Speaker Clustering of Unknown Utterances Based on Maximum Purity Estimation

Wei-Ho Tsai and Hsin-Min Wang
Institute of Information Science, Academia Sinica, Taipei, Taiwan, Republic of China
{wesley,whm}@iis.sinica.edu.tw

Abstract

This paper addresses the problem of automatically grouping unknown speech utterances that are from the same speaker. A clustering method based on maximum purity estimation is proposed, with the aim of maximizing the similarities of voice characteristics between utterances within all the clusters. This method employs a genetic algorithm to determine the cluster where each utterance should be located, which overcomes the limitation of conventional hierarchical clustering that the final result can only reach the local optimum. The proposed clustering method also incorporates a Bayesian information criterion to determine how many clusters should be created.

1. Introduction

Speaker clustering refers to the task of grouping unknown speech utterances together based on their associated speakers. With the burgeoning availability of digital audio material, speaker clustering is gaining importance as a means to index the voluminous spoken data accumulated daily for archival use [1]. It is hoped that by grouping same-speaker utterances into clusters, the human efforts required for indexing can be greatly reduced, from having to listen to each utterance to only having to check few utterances in each cluster.

Currently, most speaker-clustering methods follow a hierarchical clustering (HC) framework [2-12], which computes the similarities of voice characteristics between utterances, and then sequentially merges the utterances deemed similar to each other (agglomerative clustering), or alternatively, separates the utterances deemed dissimilar to each other (divisive clustering). During the procedure of agglomeration or division, the nearest neighborhood selection rule is usually employed in an attempt to maximize the similarities between all the utterances in each cluster. However, since no consideration is taken in respect of the interaction between clusters, HC can only make each individual cluster as homogeneous as possible, but cannot guarantee that the homogeneity for all the clusters can be summed to reach a maximum. In particular, mis-clustering errors, arising from grouping different-speaker utterances together or segregating same-speaker utterances apart, can propagate down the whole process, and hence limit the clustering performance.

To overcome the HC's limitation, this study develops a new clustering method which aims to maximize the total number of within-cluster utterances from the same speakers. In distinction to HC, which performs optimization in a cluster-by-cluster manner, the proposed method searches for the best partitioning of utterances by considering all the clusters at the same time. This is done with the estimation of the so-called *cluster purity* [5], in conjunction with an optimization process based on a *genetic algorithm* [13] for attaining maximal cluster purity.

2. Method Overview

Our speaker-clustering system is designed with an aim to take as input N isolated utterances produced by P unknown speakers[1], where P is also unknown, and to provide as output M clusters satisfying that $M = P$ and each cluster consists exclusively of utterances from only one speaker. The clustering procedure begins with the computation of inter-utterance similarities, followed by the determination of which utterances are similar enough to be grouped into a cluster. Since no information about the speaker population is available beforehand, the procedure also includes the estimation of the optimal number of clusters.

The performance of speaker clustering is evaluated on the basis of cluster purity [5], defined by

$$\rho_m = \sum_{p=1}^{P} \left(\frac{n_{mp}}{n_m} \right)^2, \tag{1}$$

where ρ_m is the purity of the m-th cluster, n_m is the total number of utterances in the m-th cluster, and n_{mp} is the number of utterances in the m-th cluster that are produced by the p-th speaker. Eq. (1) follows that $n_m^{-1} \le \rho_m \le 1$, in which the upper bound and lower bound reflect that all the within-cluster utterances are from the same speaker or completely different speakers, respectively. To evaluate the overall performance of M-clustering, we compute an average cluster purity:

$$\overline{\rho} = \frac{1}{N} \sum_{m=1}^{M} n_m \rho_m. \tag{2}$$

3. Inter-utterance Similarity Computation

The method for measuring the inter-utterance similarities is adapted from our previous work on "eigenvoice-motivated vector space" [12]. To begin, a Gaussian mixture model (GMM), which represents the generic characteristics of speakers' voices, is created using the cepstral features of all the utterances to be clustered. This GMM is then adapted to model the individual voice characteristics of each utterance using *maximum a posteriori* estimation [14], and therefore N utterance-dependent GMMs $\lambda_1, \lambda_2, \ldots, \lambda_N$ are generated.

