Time-Dependent Cross-Probability Model for Multi-Environment Model Based LInear Normalization

Luis Buera, Eduardo Lleida, Juan A. Nolazco-Flores, Antonio Miguel, Alfonso Ortega*

Communication Technologies Group (GTC)
I3A, University of Zaragoza, Spain
{amiguel,lleida,lbuera,ortega}@unizar.es

* Computer Science Department, ITESM,
Campus Monterrey, Monterrey, Mexico
jnolazco@itesm.mx

Abstract

In a previous work, Multi-Environment Model based LInear Normalization, MEMLIN, was presented and it was proved to be effective to compensate environment mismatch. MEMLIN is an empirical feature vector normalization which models clean and noisy spaces by Gaussian Mixture Models (GMMs). In this algorithm, the probability of the clean model Gaussian, given the noisy model one and the noisy feature vector (cross-probability model) is a critical point. In the previous work the cross-model probability was approximated as time-independent. In this paper, a time-dependent estimation of the cross-probability model based on GMM is proposed. Some experiments with SpeechDat Car database were carried out in order to study the performance of the proposed estimation in a real acoustic environment. MEMLIN with time-independent cross-probability model reached 70.21% of mean improvement in Word Error Rate (WER), however, when time-dependent cross-probability model based on GMM was applied, the mean improvement in WER went up to 78.47%.

Index Terms: robust speech recognition, feature normalization.

1. Introduction

When training and testing acoustic conditions differ, the accuracy of speech recognition systems rapidly degrades. To compensate for this mismatch, robustness techniques have been developed along the following two main lines of research: acoustic model adaptation methods, and feature vector normalization methods. In general, acoustic model adaptation methods produce the best results [1] because they can model the uncertainty caused by the noise statistics. However, these methods require more data and computing time than do feature vector normalization methods, which do not produce as good results but provide more on line solutions. Hybrid techniques also exist [2].

There are several feature vector normalization families [3], but independently of the family, some algorithms assume a prior probability density function (pdf) for the estimation variable. In those cases, a Bayesian estimator can be used to estimate the clean feature vector. The most commonly used criterion is to minimize the Mean Square Error (MSE), and the optimal estimator for this criterion, Minimum Mean Square Error (MMSE), is the mean of the posterior pdf. Methods, such as Stereo-based Piecewise Linear Compensation for Environments (SPLICE) [4], or Multi-Environment Model-based LInear Normalization (MEMLIN) [5] use the MMSE estimator to compute the estimated clean feature vector.

This work has been supported by the national project TIN 2005-08660-C04-01

A previous work [5] shows that MEMLIN is effective to compensate the effects of dynamic and adverse car conditions. MEMLIN is an empirical feature vector normalization based on stereo data and the MMSE estimator, with joint modelling of clean and noisy spaces by Gaussian Mixture Models (GMMs). Therefore, a bias vector transformation is associated with each pair of Gaussians from the clean and the noisy spaces. A critical point in MEMLIN is the estimation of the probability of the clean model Gaussian, given the noisy model one and the noisy feature vector (cross-probability model). In [5], a time-independent solution is considered. This work focuses on this term and it is proposed a time-dependent solution, modelling the noisy feature vectors associated to each pair of Gaussians from the clean and the noisy spaces with a GMM.

This paper is organized as follows: In Section 2, an overview of MEMLIN is detailed. In Section 3, some experiments are presented to show the importance of the cross-probability model estimation. The new proposed cross-probability model based on GMM is explained in Section 4. The results with Spanish SpeechDat Car database [6] are included in Section 5. Finally, the conclusions are presented in Section 6.

2. MEMLIN overview

2.1. MEMLIN approximations

• Clean feature vectors, $\mathbf{x}_t$, are modelled using a GMM of C components

$$p(\mathbf{x}_t) = \sum_{s_x=1}^{C} p(\mathbf{x}_t|s_x)p(s_x), \tag{1}$$

$$p(\mathbf{x}_t|s_x) = N(\mathbf{x}_t; \mu_{s_x}, \mathbf{\Sigma}_{s_x}), \tag{2}$$

where μ_{s_x}, $\mathbf{\Sigma}_{s_x}$, and $p(s_x)$ are the mean vector, the diagonal covariance matrix, and the a priori probability associated with the clean model Gaussian s_x.

