EIGEN-MLLRS APPLIED TO UNSUPERVISED SPEAKER ENROLLMENT FOR LARGE VOCABULARY CONTINUOUS SPEECH RECOGNITION

Xavier L. Aubert

Philips Research Laboratories
Weisshausstrasse 2, 52066 Aachen, Germany
xavier.aubert@philips.com

ABSTRACT

The concept of Eigen-MLLRs [1, 2], a variant of the Eigen-Voice method, is applied to unsupervised speaker enrollment in a large vocabulary CSR system. The emphasis is on fast adaptation. Two ways of estimating multiple Eigen-MLLR transformations are introduced, either joint or separated with respect to the Eigen-MLLR vector space. The first case allows multiple transforms to be robustly estimated from sparse data while the second achieves more accurate adaptation when more samples become available. The first decoded words spoken by a new test speaker are used to adapt the speaker-independent HMM means. The impact of this new enrollment algorithm is evaluated over a large real-life database dealing with professional medical transcriptions. Significant reductions of word-error-rates are achieved with less than 10 seconds of enrollment speech and without any supervision.

1. INTRODUCTION

An important question inherent to ASR adaptation techniques is how to structure the information contained in the training data, such that this *prior* knowledge can be efficiently exploited during decoding to cope with sources of variabilities like channel or environmental changes and outliers regarding voice, speaking mode or native and non-native accents.

In 1998, the Eigen-Voice method has been introduced for fast acoustic adaptation [3]. The key point lies in a specific organization of the acoustic-phonetic knowledge drawn from the training data such that a new speaker can be characterized with a small number of parameters in this "prior space". The Eigen-Voice technique is based on a principal component analysis (PCA) in the hyper-space of the HMM density means. This vector space is obtained by concatenating all acoustic model means into a mean "super-vector". An unknown speaker can be "projected" on the PCA orthonormal basis leading to speaker-dependent (SD) eigen coefficients. An adapted acoustic model is then easily obtained by linearly combining the eigen mean vectors with these SD weights. Though the dimension of the mean super-vector may reach a million or more in a large ASR system, a few tens of eigen coefficients might already provide an appropriate combination to cast a new speaker in the eigen mean space. The main drawback of this approach is its huge memory needs to store the mean super-vector basis as well as the significant computational complexity to get the matrix of the system determining the eigen-coefficients [4].

An alternative consists in applying the same PCA principle to the MLLR space rather than directly to the HMM means [1, 2]. MLLR [5] is a very popular and effective adaptation technique which applies affine transformations on *groups* of mean vectors tied in so-called regression classes. In the Eigen-MLLR approach, the vast number of mean vectors characterizing a speaker is replaced by a few matrices and offset vectors such that the dimension of the super-vector is reduced by more than a factor of 100 in a large system. Moreover, a clever implementation makes the estimation of Eigen-MLLR coefficients suitable for fast adaptation. Another interesting feature is that Eigen-MLLR fits very well with the speaker adaptive training (SAT) framework introduced in [7], where MLLR transformations are used to normalize the SD acoustic distributions of the training speakers towards a so-called "golden-speaker" or canonical model. It is known, however, that SAT models are not better if not subjected to careful adaptation when processing a new speaker [9]. Lastly, the Eigen-MLLR approach is closely related to the transform-based cluster adaptive training (CAT) of [10], both methods differing mainly in the initialisation stage. PCA yields a larger number of Eigen-MLLR modes compared to the number of clusters typically obtained with agglomerative techniques.

In the present work, Eigen-MLLRs are applied to unsupervised short-term adaptation towards a fast and robust enrollment of new speakers for use in large vocabulary systems. The emphasis is on using *multiple class* MLLRs which are known to yield improved adaptation compared to single MLLR, at least in supervised mode with sufficient enrollment data. The next section presents the main lines of the mean adaptation framework which has been adopted, followed by the description of our Eigen-MLLR algorithm. Two ways of dealing with multiple MLLRs are presented, either by joining or by keeping separated the individual transforms with respect to the Eigen-MLLR vector space. The last section is devoted to the experimental part, by first validating this new algorithm for supervised adaptation and next applying it to unsupervised speaker enrollment. The task concerns automatic transcriptions of medical reports dictated spontaneously over long-distance telephone lines from all over the US [11].