Next, all the mean vectors of each utterance-dependent GMM are concatenated in the order of the mixture index to form a super-vector with dimension of D. *Principal component analysis* is then applied on the set of N super-vectors, $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_N$, obtained from $\lambda_1, \lambda_2, \ldots, \lambda_N$. This yields D eigenvectors, $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_D$, ordered by the degree of their contribution to the between-utterance covariance matrix:

$$\mathbf{B} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{V}_i - \overline{\mathbf{V}})(\mathbf{V}_i - \overline{\mathbf{V}})', \tag{3}$$

[1] Each utterance is assumed to be produced by only one speaker.

where $\overline{\mathbf{V}}$ is the mean vector of all $\mathbf{V}_i$ for $1 \le i \le N$. These eigenvectors constitute a voice characteristic space, which characterizes the relationship between utterances as their coordinates on the space. The coordinate of each utterance, $\mathbf{w}_i = (w_{i,1}, w_{i,2}, \ldots, w_{i,D})$, $1 \le i \le N$, is computed using

$$w_{i,d} = < (\mathbf{V}_i - \overline{\mathbf{V}}), \mathbf{e}_d > . \quad (4)$$

To capture the most representative voice characteristics, we only retain low-order K ($K < D$) eigenvectors with larger eigenvalues that reflect more variation between utterances. Accordingly, the similarity between any two utterances, say $\mathbf{X}_i$ and $\mathbf{X}_j$, can be computed using the cosine measure between $\mathbf{w}_i$ and $\mathbf{w}_j$, i.e.,

$$S(\mathbf{X}_i, \mathbf{X}_j) = \frac{\mathbf{w}_i \cdot \mathbf{w}_j}{\| \mathbf{w}_i \| \| \mathbf{w}_j \|}. \quad (5)$$

4. Maximum Purity Clustering

Let h_i denote the index of the cluster where an utterance $\mathbf{X}_i$ should be located, and o_i denote the true speaker of utterance $\mathbf{X}_i$. Note, h_i is an integer between 1 and M when the number of clusters M is specified *a priori*, and o_i is an integer between 1 and P if there are P speakers involved. Our aim is to find a set of cluster indices $\mathbf{H} = h_1, h_2, \ldots, h_N$ for N utterances to be clustered that maximizes the average cluster purity, i.e.,

$$\begin{aligned} \mathbf{H}^* &= \arg\max_{\mathbf{H}} \frac{1}{N} \sum_{m=1}^{M} n_m \left(\sum_{p=1}^{P} \frac{n_{mp}^2}{n_m^2} \right) \\ &= \arg\max_{\mathbf{H}} \frac{1}{N} \sum_{m=1}^{M} \frac{\sum_{p=1}^{P} \left[\sum_{i=1}^{N} \delta(h_i, m) \delta(o_i, p) \right]^2}{\sum_{i=1}^{N} \delta(h_i, m)}, \end{aligned} \quad (6)$$

where $\delta(\cdot)$ is a Kronecker Delta function.