• Noisy space is split into several basic environments, e, and the noisy feature vectors, $\mathbf{y}_t$, are modeled as a GMM of C' components for each basic environment

$$p_e(\mathbf{y}_t) = \sum_{s_y^e=1}^{C'} p(\mathbf{y}_t|s_y^e)p(s_y^e), \tag{3}$$

$$p(\mathbf{y}_t|s_y^e) = N(\mathbf{y}_t; \mu_{s_y^e}, \mathbf{\Sigma}_{s_y^e}), \tag{4}$$

where s_y^e denotes the corresponding Gaussian of the noisy model for the e basic environment; $\mu_{s_y^e}$, $\mathbf{\Sigma}_{s_y^e}$, and $p(s_y^e)$ are the mean vector, the diagonal covariance matrix, and the a priori probability associated with s_y^e.

• Clean feature vectors can be approximated as a linear function of the noisy feature vector, which depends on the basic environment and the clean and noisy model Gaussians: $\mathbf{x} \approx \Psi(\mathbf{y}_t, s_x, s_y^e) = \mathbf{y}_t - \mathbf{r}_{s_x,s_y^e}$, where $\mathbf{r}_{s_x,s_y^e}$ is a bias vector transformation between noisy and clean feature vectors for each pair of Gaussians, s_x and s_y^e.

2.2. MEMLIN enhancement

With those approximations, MEMLIN transforms the MMSE estimation expression, $\hat{\mathbf{x}}_t = E[\mathbf{x}|\mathbf{y}_t]$, into

$$\hat{\mathbf{x}}_t = \mathbf{y}_t - \sum_e \sum_{s_y^e} \sum_{s_x} \mathbf{r}_{s_x,s_y^e} p(e|\mathbf{y}_t) p(s_y^e|\mathbf{y}_t, e) p(s_x|\mathbf{y}_t, e, s_y^e), \tag{5}$$

where $p(e|\mathbf{y}_t)$ is the a posteriori probability of the basic environment; $p(s_y^e|\mathbf{y}_t, e)$ is the a posteriori probability of the noisy model Gaussian, s_y^e, given the feature vector, $\mathbf{y}_t$, and the basic environment, e. To estimate those terms, $p(e|\mathbf{y}_t)$ and $p(s_y^e|\mathbf{y}_t, e)$, equations (3) and (4) are applied as described in [5]. Finally, the cross-probability model, $p(s_x|\mathbf{y}_t, e, s_y^e)$, is the probability of the clean model Gaussian, s_x, given the feature vector, $\mathbf{y}_t$, the basic environment, e, and the noisy model Gaussian, s_y^e. The cross-probability model, along with the bias vector transformation, $\mathbf{r}_{s_x,s_y^e}$, is estimated in a training phase using stereo data, and avoiding the time dependence given by the noisy feature vector. So, $p(s_x|\mathbf{y}_t, e, s_y^e)$ can be estimated by relative frequency (time-independent cross-probability model) [5]

$$p(s_x|\mathbf{y}_t, e, s_y^e) \simeq p(s_x|s_y^e, e) = \frac{C_N(s_x|s_y^e)}{N_{s_y^e}}, \tag{6}$$

where $C_N(s_x|s_y^e)$ is the count number of times that the most probable pair of Gaussians is s_x and s_y^e for all pairs of stereo training data of the e basic environment, and $N_{s_y^e}$ is the count number of times that the most probable Gaussian for noisy training feature vectors is s_y^e for the e basic environment.

3. Cross-probability model performance

To study the performance of the cross-probability model in a qualitative way, the histograms and scattegrams between the first Mel Frequency Cepstral Coefficients (MFCCs) in non-silence frames for different signals are depicted in Fig. 1.

Figure 1.a, which represents clean and noisy in real car conditions feature vectors, shows the effects of car noise. The pdf of clean first MFCCs is clearly affected (Fig.1.a.1), and the uncertainty is increased (Fig.1.a.2).