Based on the first decoded words, the Eigen-MLLR coefficients are estimated and used to adapt the original SI means to the current speaker. The whole process is unsupervised and automatic, introducing only a minor delay at the beginning. Compared to the word error rate (WER) of the baseline without enrollment, significant improvements are observed with less than 10 seconds of adaptation speech, averaged over 22 test speakers, all US natives but one. Individual gains range between 1% and 20% relative with respect to the baseline driven by *gender-dependent* models. Further additional gains are expected when combined with SAT and followed with long(er) term adaptation techniques.

2. MEAN ADAPTATION FRAMEWORK

Let d denote the dimension of the acoustic space, M the total number of densities in the acoustic mixture model and m the index of a density with $1 \leq m \leq M$. Let o_t be the observation vector at time t and F the total number of adaptation centi-second frames such that $t \in [1, F]$. Let r be the index of one among R regression classes where each MLLR affine transformation is a full $(d \times d)$ matrix $\mathbf{A}_r$ with an offset vector $\mathbf{b}_r$, $1 \leq r \leq R$. The m_th speaker-independent mean vector is written as μ_m^{si} and belongs to the regression class $r = r(m)$. With these notations, the mean "super-vector" of the Eigen-Voice method is a $M \cdot d$ vector $\upsilon = \{\mu_1^T, \mu_2^T, ...\mu_M^T\}^T$. The Eigen-MLLR super-vector is obtained by concatenating the offset-vector with the matrix *columns* giving a $d \cdot (d+1)$ vector $T_r = \{\mathbf{b}_r{}^T, \mathbf{A}_{r,1}{}^T, \mathbf{A}_{r,2}{}^T, ...\mathbf{A}_{r,d}{}^T\}^T$ for a single transformation. The MLLR super-vector for multiple transforms is obtained by concatenating the former T_r vectors into a $R \cdot d(d+1)$ vector $\tau = \{T_1^T, T_2^T, ...T_r^T, ...T_R^T\}^T$. The orthonormal eigen-vectors generated by PCA are written as υ_i for the Eigen-Voices and τ_i for the Eigen-MLLR basis with $1 \leq i \leq N$ where N, the eigen-space dimension, depends on the number S of training speakers. To select a part of a super-vector, τ_i^r will be used to denote the parameters of the r_th regression class in τ_i and similarly for $\upsilon^m = \mu_m$.

For the efficient implementation of the Eigen-MLLR algorithm, affine transformations must be expressed in terms of super-vectors T_r rather than with $\mathbf{A}_r$ matrices. By associating a rectangular $d \times d(d+1)$ matrix $\mathbf{D}(\mu)$ to the mean vector μ (see figure [1] at the paper end), the equivalent "vectorized" form is achieved :

$$\mathbf{A}_r * \mu + \mathbf{b}_r = \mathbf{D}(\mu) * T_r \tag{1}$$

$\mathbf{D}(\mu)$ is a diagonal-block *sparse* matrix whose non-zero elements are at (i, j) index pairs such that $j \bmod d = i$, the $\mathbf{D}_{i,j}$ value being the k_th component of μ with $k =$ trunc (j/d) and $\mu[0] = 1$.

In this framework, the enrollment step consists in estimating an unknown parameter vector α such that the cumulated distance between the observed vectors and the adapted means $\hat{\mu}_m(t)$ which depend on α, is minimized :

$$\underset{\alpha}{\mathbf{MIN}} \sum_t^F \sum_m^M \gamma_{m(t)} \; ||o_t - \hat{\mu}_{m(t)}(\alpha)||^2 \tag{2}$$

For simplicity, the variances have been assumed to be constant over the densities. Likewise, maximum likelihood (ML) estimations will be considered under Viterbi approximation assuming that $\gamma_m(t)$, the posterior probability of a density m at time t, is either 0 or 1 such that one *single* density $m(t)$ is associated to each observation o_t. Depending on the functional form of $\hat{\mu}_{m(t)}(\alpha)$, several cases can be considered and their differences made clear :

1. Standard MLLR: $\hat{\mu}_{m(t)} = \mathbf{A}_{r(m(t))} * \mu_{m(t)}^{si} + \mathbf{b}_{r(m(t))}$,
2. Eigen-Voices: $\hat{\mu}_{m(t)} = \sum_{i=1}^n \alpha_i \upsilon_i^{m(t)}$, with $n \leq N$
3. Eigen-MLLRs : $\hat{\mu}_{m(t)} = \mathbf{D}(\mu_{m(t)}^{si}) * \{\sum_{i=1}^n \alpha_i \tau_i^{r(m(t))}\}$

In the standard MLLR case, the unknown vector α has a number of $R \cdot d(d+1)$ free parameters as $\alpha = \{\mathbf{A}_r, \mathbf{b}_r;\ r = 1, ...R\}$ while in the two other cases α is just the vector of eigen-coefficients of dimension $n \leq N$. This paper deals with the last case only.