However, as the computation of cluster purity requires that the true speaker of each utterance is known in advance, it is impossible to find $\mathbf{H}^*$ from Eq. (6) directly. To make this equation solvable, we need to estimate the term $\Sigma_{p=1}^{P}[\Sigma_{i=1}^{N}\delta(h_i, m)\delta(o_i, p)]^2$ in the absence of the ground truth. Since

$$\begin{aligned} &\sum_{p=1}^{P} \left[\sum_{i=1}^{N} \delta(h_i, m) \delta(o_i, p) \right]^2 \\ &= \sum_{p=1}^{P} \left[\sum_{i=1}^{N} \delta(h_i, m) \delta(o_i, p) \right] \left[\sum_{j=1}^{N} \delta(h_j, m) \delta(o_j, p) \right] \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{p=1}^{P} \delta(h_i, m) \delta(o_i, p) \delta(h_j, m) \delta(o_j, p) \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} \delta(h_i, m) \delta(h_j, m) \left[\sum_{p=1}^{P} \delta(o_i, p) \delta(o_j, p) \right] \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} \delta(h_i, m) \delta(h_j, m) \delta(o_i, o_j), \end{aligned} \quad (7)$$

the estimation of $\Sigma_{p=1}^{P}[\Sigma_{i=1}^{N}\delta(h_i, m)\delta(o_i, p)]^2$ hinges on how to determine the term $\delta(o_i, o_j)$ when utterances $\mathbf{X}_i$ and $\mathbf{X}_j$ are located in the m-th cluster. Motivated by Solomonoff *et al.*'s work [5], we determine $\delta(o_i, o_j)$ by using the following approximation:

$$\hat{\delta}(o_i, o_j) = \begin{cases} 1, & \text{if } i = j \\ \dfrac{S(\mathbf{X}_i, \mathbf{X}_j)}{S(\mathbf{X}_i, \mathbf{X}_{\xi_i})}, & \text{if } i \ne j, \text{and } R[S(\mathbf{X}_i, \mathbf{X}_j)] \le n_m \\ 0, & \text{otherwise} \end{cases}, \quad (8)$$

where $R[S(\mathbf{X}_i, \mathbf{X}_j)]$ denotes the rank of inter-utterance similarity $S(\mathbf{X}_i, \mathbf{X}_j)$ among $S(\mathbf{X}_i, \mathbf{X}_1)$, $S(\mathbf{X}_i, \mathbf{X}_2)$, …, $S(\mathbf{X}_i, \mathbf{X}_N)$ in descending order, and ξ_i is the utterance most similar to $\mathbf{X}_i$, i.e., $R[S(\mathbf{X}_i, \mathbf{X}_{\xi_i})] = 2$. Implicit in Eq. (8) is the idea that two utterances, $\mathbf{X}_i$ and $\mathbf{X}_j$, can only be considered as being from the same speaker if the similarity $S(\mathbf{X}_i, \mathbf{X}_j)$ is high enough to satisfy $R[S(\mathbf{X}_i, \mathbf{X}_j)] \le n_m$. In addition, to avoid a possible misjudgement arising from an over-large n_m, we approximate $\delta(o_i, o_j)$ as a probability that utterances $\mathbf{X}_i$ and $\mathbf{X}_j$ belong to the same speaker. This probability is measured by comparing the similarity $S(\mathbf{X}_i, \mathbf{X}_j)$ with that of the two utterances that most likely belong to the same speaker, i.e., $S(\mathbf{X}_i, \mathbf{X}_{\xi_i})$. With this approximation, an optimal $\mathbf{H}^*$ may be found according to

$$\mathbf{H}^* = \arg\max_{\mathbf{H}} \frac{1}{N} \sum_{m=1}^{M} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \delta(h_i, m) \delta(h_j, m) \hat{\delta}(o_i, o_j)}{\sum_{i=1}^{N} \delta(h_i, m)}. \quad (9)$$

We note that the solution to Eq. (9) remains non-trivial, since a gradient-based optimization cannot be used in this scenario. Moreover, it is infeasible to perform an exhaustive search that would examine all possible solutions to determine the best one, because there are M^N possible combinations of cluster indices. To overcome these difficulties, we apply the genetic algorithm (GA) [13] to find $\mathbf{H}^*$ by using GA's global scope and parallel searching power.

