lambda2(X)$ are calculated as follows:

$$\Lambda 1(X) = \log p(X|\lambda \text{ client}) - \log p(X|\lambda \text{ UBMM}) \quad (1)$$

$$\Lambda 2(X) = \log(p(X|\lambda \text{ client}) - \log (p(X|\lambda \text{ UBMF})) \quad (2)$$

We compared to a threshold θ: If $\Lambda(X) > \theta$ client acces Else impostor.

- **Second Case :** with Close Speaker Models

$$\Lambda 3(X) = \log(p(X|\lambda \text{ CSMM}) - \log (p(X|\lambda \text{UBMM})) \quad (3)$$

$$\Lambda 4(X) = \log(p(X|\lambda \text{ CSF}) - \log (p(X|\lambda \text{ UBMF})) \quad (4)$$

If (LLR ($p(X|\lambda$client) less than LLR($p(X|\lambda$ CSM)), we consider that either give the test data are corrupt or the deviation between the training data and test. In this case, we compute the score with formula (3) and (4).

4. Experimental Results

The described speaker verification system has been developed and evaluated using recorded Arabic speech database of 56 speakers. In this section we will describe the corpus, the voice module and the performance of speaker verification system based on clustering algorithm.

4.1 Database and Baseline System

The database is recorded with Goldwave at 16KHz frequency during 30s for each speaker in training and 10s in the test. The UBM population is 15 men's and 15 women. Three sessions are recorded for each speaker with 10 utterances at interval of 1 month. Ten clients have been registered in database (5 male and 5 female speakers). The threshold has been computed with 16 speakers (8males and 8females).

4.2 Comparative Study of Client and CSM Likelihood

The Figure 4 shows the LLR($p(X/\lambda$ client)) < LLR($p(X/\lambda$ CSM)), and we have to improve client model or change it in few client data condition. In this case, client model is weak. We confirm our assumption in this example.

Figure 4 also demonstrates three females speakers taken as example ,the first one is the log likelihood value and the second represents a close Speaker Model likelihood constituted with a set of two real speakers 1 and 2 and the third one is the second close speaker model likelihood grouping real speaker 1 and 3. The score of CSM 1 and 2 is better than other models; however, the speaker model (blue) should be higher because this is the real target speaker. We conclude the CSM 1&2 can easily replace the target speaker even this is an impostor.

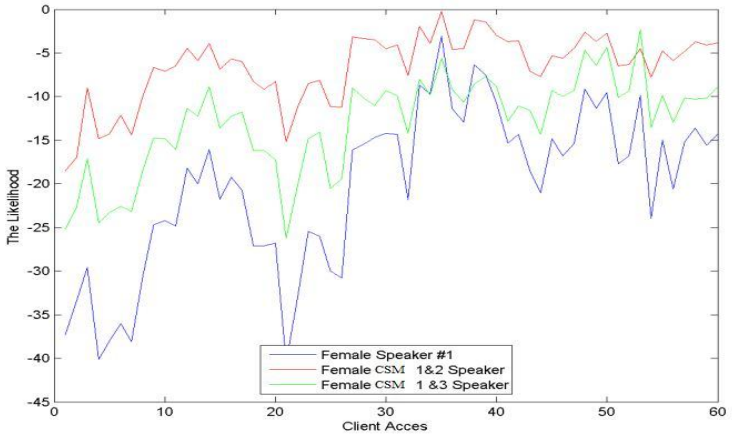

Figure 4. Log Likelihood of Client and CSM Model

4.3 Speakers Models

Three models were built, speakers models, UBM and CSM models, We carry out the training from the nearest customer by voice module (F0 and formant F1, F2, F3, F4). Each client is trained by MAP adaptation. We built UBM models from 30 arabic speakers; UBM male with 15 male speakers and UBM female from 15 female speakers.

We test 8, 16, 32, 64, 128 Gaussians and classify them by gender (male, female) with vocal module. The global threshold is computed from other database: 8 male and 8 female speakers. We get for GMM MAP models the result in table 1 with different mixtures sizes. The Euclidean distance calculates the distance values of different speakers.