Figure 1.b represents clean and normalized with MEMLIN feature vectors. MEMLIN is applied with 128 Gaussians. The pdf of normalized first MFCCs has been approximated to the clean signal one (Fig. 1.b.1), and the uncertainty has been reduced (Fig. 1.b.2). The peak that appears in Fig. 1.b.1 is due to the transformation of noisy feature vectors towards the clean silence.

Finally, Fig. 1.c represents clean and normalized with MEMLIN feature vectors where the cross-probability model is computed with the corresponding clean feature vector as (7). MEMLIN is applied with 128 Gaussians. In this case the pdf of the normalized signal is almost the same that the clean one (Fig. 1.c.1) and the uncertainty is drastically reduced (Fig. 1.c.2). These results verify the importance of the estimation of the cross-probability model in MEMLIN algorithm.

$$p(s_x|\mathbf{y}_t, e, s_y^e) \simeq \frac{p(s_x)p(\mathbf{x}_t|s_x)}{\sum_{s_x} p(s_x)p(\mathbf{x}_t|s_x)}. \tag{7}$$

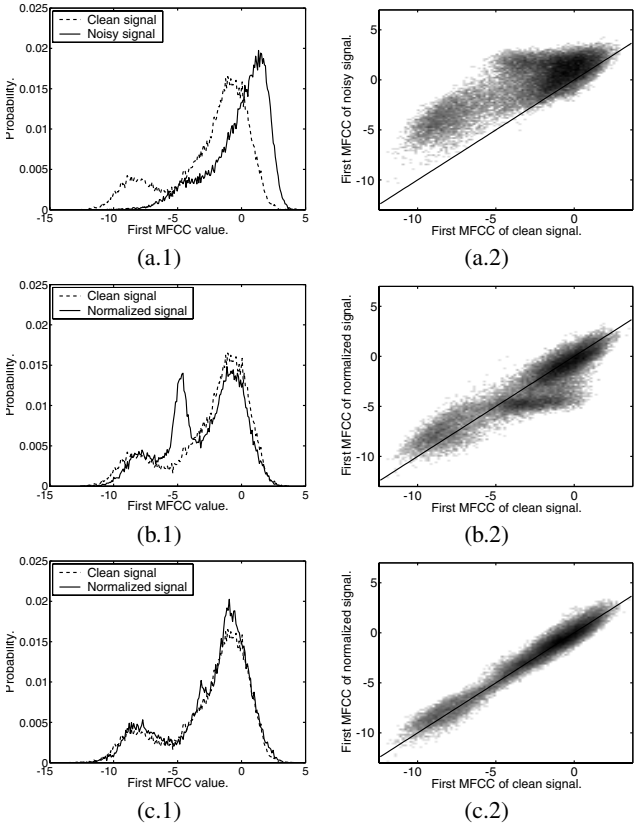

Figure 1: Scattegrams and histograms between the first MFCC in non-silence frames for different signals. The line in the scattergrams represents the function $x = y$.

4. Cross-probability model based on GMM

To improve the time-independent cross-probability model (6), we propose to model the noisy feature vectors associated to a pair of Gaussians (s_x and s_y) with a GMM of C'' components

$$p(\mathbf{y}_t|s_x, s_y) = \sum_{s'_y=1}^{C''} p(\mathbf{y}_t|s_x, s_y, s'_y) p(s'_y|s_x, s_y), \tag{8}$$

$$p(\mathbf{y}_t|s_x, s_y, s'_y) = N(\mathbf{y}_t; \mu_{s_x,s_y,s'_y}, \mathbf{\Sigma}_{s_x,s_y,s'_y}), \tag{9}$$

where μ_{s_x,s_y,s'_y}, $\mathbf{\Sigma}_{s_x,s_y,s'_y}$, and $p(s'_y|s_x, s_y)$ are the mean, the diagonal covariance matrix, and the a priori probability associated with s'_y Gaussian of the cross-probability GMM associated with s_x and s_y. To train these three parameters, the EM algorithm [7] is applied. The basic environments are not indexed for clarity: they are considered independently.