The basic operation of the GA is to explore a given search space in parallel by means of iterative modifications of a population of chromosomes. Each chromosome, encoded as a string of alphabets or real numbers called *genes*, represents a potential solution to a given problem. In our task, a chromosome is exactly a legitimate $\mathbf{H}$, and a gene corresponds to a cluster index associated with an utterance. The GA optimization starts with a random generation of chromosomes according to a certain population size. Then, the fitness of all chromosomes is evaluated and ranked on the basis of the estimated average purity, i.e.,

$$\overline{\rho}(\mathbf{H}) = \frac{1}{N} \sum_{m=1}^{M} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \delta(h_i, m) \delta(h_j, m) \hat{\delta}(o_i, o_j)}{\sum_{i=1}^{N} \delta(h_i, m)}. \quad (10)$$

As a result of this evaluation, a particular group of chromosomes is selected from the population to generate offspring by subsequent recombination.

Next, crossover among the selected chromosomes proceeds by exchanging the substrings of two chromosomes between two randomly selected crossover points. A crossover probability is assigned to control the ratio of the number of offspring produced in each generation to the population size. After crossover, a mutation operator is used to introduce random variations into the genetic structure of the chromosomes. This is done by generating a random number and then replacing one gene of an existing chromosome with a mutation probability. The procedure of fitness evaluation, selection, crossover, and mutation is repeated continuously. It is hoped that the average purity of the clustering will increase from generation to generation. When the maximum number of generations is reached, the best chromosome in the final population is taken as the solution of $\mathbf{H}^*$.

5. Speaker Population Estimation

In general, the more clusters we generate, the larger the value of purity we can obtain. However, if we generate too many clusters, a single speaker's utterances would be split across multiple clusters, so the speaker clustering would be

incomplete. Clearly, the optimal number of clusters is equal to the speaker population, which is unknown and needs to be estimated.

Our basic strategy for estimating the speaker population is to define a score for assessing a partitioning of the utterances based on how a large average purity can be achieved at the expense of increasing the number of clusters. This problem may be tackled from the standpoint of model selection. Specifically, if each of the possible partitionings with different numbers of clusters is considered as a model for characterizing the speaker information of the utterances, we choose the model that can produce the largest average purity and has the smallest number of clusters. Viewed in this manner, the Bayesian information criterion (BIC) [15], which is popular for solving model-selection problems, could be used to assess the clustering.

The BIC scores a parametric model based on how well the model fits a data set, and how simple the model is:

$$\mathrm{BIC}(\Lambda) = \log \Pr(\mathbf{O} \mid \Lambda) - 0.5\,\gamma\,\#(\Lambda)\log|\mathbf{O}|, \quad (11)$$

where $\#(\Lambda)$ denotes the number of free parameters in model Λ, $|\mathbf{O}|$ is the size of the data set $\mathbf{O}$, and γ is a penalty factor. The larger the value of $\mathrm{BIC}(\Lambda)$, the better model Λ will perform. In another work on speaker clustering [6], BIC is applied to score a partitioning of an utterance collection, in which a cluster is represented by a uni-Gaussian density estimated from the feature vectors of the utterances, and the model Λ is a set of Gaussian densities. Since we convert each utterance from the feature vectors into a coordinate, our work differs from [6] by the way clusters are modeled, which is directly related to the clustering performance.

Consider a model Λ, consisting of M parameters for classifying a set of N utterances from P unknown speakers. Each of the parameters represents an integer index to tag each of the utterances. The model is designed with such an aim that, by having all the utterances tagged, the utterances belonging to the same speakers are tagged with the same index. Thus, the likelihood $\Pr(\mathbf{O}|\Lambda)$, which measures how well the model fits the data, is concerned with the probability that, given N indices $h_1, h_2, \ldots, h_N$ for the N utterances, $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$, the utterances tagged with the same indices come from the same speakers. Suppose that the true speakers of N utterances, $o_1, o_2, \ldots, o_N$, are statistically independent with each other. We compute the likelihood $\Pr(\mathbf{O}|\Lambda)$ by