3. EIGEN-MLLR ALGORITHM

3.1. PCA in MLLR Vector-Space

For each training speaker $s, 1 \leq s \leq S$, a set of MLLR transformations is computed using standard ML estimation and processed into a super-vector $\tau(s)$ of dimension $R \cdot d(d+1)$. PCA is carried out on the set of S MLLR super-vectors using a Gram-Schmidt orthonormalization followed by an eigen value decomposition [4]. The Eigen-MLLR basis is provided by the eigen-vectors sorted on decreasing eigen values. Note that PCA is applied on "centered" vectors, the averaged MLLR super-vector $\overline{\tau}$ being subtracted prior to orthonormalization. The outcome is a sorted basis of *delta* MLLR eigen-vectors τ_i , $1 \leq i \leq N$ where $N = S - 1$ when no speaker-clustering is performed.

3.2. ML Estimation of Linear Decomposition Coefficients

Given a sequence of observations $o_1, ...o_t, ...o_F$ produced by an unknown speaker, the Eigen-MLLR coefficients α_i, $1 \leq i \leq n$ are obtained by solving a set of n *linear* equations derived from equation (2) coupled with the mean adaptation rule for Eigen-MLLRs, taking account of the "centering" step :

$$\hat{\mu}_{m(t)} = \mathbf{D}(\mu_{m(t)}^{si}) * \{\overline{\tau}^{r(m)} + \sum_{i=1}^n \alpha_i \tau_i^{r(m(t))}\}, \;\; 1 \leq t \leq F \tag{3}$$

The linear equations are obtained straightforwardly by inserting (3) into (2), expanding the L2 norm as a scalar product and setting the partial derivatives $\partial(.)/\partial\alpha_j$, $1 \leq j \leq n$ to zero. The estimated α_i coefficients serve to generate a set of affine transforms, by linearly combining the MLLR eigen-vectors τ_i, which are further used to adapt *all* SI mean vectors to the new speaker.

3.2.1. Joint Handling of Multiple Regression Classes

In the base Eigen-MLLR algorithm, multiple regression-classes are processed in one *single* super-vector, the individual transformations T_r being concatenated as explained in section 2. This defines the *joint* handling of multiple MLLRs which implicitly assumes intra-speaker correlation among respective transforms [6]. Note, that this is not the usual way of computing multiple MLLRs that are kept independent and separately estimated [5]. This leads to a single system of n linear equations in α_i for $i = 1, ...n$:

$$\sum_t^F \tilde{o}_t^T \; \mathbf{D}(\mu_{m(t)}^{si}) \; \tau_i^{r(m(t))} = \sum_t^F \sum_j^n \alpha_j \; [\tau_j^{r(m(t))}]^T \; \{\mathbf{D}(\mu_{m(t)}^{si})^T \mathbf{D}(\mu_{m(t)}^{si})\} \; \tau_i^{r(m(t))}, \tag{4}$$

where $\tilde{o}_t = o_t - \mathbf{D}(\mu_m^{si}) \, \overline{\tau}^{r(m(t))}$ to account for centering. Upon grouping the observations mapped to the same regression class r which is noted $o_t \to r(m(t))$, leads to the equivalent system (5) :

$$\sum_r^R X^T(r) \; \tau_i^{r(m(t))} = \sum_j^n \alpha_j \sum_r^R [\tau_j^{r(m(t))}]^T \; \mathbf{Z}(r) \; \tau_i^{r(m(t))},$$

where the auxiliary vectors $X(r)$ and matrices $\mathbf{Z}(r)$ are defined as:

$$X(r) = \sum_t^{o_t \to r(m(t))} [\tilde{o_t}^T \; \mathbf{D}(\mu_{m(t)}^{si})]^T, \;\; \text{of dimension} \;\; d(d+1),$$

$$\mathbf{Z}(r) = \sum_t^{o_t \to r(m(t))} \mathbf{D}(\mu_{m(t)}^{si})^T \mathbf{D}(\mu_{m(t)}^{si}) \;\; \text{of order} \; d(d+1)$$

The solution of equation (5) is given by $\alpha = \mathbf{B}^{-1} * A$ where

$$\mathbf{B}_{i,j} = \sum_{r}^{R} [\tau_j^{r(m(t))}]^T \mathbf{Z}(r) \tau_i^{r(m(t))}, \quad A_j = \sum_{r}^{R} X^T(r) \tau_j^{r(m(t))}$$

It shows that the coefficients of the linear system for α are given as a *sum* over the contributions of the respective regression classes when using multiple MLLRs *jointly*, and the solution is achieved with one single matrix inversion of order n, independently of R.