Let a set of clean and noisy stereo data available to learn the corresponding cross-probability GMM parameters $(\mathbf{X}, \mathbf{Y}) = \{(\mathbf{x}_1, \mathbf{y}_1), ...(\mathbf{x}_n, \mathbf{y}_n)..., (\mathbf{x}_N, \mathbf{y}_N)\}$. Each $\mathbf{y}_n$ can be seen as an incomplete component-labelled frame, which is completed by two indicator vectors. The first one is $\mathbf{w}_n \in \{0,1\}^{C'}$, with 1 in the position corresponding to the s_y Gaussian generating $\mathbf{y}_n$ and zeros elsewhere ($\mathbf{W} = \{\mathbf{w}_1, ..., \mathbf{w}_N\}$). The second indicator vector is $\mathbf{z_n} \in \{0,1\}^{C''}$, with 1 in the position corresponding to the s'_y Gaussian of the cross-probability GMM generating $\mathbf{y}_n$ and zeros elsewhere ($\mathbf{Z} = \{\mathbf{z}_1, ..., \mathbf{z}_N\}$). Each $\mathbf{x}_n$ can be seen also as

Table 1: WER baseline results, in %, from the different basic environments (E1,..., E7).

Train	Test	E1	E2	E3	E4	E5	E6	E7	MWER (%)
CLK	CLK	1.90	2.64	1.81	1.75	1.62	0.64	0.35	1.75
CLK	HF	5.91	14.49	14.55	20.17	21.07	16.19	35.71	16.21
HF	HF	6.67	14.24	12.73	12.91	14.97	9.68	8.50	11.81
†HF	HF	2.86	7.12	4.34	4.39	7.63	4.60	4.76	5.30

an incomplete component-labelled frame, which is completed by one indicator vector: $\mathbf{v}_n \in \{0,1\}^C$, with 1 in the position corresponding to the s_x Gaussian generating $\mathbf{x}_n$ and zeros elsewhere ($\mathbf{V} = \{\mathbf{v}_1, ..., \mathbf{v}_N\}$). The indicator vectors are called missing data, too. So, the complete data pdf is

$$p(\mathbf{x},\mathbf{y},\mathbf{v},\mathbf{w},\mathbf{z}) \simeq p(\mathbf{v},\mathbf{w})p(\mathbf{x}|\mathbf{v},\mathbf{w}) \times p(\mathbf{v},\mathbf{w},\mathbf{z})p(\mathbf{y}|\mathbf{v},\mathbf{w},\mathbf{z}), \quad (10)$$

where it is assumed that $\mathbf{x}$ is independent of $\mathbf{y}$ and $\mathbf{z}$. Since the indicator vectors are Multinomial, the complete data pdf can be expressed as (11), where v_{s_x}, w_{s_y} and $z_{s'_y}$ are the components of $\mathbf{v}$, $\mathbf{x}$ and $\mathbf{z}$ associated to the Gaussians s_x, s_y and s'_y, respectively.

The EM algorithm is applied iteratively in two steps. The Expectation (E) step, which estimates the expected values of the missing data, and the Maximization (M) step, which obtains the parameters of the cross-probability GMM using the estimated missing data.

4.1. The E step

To evaluate the E step, the function $Q(\mathbf{\Theta}|\mathbf{\Theta}^{(k)})$ is defined as $Q(\mathbf{\Theta}|\mathbf{\Theta}^{(k)}) = E[log(p(\mathbf{X},\mathbf{Y},\mathbf{V},\mathbf{W},\mathbf{Z}|\mathbf{\Theta}))|\mathbf{X},\mathbf{Y},\mathbf{\Theta}^{(k)}]$, where $E[\bullet]$ is the expected value, k is the iteration index and $\mathbf{\Theta}$ includes the unknown parameters of the cross-probability GMM. It is expressed as (12), where

$$(v_{s_x} w_{s_y})^{(k)} \simeq E[v_{s_x}|\mathbf{x}_n] E[w_{s_y}|\mathbf{y}_n], \quad (13)$$