Table 1. GMM MAP Baseline Performance

#Gaussians	8	16	32	64	128
GMMMAP%	19.16	36.12	35.2	35	36.04

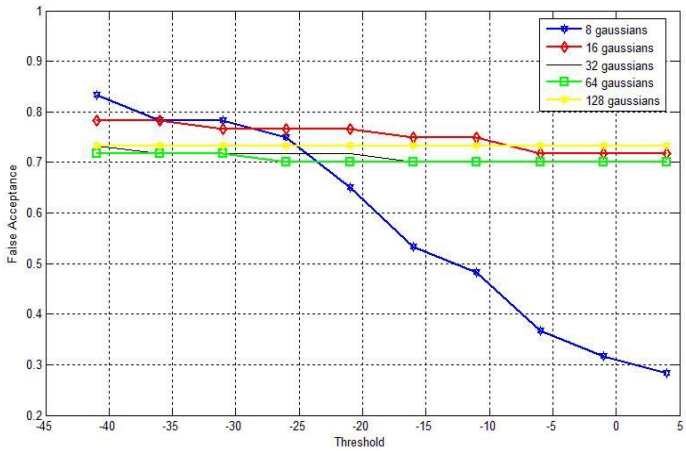

Figure 5. GMM MAP Baseline System False acceptance Error

4.4 Voice Module

Pitch Extraction for gender detection: The speech signal is divided into segments of 60 ms, each segment is extracted every 50ms interval and requires a function autocorrelation pitch to estimate the fundamental Frequency of this segment. This algorithm was tested on speech samples from people of different gender from the basis with 16khz sampling frequency. Detection Gender Errors are 2%.

We used two main matlab programs, one for extracting the fundamental frequency (average pitch) and the second calculate the parameters F1, F2, F3, F4, gender. For each speaker, we tried 5 enrollments, therefore, there are intervals for which formant belongs and used to identify the speaker. Table 2 shows the result of 3 males (M) and 3 females speaker(F), under praat2 , three formant F1,F2,F3,F4.

Table 2. Pitch Speakers Values

Number gender	F0 Praat
1M	174,763
2M	229,254
3M	278,615
1F	172,379
2F	202,008
3F	199,219

4.5 Speaker Intervariability

This algorithm is applied to 8 speakers with formant characteristics. At each iteration, it found the two nearest speakers with euclidean distance. It stores their index, then repeat the process until there is no speaker. We sort the distance increasely and move on to next speaker and repeat the process.

Table 3. Speaker Inter Variability Results with Distance Measure

Speakers #	Distance	Close Speakers	After .1 th iteration	Store
1	637,42	3	1	1 1,3 store
2	745,27	7	2	2 2,7 store
3	717,88	4	3	3 3,4 store
4	712,79	6	4	4 4,6 store
5	879,44	7	5	5 5,7 store
6	776,80	7	6	6 6,7 store

5. Discussions

We make the following observations: table 1 shows GMMMAP modeling approach achieves 19.16% of equal error rate for M=8 mixtures where the values of M=16, 32, 64, 128 mixtures vary between [35% - 36.12%]. The performance of the baseline system is low for m greater than 8 mixtures. The reason is the small quantity of data during training.

The results at figure 5 show the false acceptance errors of the GMMMAP Baseline Models. It indicates the lowest error FA=23.33% is given by M=8 mixtures where the threshold=4. Figure 7 indicates the false acceptance errors of the Close Speakers Models and we observe that the values are very close to GMMMAP for 8 mixtures.

Figure 8 gives the false rejection of CSM models and we have to consider the compromise between false acceptance and false rejection. The Equal Error Rate (value of threshold where FA=FR) is 42% for model order=8 gaussians.