3.2.2. Separated Handling of Multiple Regression Classes

For handling multiple MLLR transforms *separately*, the PCA preprocessing has first to be run R times on single-MLLR supervectors of dimension $d(d+1)$, thus providing a specific Eigen-MLLR orthonormal basis for each regression class in its respective vector space. Next, a linear system has to be set up for each parameter vector $\alpha^r, 1 \leq r \leq R$ but this is straigthforward based on the previous equations. Indeed, each regression-class transformation is obtained with $\alpha^r = (\mathbf{B^r})^{-1} * A^r$ where the system coefficients are simply given by

$$\mathbf{B^r}_{i,j} = [\tau_j^{r(m(t))}]^T \mathbf{Z}(r)\, \tau_i^{r(m(t))}, \quad A_j^r = X^T(r)\, \tau_j^{r(m(t))}$$

Apart from having to invert R matrices of order n, the main computing costs implied by the calculations of the system coefficients are identical to the case of handling multiple MLLRs jointly. In the present case, the regression-class "counts" are just kept separated and fed into specific linear systems. This offers the possibility of using a different number of eigen-modes n for each regression-class, for example, by making $n(r)$ dependent on the number of observations collected for that class. Likewise, this suggests possible interpolation schemes if some regression-class gets (too) sparse data.

4. EXPERIMENTAL RESULTS

4.1. System and Task Description

The baseline is a recent upgrade of the Philips Research LVCSR system for the transcription of spontaneous speech tasks like medical reporting, where filled pauses are *explicitly* taken into account at all modeling levels [11]. The baseline main components are as follows:

- The signal front-end relies on MFCC vectors augmented with voicing features and subjected to LDA [12].
- Continuous mixtures of Laplacian densities are estimated gender dependently with ML Viterbi training.
- Multiple pronunciations representing one fifth of all lexical entries are weighted by their unigram priors.
- Filler specific phones are introduced together with longer than average minimum filler phone durations.
- Decoding proceeds from left to right using a prefix-tree lexicon and contributions of simultaneously active pronunciations of the same word are summed up [11].

All results reported here are produced with a single decoding pass and have been carried out using an inhouse data collection of real-life recordings of medical reports. The acoustic training corpus consists of about 160h of data (375 speakers, 1.4M words), where filled pauses and non-speech events are annotated. The development corpus (DEV set) consists of 11 speakers, 38k spoken words and the evaluation corpus (EVAL set) of 11 speakers, 27k words.

4.2. Validation for Supervised Adaptation

The Eigen-MLLR method described in the previous sections has been first applied to *supervised* speaker adaptation and compared with standard MLLR estimations for validating the new algorithm and gaining more insight into its behavior. A subset of "difficult" speakers has been considered for which about 5 minutes of *transcribed* data are available in addition to another 15 minutes for testing. A first point concerns the "optimal" number of Eigen-MLLR modes leading to the best adaptation improvements. A second point concerns the use of multiple MLLRs which are based on a tree organization of regression classes [5, 9]. The tree leaves are phone subsets defined on broad phonetic articulatory features with a maximum of 25 regression classes. For standard MLLR estimation, the number of "active" regression classes is controlled by a threshold of minimum observations and in the present setup 12 distinct MLLR classes are typically estimated. Concerning the Eigen-MLLR estimation, the PCA space is based on these 12 broad-phonetic regression classes implying a dimension of 15120 for the joint super-vector. When keeping the multiple Eigen-MLLRs separate, the number of eigen coefficients depends on the number of observations assigned to each class. The table below summarizes these supervised adaptation experiments.