$$(v_{s_x} w_{s_y} z_{s'_y})^{(k)} \simeq (v_{s_x} w_{s_y})^{(k)} E[z_{s'_y}|\mathbf{y}_n, v_{s_x}, w_{s_y}, \mathbf{\Theta}^{(k)}], \quad (14)$$

where it is assumed that v_{s_x} and w_{s_y} are independent, $E[v_{s_x}|\mathbf{x}_n, \mathbf{y}_n, \mathbf{\Theta}^{(k)}] \simeq E[v_{s_x}|\mathbf{x}_n]$ and $E[w_{s_y}|\mathbf{x}_n, \mathbf{y}_n, \mathbf{\Theta}^{(k)}] \simeq E[w_{s_y}|\mathbf{y}_n]$. $E[z_{s'_y}|\mathbf{y}_n, v_{s_x}, w_{s_y}, \mathbf{\Theta}^{(k)}]$ is estimated with (8) and (9) as (15), and $E[v_{s_x}|\mathbf{x}_n]$ and $E[w_{s_y}|\mathbf{y}_n]$ are computed in a similar way with (1) and (2), and with (3) and (4), respectively. Although, in this work, to simplify, $E[v_{s_x}|\mathbf{x}_n]$ and $E[w_{s_y}|\mathbf{y}_n]$ values are 1, if the corresponding Gaussians are the most probable ones, and 0 in any other case (hard Gaussian estimation approach).

4.2. The M step

To obtain the maximum likelihood estimates for the parameters of the cross-probability GMM, $Q(\mathbf{\Theta}|\mathbf{\Theta}^{(k)})$ is maximized with respect to them. So, the corresponding expressions for the $(k+1)$th iteration are

$$p(s'_y|s_x, s_y)^{(k+1)} = \frac{\sum_n (v_{s_x} w_{s_y} z_{s'_y})^{(k)}}{\sum_n \sum_{s'_y} (v_{s_x} w_{s_y} z_{s'_y})^{(k)}}. \quad (16)$$

$$\mu^{(k+1)}_{s_x,s_y,s'_y} = \frac{\sum_n (v_{s_x} w_{s_y} z_{s'_y})^{(k)} \mathbf{y}_n}{\sum_n (v_{s_x} w_{s_y} z_{s'_y})^{(k)}}. \quad (17)$$

$$\mathbf{\Sigma}^{(k+1)}_{s_x,s_y,s'_y} = \frac{1}{\sum_n (v_{s_x} w_{s_y} z_{s'_y})^{(k)}} \times \sum_n (v_{s_x} w_{s_y} z_{s'_y})^{(k)} (\mathbf{y}_n - \mu^{(k)}_{s_x,s_y,s'_y})(\mathbf{y}_n - \mu^{(k)}_{s_x,s_y,s'_y})^t. \quad (18)$$

Once the cross-probability GMM parameters are estimated for each basic environment, $p(s_x|\mathbf{y}_t, e, s^e_y)$ can be obtained with (8) as (19). Note that the time-independent assumption has been avoided.

$$p(s_x|\mathbf{y}_t, e, s^e_y) = \frac{p(\mathbf{y}_t|s_x, s^e_y)}{\sum_{s_x} p(\mathbf{y}_t|s_x, s^e_y)}. \quad (19)$$

Observe that if the hard Gaussian estimation approach is considered and the noisy feature vectors are modelled in (8) with the same uniform pdf for all the pairs of Gaussians (s_x and s^e_y), instead of a GMM for each one, the cross-probability model is (6).

5. Results

To observe the performance of the cross-probability GMM proposed in a real, dynamic, and complex environment, a set of experiments were carried out using the Spanish SpeechDat Car database [6]. Seven basic environments were defined: car stopped, motor running (E1), town traffic, windows close and climatizer off (silent conditions) (E2), town traffic and noisy conditions: windows open and/or climatizer on (E3), low speed, rough road, and silent conditions (E4), low speed, rough road, and noisy conditions (E5), high speed, good road, and silent conditions (E6), and high speed, good road, and noisy conditions (E7).