$$\begin{aligned}\Pr(\mathbf{O} \mid \Lambda) &= \prod_{i=1}^{N} \Pr(o_i = O(\mathbf{X}) \mid h_i = C(\mathbf{X})) \\ &\approx \left[\Pr(O(\dot{\mathbf{X}}) = O(\mathbf{X}) \mid C(\dot{\mathbf{X}}) = C(\mathbf{X}))\right]^N,\end{aligned} \quad (12)$$

where $O(\mathbf{X})$ denotes the true speaker of an arbitrary utterance $\mathbf{X}$ tagged with index $C(\mathbf{X})$, and $\Pr(o_i = O(\mathbf{X}) | h_i = C(\mathbf{X}))$ represents the probability that, given an arbitrary utterance $\mathbf{X}$ tagged with the same index as utterance $\mathbf{X}_i$, the true speakers of utterances $\mathbf{X}_i$ and $\mathbf{X}$ are the same. For computational efficiency, all the probabilities $\Pr(o_i = O(\mathbf{X}) | h_i = C(\mathbf{X}))$, $1 \le i \le N$, are approximated by $\Pr(O(\dot{\mathbf{X}}) = O(\mathbf{X}) \mid C(\dot{\mathbf{X}}) = C(\mathbf{X}))$, which is the probability that any two utterances $\mathbf{X}$ and $\dot{\mathbf{X}}$ tagged with the same index come from the same speaker.

Assume that there are n_m utterances tagged with m, and among these n_m utterances there are n_{mp} utterances from the p-th speaker. If we pick up one of the n_m utterances twice at random, with replacement, the probability that both of the chosen utterances come from the p-th speaker is $(n_{mp}/n_m) \times (n_{mp}/n_m)$. Thus, the probability that two utterances tagged with m come from the same speaker is $\Sigma_{p=1}^{P}(n_{mp}/n_m)^2$, which is also the cluster purity ρ_m defined in Eq. (1). Since the probability that one utterance tagged with m is n_m/N, we can estimate the probability that any two utterances $\mathbf{X}$ and $\dot{\mathbf{X}}$ tagged with the same index come from the same speaker by

$$\Pr(O(\dot{\mathbf{X}}) = O(\mathbf{X}) \mid C(\dot{\mathbf{X}}) = C(\mathbf{X})) = \sum_{m=1}^{M} \frac{n_m}{N}\left[\sum_{p=1}^{P}\left(\frac{n_{mp}}{n_m}\right)^2\right] = \bar{\rho}. \quad (13)$$

Approximating $\bar{\rho}$ as $\bar{\rho}(\mathbf{H})$ in Eq. (10), the likelihood $\Pr(\mathbf{O}|\Lambda)$ can be obtained with $\bar{\rho}(\mathbf{H})^N$. Accordingly, we can score a partitioning of N utterances having M clusters via

$$\mathrm{BIC}(M\text{-Clustering}) = N \log \bar{\rho}(\mathbf{H}) - 0.5\,\gamma\,M \log N. \quad (14)$$

The BIC value should increase with the increase of the M value in the beginning, but will decline significantly after an excess of clusters is created. A reasonable number of clusters can, thus, be determined by

$$M^* = \arg\max_{2 \le M \le N} \mathrm{BIC}(M\text{-Clustering}). \quad (15)$$

6. Experimental Results

Our speech data was chosen from the test set of the *2001 NIST Speaker Recognition Evaluation Corpus* [16]. It consisted of 197 utterances spoken by 15 male speakers, each of whom spoke 5 to 39 utterances. The speech features including 24 Mel-scale frequency cepstral coefficients were extracted from these data for every 20-ms Hamming-windowed frame with 10-ms frame shifts.