CASE	ML-Score	#Params	WER%	Rel.Improv.
NO ADAPT	-163.8	0	35.3%	Baseline
STD-01	-160.0	1260	29.9%	15.3%
STD-12	-157.3	≈15000	27.7%	21.5%
EIG-01	-161.7	50	30.9%	12.5%
EIG-01	-160.3	300	29.4%	16.7%
EIG-12 Joint	-159.8	350	28.3%	19.8%
EIG-12 Sep.	-157.5	≈2500	27.2%	23.0%

Table 1. *Supervised speaker-adaptation tests*

The second column gives the scaled log-likelihood score values achieved on the adaptation data while the third column contains the number of adaptation parameters that are numerically estimated on these data. As shown in the table, for a single MLLR transform (STD-01 and EIG-01 cases), the number of eigen-coefficients has to be rather high, i.e. close to the number of PCA dimensions unless the results are sub-optimal[1]. When 300 eigen-modes are taken (or more) both methods achieve similar results, the Eigen-MLLR reaching a slightly lower error rate which indicates that the priors are well exploited. For the multiple MLLR case, however, the *joint* eigen estimation (case EIG-12 Joint) is unable to get the improvements obtained by a standard ML estimation of 12 affine transformations. This is, to a large extent, due to the small number of free parameters of the joint Eigen-MLLRs, which is not rewarding when dealing with several minutes of enrollment data. When estimating the Eigen-MLLR coefficients *separately* for each class (case EIG-12 Sep.), the standard multi-MLLR case is outperformed with about one sixth of the parameter number. This is yet another illustration of the tradeoff between trainability and specificity of statistical models: by joining multiple MLLRs in one super-vector, piecewise linear transformations of means can be estimated robustly from very little data while improved adaptation is

[1] Using only a few tens of eigen-modes as seen in some publications might occasionally be too small for optimal results as pointed out in [13].

achieved when more parameters can be estimated from more data.

4.3. Unsupervised Speaker Enrollment

For speaker enrollment, a critical parameter concerns the amount of adaptation data. Four durations have been considered, namely 8, 15, 30 and 60 seconds taken from the first test utterance(s). However, due to pause intervals at sentence start and the delay introduced by partial traceback, the true amount of speech samples is about half of the signal duration. Thus the effective enrollment times averaged over the test speakers are around 4.0, 7.5, 18 and 38 seconds of spoken words. When less than 5 seconds of speech are available, the Eigen-MLLR algorithm falls back to a single transformation while with more data, multiple MLLRs are estimated jointly with an increasing number of eigen-modes between 150 and 250. No further adaptation technique is applied whatsoever after the single enrollment step has been completed, to focus on the impact of the current technique. The table below summarizes the enrollment results in terms of word error-rate (%) for the DEV and EVAL sets, each one involving 11 males speakers.

SET	Enroll Time:	0.0 Sec	4 Sec	7.5 Sec	18 Sec	38 Sec
DEV	Word-Error:	20.1%	19.55%	19.35%	18.95%	18.8%
	Rel.Improv:	-	2.6%	3.7%	5.7%	6.5%
EVAL	Word-Error:	26.7%	25.7%	25.2%	24.85%	24.65%
	Rel.Improv:	-	3.8%	5.7%	6.9%	7.5%

Table 2. *Unsupervised speaker-enrollment (R=1 or 6, no SAT)*

Generally speaking, all speakers are improved, the individual relative gains ranging from 1% to 20%. The EVAL set contains a few "outlier" speakers including one non-native which explains the higher baseline error-rate and also the larger gains especially with just 4 or 7.5 seconds of enrollment data. Even with about 4 seconds of speech, all EVAL speakers are improved but one and the average error reduction is of 3.8% relative. When the enrollment time is increased beyond 20 seconds, the additional improvements appear small which might indicate that the current control strategy of the type and number of eigen-modes is not suited for these operating conditions.

A number of contrast experiments have been run using standard MLLR for *unsupervised* enrollment with a single transformation. A special block-matrix structure is applied for improved robustness. With less than 5 seconds enrollment, the latter method appears unreliable, several speakers being actually degraded. With 7.5 seconds of speech, the improvements observed on the same data under identical conditions are of 1.5% relative only, a factor of 3 smaller than the average Eigen-MLLR gains. When longer enrollment times are considered, the differences between standard and eigen-MLLRs get smaller but remain in favor of the latter. The eigen-MLLR approach is definitely superior as far as adaptation *speed* is concerned.