The clean signals are recorded with a CLose talK (CLK) microphone (Shune SM-10A), and the noisy ones are recorded by a Hands-Free (HF) microphone placed on the ceiling in front of the driver (Peiker ME15/V520-1). The SNR range for CLK signals goes from 20 to 30 dB, and for HF ones goes from 5 to 20 dB.

For speech recognition, the feature vectors are composed of the 12 MFCCs, first and second derivatives and the delta energy, giving a final feature vector of 37 coefficients computed every 10 ms using a 25 ms Hamming window. On the other hand, in this work, the feature vector normalization methods are applied only to the 12 MFCCs and energy, whereas the derivatives are computed over the normalized static coefficients

The recognition task is isolated and continuous digits recognition. Three-state 16 Gaussian continuous density HMM to model the 25 Spanish phonemes and 2 silence models for long and inter-word silences are used in this task.

The Word Error Rate (WER) baseline results for each basic environment are presented in Table 1, where MWER is the Mean WER computed proportionally to the number of utterances in each basic environment. Cepstral mean normalization is applied to testing and training data. "Train" column refers to the signals used to obtain the corresponding acoustic HMMs: CLK if they are trained with all clean training utterances, and HF and if they are trained with all noisy ones. HF† indicates that specific acoustic HMMs for each basic environment are applied in the recognition task (environment match condition). "Test" column indicates which signals are used for recognition: clean, CLK, or noisy, HF.

Table 1 shows the effect of real car conditions, which increases the WER in all of the basic environments, (Train CLK, Test HF), concerning the rates for clean conditions, (Train CLK, Test CLK).

$$p(\mathbf{x},\mathbf{y},\mathbf{v},\mathbf{w},\mathbf{z}) \simeq \prod_{s_x}\prod_{s_y}[p(v_{s_x}=1,w_{s_y}=1)p(\mathbf{x}|v_{s_x}=1,w_{s_y}=1)]^{v_{s_x}w_{s_y}} \times \prod_{s_x}\prod_{s_y}\prod_{s'_y}[p(v_{s_x}=1,w_{s_y}=1,z_{s'_y}=1)p(\mathbf{y}|v_{s_x}=1,w_{s_y}=1,z_{s'_y}=1)]^{v_{s_x}w_{s_y}z_{s'_y}}. \quad (11)$$

$$Q(\mathbf{\Theta}|\mathbf{\Theta}^{(k)}) = \sum_n \sum_{s_x}\sum_{s_y}(v_{s_x}w_{s_y})^{(k)}[log(p(s_x)p(s_y)) + log(p(\mathbf{x}_n|v_{s_x}=1,w_{s_y}=1))] + \sum_n \sum_{s_x}\sum_{s_y}\sum_{s'_y}(v_{s_x}w_{s_y}z_{s'_y})^{(k)}[log(p(s_x)p(s_y)p(s'_y|s_x,s_y)) + log(p(y_n|v_{s_x}=1,w_{s_y}=1,z_{s'_y}=1))]. \quad (12)$$

$$E[z_{s'_y}|\mathbf{y}_n, v_{s_x}, w_{s_y}, \mathbf{\Theta}^{(k)}] = \frac{p(s'_y|s_x,s_y)^{(k)}N(\mathbf{y}_n|\mu^{(k)}_{s_x,s_y,s'_y}, \mathbf{\Sigma}^{(k)}_{s_x,s_y,s'_y})}{\sum_{s'_y} p(s'_y|s_x,s_y)^{(k)}N(\mathbf{y}_n|\mu^{(k)}_{s_x,s_y,s'_y}, \mathbf{\Sigma}^{(k)}_{s_x,s_y,s'_y})}. \quad (15)$$

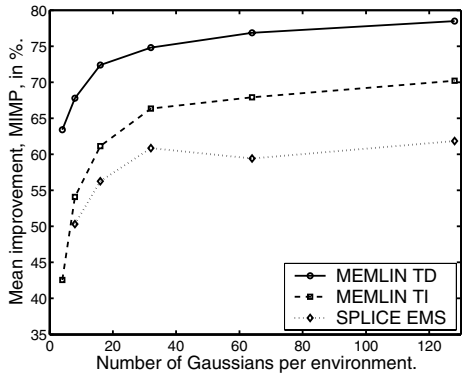

Figure 2: Mean improvement in WER, MIMP, in % for different normalization techniques.