In computing the inter-utterance similarities, the numbers of mixtures in the GMMs and the eigenvectors were empirically determined to be 128 and 150, respectively. In the GA optimization, the empirical parameter values used for the maximum number of generations, the population size, the crossover probability, and the mutation probability were 2000, 5000, 0.5, and 0.1, respectively. For performance comparison, an agglomerative hierarchical clustering method was also implemented, in which the similarities between clusters were computed using the *complete linkage* of the inter-utterance similarities. Fig. 1. shows the speaker-clustering results as a function of the number of clusters. Here, "HC-GLR" and "HC-Eigenvoice" denote the agglomerative hierarchical clustering with the inter-utterance similarities computed using the *generalized likelihood ratio* [2,5,10] and our eigenvoice-motivated approach [12], respectively. "MPE-Eigenvoice" denotes the maximum purity clustering with the eigenvoice-based inter-utterance similarities. We can see that the proposed maximum purity clustering consistently yields better performance than the methods based on hierarchical clustering. When the number of clusters was specified as equal to the speaker population ($M = P = 15$), the best average cluster purity of 0.81 was achieved with MPE-Eigenvoice, which signifies a relative improvement of more than 10%, compared to 0.72, obtained with HC-Eigenvoice.

Next, to investigate the problem of speaker population estimation, the database was divided into several subsets involving speaker population sizes of 3, 6, and 9. A series of clustering experiments were performed on these subsets separately to examine if the optimal numbers of clusters

determined by using Eq. (14) could be around 3, 6, and 9, respectively. For each of the speaker population, we organized three subsets that were mutually distinct, as far as possible, in terms of speakers comprised. The penalty factor γ in Eq. (15) was empirically set to 1.4 throughout this experiment. Fig. 2 shows the resulting BIC values as a function of the number of clusters. The peak of each curve in the figure indicates the optimal number of clusters according to the BIC criterion. We can see that most of the peaks appeared near the actual number of speakers, and the BIC values declined significantly after an excess of clusters was created. This validates that the proposed method is capable of estimating the speaker population size.

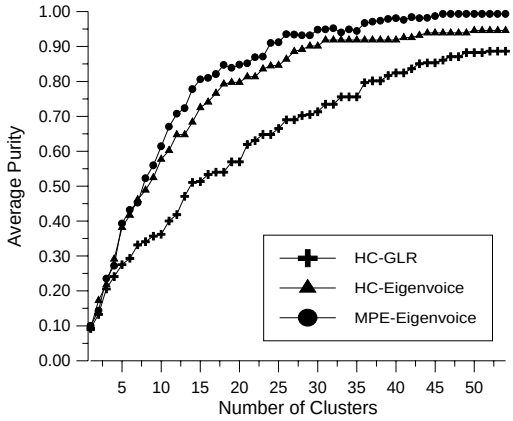
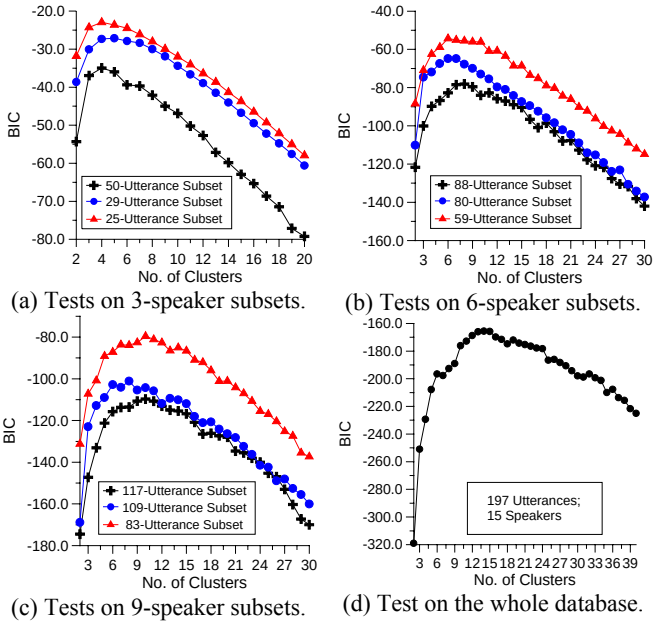


Figure 1: Performance of speaker clustering.