5. SUMMARY

In this paper, the concept of Eigen-MLLRs has been applied to unsupervised speaker enrollment for use in large vocabulary tasks. Two ways of dealing with multiple MLLR transforms have been presented, their respective merits depending on the amount of adaptation data. This new algorithm has been validated with supervised adaptation experiments. Concerning unsupervised speaker enrollment, significant gains have been achieved with less than 10 seconds of adaptation speech. Contrast runs with standard MLLR show that the present algorithm is significantly better for fast enrollment. Larger gains are expected when applied to clear outliers like non-native with strong accents, which was not the case in the current testing conditions.

On the other hand, when more enrollment data is considered, the benefit of the present method has not been thoroughly evaluated yet. Prior information is known to be most useful when little data from the speaker at hand is available which suggests that Eigen-MLLRs would be profitably combined with longer term adaptation. This technique is currently being evaluated with SAT models for which the MLLR-PCA approach appears to be a promising complement.

Acknowledgements : I would like to thank my colleagues Henrik Botterweck and Christoph Neukirchen for helpful discussions.

6. REFERENCES

[1] Kuan-ting Chen, Wen-wei Liau, Hsin-min Wang and Lin-shan Lee, "Fast Speaker Adaptation using Eigenspace-Based Maximum Likelihood Linear Regression", in Proc. ICSLP 2000, pp. 742-745, Bejing, China, October 2000.

[2] Nick J.-C. Wang, Sammy S.-M. Lee, Frank Seide and Lin-Shan Lee, "Rapid Speaker Adaptation using a priori knowledge by eigen-space analysis of MLLR parameters", Proc. ICASSP 2001, pp. 345-349, Salt-Lake-City, US, May 2001.

[3] R. Kuhn, P. Nguyen, J.-C. Junqua, L. Goldwasser, N. Niedzielski, S. Fincke, K.Field, M. Contolini, "Eigenvoices for Speaker Adaptation" in Proc. ICSLP 1998, pp. 1771-1774, Sydney, Australia, November 1998.

[4] Henrik Botterweck, "Very Fast Adaptation for Large Vocabulary Continuous Speech Recognition using Eigenvoices", in Proc. ICSLP 2000, IV, pp 354-357, Bejing, China, Oct'2000.

[5] C.J. Leggetter, P.C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models", Computer Speech and Language, Vol. 9 (2), pp. 171-185, 1995.

[6] Sam-Joo Doh and Richard M. Stern, "Inter-Class MLLR for Speaker Adaptation", in Proc. ICASSP 2000 pp. 1543-1546, Istanbul, Turkey, May 2000.

[7] Tasos Anastasakos, John McDonough, Richard Schwartz, John Makhoul, "A Compact Model for Speaker-Adaptive Training", Proc. ICSLP, pp. 1137-1140, Philadelphia, USA, October 1996.

[8] Spyros Matsoukas, Rich Schwartz, Hubert Jin, Long Nguyen, "Practical Implementations of Speaker-Adaptive Training", DARPA Speech Recognition Workshop, Chantilly, Virginia, USA, February 1997.

[9] Xavier L. Aubert and Eric Thelen, "Speaker Adaptive Training Applied to Continuous Mixture Density Modeling", in Proc. Eurospeech 1997, pp. 1851-1854, Rhodos, Greece, September 1997.

[10] Mark J. F. Gales, "Cluster Adaptive Training of Hidden Markov Models", Technical Report, Cambridge University Engineering Department, Cambridge, England, August 1999.

[11] H. Schramm, X. L. Aubert, C. Meyer, J. Peters, "Filled-Pause Modeling for Medical Transcriptions" in SSPR Workshop, Tokyo, April 2003.

[12] X. L. Aubert, "Voicing Features in MFCC-LDA framework applied to large vocabulary automatic transcriptions of Medical Reports", Philips Research Technical Report, April 2003.

[13] P. Kenny, M. Mihoubi and P. Dumouchel, "New MAP Estimators for Speaker Recognition", in Proc. Eurospeech 2003, pp. 2961-2964, Geneva, Switzerland, September 2003.

Figure 1: *Structure of the sparse 3x12 matrix $D(\mu)$ for d=3*

1 0 0	$\mu[1]$ 0 0	$\mu[2]$ 0 0	$\mu[3]$ 0 0
0 1 0	0 $\mu[1]$ 0	0 $\mu[2]$ 0	0 $\mu[3]$ 0
0 0 1	0 0 $\mu[1]$	0 0 $\mu[2]$	0 0 $\mu[3]$