When acoustic models are retrained using all basic environment signals, (Train HF) MWER decreases. Finally, 5.30% of MWER is obtained for environment match condition.

Figure 2 shows the mean improvement in WER (MIMP) in % for MEMLIN with Time-Independent cross-probability model (MEMLIN TI) and with Time-Dependent cross-probability GMM (MEMLIN TD). Also the results with Environmental Model Selection (SPLICE EMS) [4] are included. A 100% MIMP would be achieved when MWER equals the same of clean conditions. The cross-probability GMMs are composed by 2 Gaussians. It can be observed the important improvement of MEMLIN TD concerning MEMLIN TI: from 42.55% to 63.38% with 4 Gaussians per basic environment and from 70.58% to 78.47% with 128 Gaussians. Although the number of Gaussians to model the basic environments could be the same for MEMLIN TI and MEMLIN TD, the computing time is not the same. To reduce it, only the cross-probability GMMs of the most probable pairs of Gaussians can be computed in normalization. In this case, for each noisy feature vector, the most probable Noisy model Gaussians (♯NG) can be obtained with (3) and (4), and for each one, the corresponding most probable Clean model Gaussians (♯CG) are obtained with (6). Table 2 shows the results for MEMLIN TD for different ♯NG and ♯CG. In all cases, the clean and noisy basic environments are modelled with 128 Gaussians, and the cross-probability GMMs are composed by 2 Gaussians. It can be observed that the results always improve the ones obtained with MEMLIN TI with 128 Gaussians per basic environment (70.21%).

6. Conclusions

In this paper we have presented an approach of MEMLIN where the cross-probability model is estimated by modelling the noisy feature vectors associated to each pair of Gaussians from the clean and the noisy spaces with a GMM. MEMLIN obtains an improvement in WER of 70.21% with 128 Gaussians per environment,

Table 2: Mean WER (MWER) and mean improvement in WER (MIMP) in % when different Gaussians of cross-probability GMM are computed.

	♯NG	♯CG	MWER	MIMP
MEMLIN TD 128-128	4	4	5.39	74.81
MEMLIN TD 128-128	8	8	5.53	73.85
MEMLIN TD 128-128	16	16	5.41	74.69
MEMLIN TD 128-128	32	32	5.11	76.77
MEMLIN TD 128-128	64	64	4.86	78.47
MEMLIN TD 128-128	128	128	4.86	78.47

whereas MEMLIN with cross-probability GMM reaches 78.47% for the same number of Gaussians to model each basic environment. Since the computing cost for the proposed approach is higher, an alternative is considered: only the most probable pair of Gaussians of the cross-probability GMM are computed. So, only with the 16 most probable pair of Gaussians, an improvement of 74.81% is obtained, when 128 Gaussians per basic environment are used.

7. References

[1] L. Neumeyer and M. Weintraub, "Robust Speech Recognition in Noise Using Adaptation and Mapping Techniques," in *Proceedings of ICASSP* , vol. 1, 1995, pp. 141–144.

[2] A. Sankar and C. Lee, "A maximum-likelihood approach to stochastic matching for robust speech recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 4, pp. 190– 202, May 1996.

[3] R. M. Stern, B. Raj, and P. J. Moreno, "Compensation for environmental degradation in automatic speech recognition," in *ESCA Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels*, 1997, pp. 33–42.

[4] J. Droppo, L. Deng, and A. Acero, "Evaluation of the SPLICE algorithm on the AURORA2 database," in *Eurospeech*, 2001.

[5] L. Buera, E. Lleida, A. Miguel, and A. Ortega, "Multi-environment models based linear normalization for robust speech recognition in car conditions," in *ICASSP*, 2004.

[6] H. van den Heuvel, J. Boudy, R. Comeyne, S. Euler, A. Moreno, and G. Richard, "The speechdat-car multilingual speech databases for in-car applications: some first validation results," in *Eurospeech*, 1999.

[7] A. P. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from imcomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 9, no. 1, pp. 1–37, 1977.