Table 4. False Acceptance of GMMMAP and CSM MAP Models

Threshold	GMMMAP	CSM MAP
	FA%	FA%
-41	85.00	83.33
-36	80.00	78.33
-31	80.00	78.33
-26	73.33	75.00
-21	61.67	65.00
-16	55.00	53.33
-11	50.00	48.33
-6	35.00	36.67
-1	28.33	31.67
4	23.33	28.33

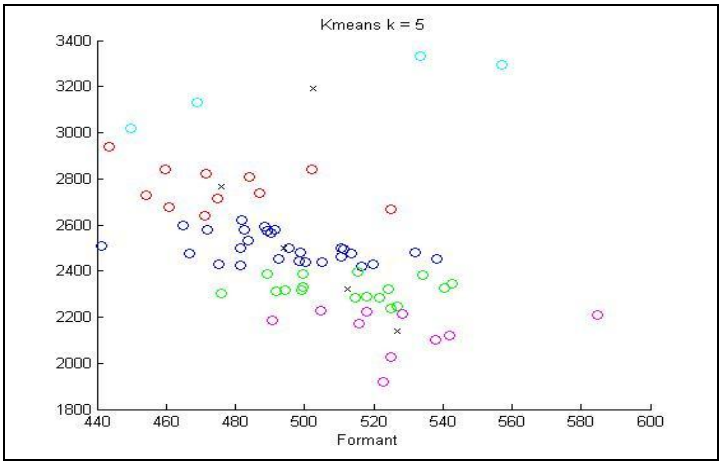

Figure 6. kmeans k=5, F2 = function(F1)

Algorithm 1: The values of formants are close to each others from one session to another for the same speaker. As example the figure 6 show us the intra speaker clustering is not significant for k greater than 3 , for this reason, we choose k=2 in other to get the distance as wide as possible for keeping the maximum variance of data(formant). The clustering built a subset of customers and contributes in well speakers modeling.

Algorithm 2: The results of this algorithm is shown in table 3, the nearest speakers are numbered like follows (1,3) , (4,6) , (3,4) , (2,7), (6,7) . We constitute the CSM models from this subset of speakers. For example, we create the GMMMAP of both speaker 1 and 3

called CSM Model. We do the same thing for (4,6),(3,4),(2,7),(6,7). We observe the speakers(1,3) and (3,4) are close therefore we create CSMMAP of this three speakers(1,3,4).

We observe that the baseline GMM MAP system is better than CSM MAP model. Equal Error Rate EER =19% for GMMMAP when EER = 42% for CSM MAP. This involves that the CSM model cannot efficiently replace speaker model. In addition to that, we observe the value of EER for GMMMAP is worst for model order= 16, 32, 64, 128.

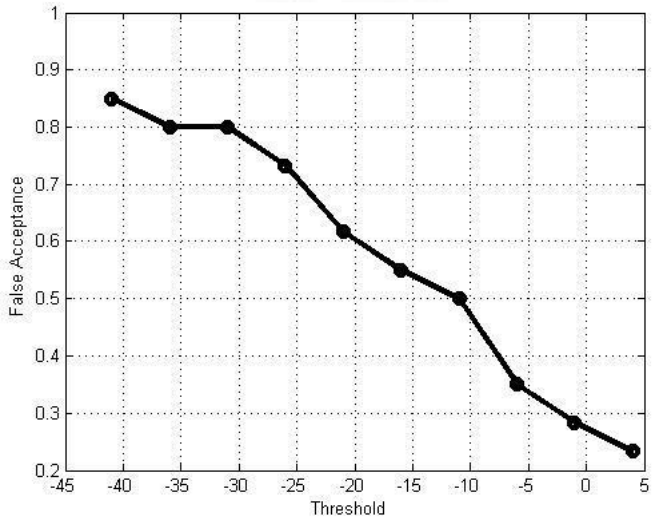

Figure 7. CSM MAP False Acceptance (FA)