(a) Tests on 3-speaker subsets. (b) Tests on 6-speaker subsets.

(c) Tests on 9-speaker subsets. (d) Test on the whole database.

Figure 2: BIC values as a function of number of clusters.

7. Conclusions

This study has investigated the problem of how to gather speech utterances into clusters such that all the within-cluster utterances can be, as far as possible, from one single speaker. This requirement has been formulated as a problem of estimating and maximizing the overall cluster purity. By representing cluster purity as a function of inter-utterance similarity, and applying the genetic algorithm to find the solution of this function, we have demonstrated a noticeable improvement in the speaker-clustering performance, compared to the conventional agglomerative hierarchical approach. Furthermore, the clustering method has been incorporated with the Bayesian information criterion to determine how many clusters should be generated. Experimental results showed that the automatically-determined number of clusters can approximate the actual speaker population. With regard to practicability, our future work will extend the current speaker-clustering methods to deal with speech data containing multiple non-simultaneous or simultaneous speakers.

8. Acknowledgement

This work was supported in part by the National Science Council, Taiwan, under Grant NSC93-2213-E-001-017.

9. References

[1] Makhoul, J., Kubala, F., Leek, T., Liu, D., Nguyen, L., Schwartz, R., and Srivastava, A., "Speech and language technologies for audio indexing and retrieval", *PROCEEDINGS OF IEEE*, 88(8):1338- 1353, 2000.

[2] Gish, Herbert, Siu, M. H., and Rohlicek, R., "Segregation of speakers for speech recognition and speaker identification", *ICASSP'91*.

[3] Siegler, M. A., Jain, U., Raj, B., and Stern, R. M., "Automatic segmentation, classification and clustering of broadcast news audio", *DARPA Speech Recognition Workshop*, 1997.

[4] Jin, H., Kubala, F., and Schwartz, R., "Automatic speaker clustering", *DARPA Speech Recognition Workshop*, 1997.

[5] Solomonoff, A., Mielke, A., Schmidt, M., and Gish, H., "Clustering speakers by their voices", *ICASSP'98*.

[6] Chen, S. S. and Gopalakrishnan, P. S., "Clustering via the Bayesian information criterion with applications in speech recognition", *ICASSP'98*.

[7] Reynolds, D. A., Singer, E., Carson, B. A., O'Leary, G. C., McLaughlin, J. J., and Zissman, M. A., "Blind clustering of speech utterances based on speaker and language characteristics", *ICSLP'98*.

[8] Johnson, S. E., "Who spoke when? – Automatic segmentation and clustering for determining speaker turns", *Eurospeech'99*.

[9] Moh, Y., Nguyen, P., and Junqua, J. C., "Towards domain independent speaker clustering", *ICASSP'03*.

[10] Liu, D., and Kubala, F. "Online speaker clustering", *ICASSP'03*.

[11] Valente, F., and Wellekens, C., "Scoring unknown speaker clustering: VB vs. BIC", *ICSLP'04*.

[12] Tsai, W. H., Cheng, S. S., Chao, Y. H., Wang, H. M., "Clustering speech utterances by speaker using eigenvoice-motivated vector space model", *ICASSP'05*.

[13] Goldberg, D. E. *Genetic Algorithm in Search, Optimization and Machine Learning*. New York: Addison-Wesley, 1989.

[14] Reynolds, D. A., Quatieri, T. F., and Dunn, R. B., "Speaker verification using adapted Gaussian mixture models", *Digital Signal Processing*, 10:19-41, 2000.

[15] Schwarz, G. "Estimating the Dimension of a Model", *The Annals of Statistics*, 6:461–464, 1978.

[16] *http://www.nist.gov/speech/tests/index.htm*.