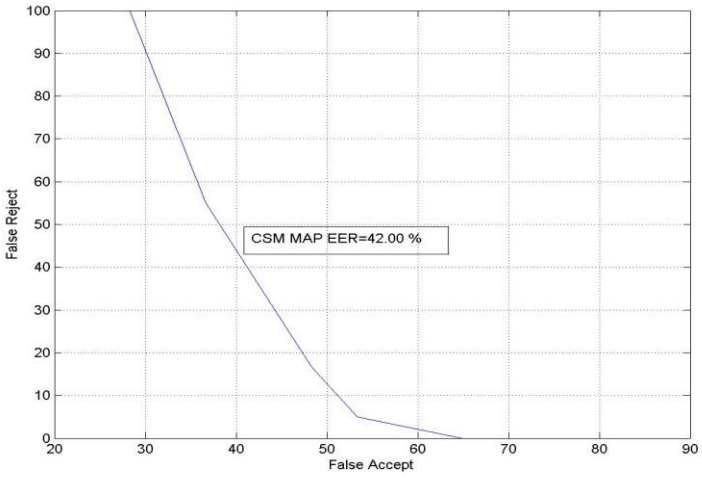

Figure 8. Close Speaker Models Equal Error Rate

6. Conclusion

We aim to improve the score and thus the final decision with Close speaker models helped by voice module. The idea is to construct a family of speakers near the customer able to replace our target speaker model if client is inadequate (no sufficient data or bad records). We

proved the Speaker MAP Model give us the LLR less than the CSM Model in condition of few speaker data and we have to improve speaker model. The close speaker model gives EER=42% for 8 mixtures.

The clustering algorithm in modeling can affect the global accuracy of the ASV system because his ability to create a CSM models which are close to the target speaker and can lead to accept the CSM instead of our client. Preliminary results indicate that MAP models are better but we cannot generalize because the first reason is the no sufficient size of client , we experiment only ten speakers and second reason is we should complete the test with model order M=16,32,64,128 mixtures.

In this paper, we obtain the close speaker model don't offer better accuracy when the speaker dependent model is weak in few data condition. We analyze the impact of clustering algorithms in modeling and achieved the CSM model which EER is worst but close to GMMMAP technique error. We have to investigate the use of supplementary acoustics parameters in voice module to improve the CSM model.

References

[1] J. P. Campbell, "Speaker Recognition: a Tutorial", Proceeding IEEE, vol. 85, No. 9, **(1997)**, pp. 1437-1462.

[2] T. Kinnunen and H. Li, "An Overview of Text Independent Speaker Recognition from Features to Supervectors", Speech Communi., 52(1), **(2007)**, pp. 12-40.

[3] D. A. Reynolds, T. F. Quatieri and R. B. Dunn, "Speaker verification using adapted Gaussian Mixture Models", Digital Signal Process, 10, **(2000)**, pp. 19–41.

[4] D. A. Reynolds and W. M. Campbell, "Text Independent Speaker Recognition", Springer Handbook of Speech Processing", Springer Handbook, (**2008),** pp. 763-779.

[5] J. M. Karen, K. J. Epps, M. Nosratighods, E. Ambikairajah and E. H. C. Choi, "Using Clustering Comparison Measures for Speaker Recognition", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, **(2011**), pp. 5452-5455.

[6] L. Qi, J. Tsai, A. Tsai and Q. Zhou, "Robust Endpoint Detection and Energy Normalization for Real Time Speech and Speaker Recognition", IEEE Trans on Speech and Audio Processing, vol. 10, no. 3, **(2002)**, pp. 146-157.

[7] R. A. Finan, A. T. Sapeluk and R. I. Damper, "Impostor Cohort Selection for Score Normalization in Speaker Verification", Pattern Recognition Letter, **(1997),** pp. 881-888.

[8] D. Sturim and D. A. Reynolds, "Speaker Adaptative Cohort Selection for Tnorm in Text Independent Speaker Verification", in Proceeding of the IEEE International Conference on Acoustics, Speech, and Signal Processing, **(2005)**, pp. 741-744.

[9] A. Adami, R. Mihaescu, D. A.Reynolds and J. J. Godfrey, "Modeling Prosodic Dynamics for Speaker Recognition", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 788-791, Hong Kong, China, **(2003).**

[10] M. Kamran and I. C. Bruce, "Robust Formant Tracking for Continous Speech with Speaker Variability", IEEE Transactions on Speech and Audio Processing, vol. 14, (**2006),** pp. 435-444.

[11] V. Hautamäki, T. Kinnunen, I. Kärkkäinen, J. Sa,astamoinen, M. Tuononen and P. Fränti, "Maximum a Posteriori Adaptation of the Centroid Model for Speaker Verification", *IEEE Signal Processing Letters*, Vol. 15, **(2008),** pp 162-165.

[12] K. Kasi, "Yet Another Algorithm For Pitch Tracking", Thesis Andhra University, India. Master Of Science Electrical Engineering, **(2009)**.

[13] R. P. Ramachandran, K. R. Farrell and R. J. Mammone, "Speaker Recognition General Classifier Approaches and Data Fusion Methods", Pattern Recognition, 35(12), pp. 2801-2821, **(2002)**.

[14] B. Vesnicer and F.Mihelic, "The likelihood Ratio Decision Criterion for Nuisance Attribute Projection in GMM Speaker Verification", Hindawi Publishing Corporation Eurasip Journal On advances in Signal Processing vol. 2008, pp 1-10, **(2008)**.

[15] A. Preti, "Surveillance de Réseaux Professionnels de Communication par la Reconnaissance du Locuteur", thesis Académie d'Aix Marseille, Laboratoire d'informatique d'Avignon, **(2008)**.

[16] F. Bimbot, J. F. Bonastre C. Fredouille. G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz and D. A. Reynolds, "A Tutorial on Text-Independent Speaker Verification", Journal Applied Signal Processing,Vol. 4, pp. 430–451, **(2005)**.

[18] T. Kinnunen, I. Sidoroff, M. Tuononen and P. Franti, "Comparison of Clustering Algorithm Methods: A Case Study of Text Independent Speaker Modeling", Pattern Recognition Letters, 32, pp. 1604-1617, **(2011)**.

[19] T. Kinnunen, Kilpelainen and P. Franti, "Comparison of Clustering Algorithm in Speaker Identification", In Proceeding 28 International. Conference of Signal Processing and Communication, **(2000)**, pp. 222-227. Marbella, Spain.

[20] A. K. Jain and R. C. Dubes, "Algorithms for Clustering Data", Prentice Hall, **(1988)**.

[21] E. W. Forgy, "Cluster Analysis of Multivariate Data: Efficiency Versus Interpretability of Classification", Biometrics, (**1965**).

[22] A. P. Dempster and B. Series, "Maximum Likelihood from Incomplete Data via the EM Algorithm", Journal of the Royal Statistical Society, vol. 39, no 1, pp. 1–38, (**1977)**.

[23] K. Anil and Jain, "Data Clustering: 50 Years Beyond K-Means", Pattern Recognition Letters, Vol. 31, No. 8. pp. 651-666, **(2010)**.

[24] Y. Linde, A. Buzo and R. M. Gray, "An Algorithm for Vector Quantizer Design", IEEE Trans. Commun.28 (1), pp. 84-95, **(1980)**.

Authors

DJELLALI Hayet

DJELLALI Hayet was born on 16th of December 1969 in Annaba from Algeria; she graduated from the University of Badji Mokhtar Annaba, Algeria, with a state Engineering degree in Computer Science, in June 1994. She is a doctoral student since 2008, and joined the department of computer science in her original university of Badji Mokhtar Annaba in 2009 as an assistant teacher. Her main area of expertise is speaker recognition.

LASKRI Mohamed Tayeb

Mohamed Tayeb Laskri was born at Annaba (Algeria) in 1958. He obtained a 3rd cycle doctorate on computer science (France, 1987). He obtained his PhD degree from the Annaba University (Algeria, 1995). He is chief of the Research Group in Artificial Intelligence within laboratory LRI. Its current research takes the reasoning in artificial intelligence like field of application privileged in particular in the image processing, the multi-agents systems, the engineering of the Human-machine interfaces and the automatic treatment of the natural language. He was President of the University of Annaba and the Scientific Council of the University of Annaba on 2002-2010. He was President of the Consortium CEMUR «collaboration Europe-Maghreb of Universities in Network» in 2009. He is Founder member of Arab Computer Society (ACS).