An $O(1)$ algorithm to find the adaptation coefficients for MAP adaptation os ASR systems operating in noisy conditions

Carlos Alberto Ynoguti & Tatiane Melo Vital

Abstract— **The mismatch of noise conditions between the training and testing utterances is one of the reasons for the dramatic performance degradation experienced by automatic speech recognition system when operating in real world conditions. The Maximum a Posteriori Adaptation is one of the techniques used to face this degradation. A problem with this method is the search for good adaptation coefficients as they have to be found by a scanning process. In this work, we provide an algorithm based on parametric adjustment (using a logistic curve) that returns good adaptation coefficients for this technique.**

Index Terms— **Automatic Speech Recognition, Maximum a Posteriori Adaptation, Multi-Style Training, Parametric Adjustment.**

I. Introduction

It is widely known that ASR (Automatic Speech Recognition) systems performance degrades when operating under noisy conditions [1] and one reason for this fact is the mismatch between the training and the testing acoustic conditions [2].

There are several approaches that try to minimize the effects of background disturbance even in unknown noisy conditions [3], and they can be divided into three main classes:

- the first one is applied before acoustic modeling in front-end signal preprocessing. PLPs or MFCCs helps to minimize the effect of speaker variability. In front-end signal processing, noise suppression methods such as Subtraction Spectral (SS), Wiener filtering and Minimum Mean Square Error (MMSE) estimation are effective to reduce the intensity of noise in speech;
- the second one take into account methods that act in the modeling phase. In this case, clean speech is used in the training phase to ensure a high quality of the final speech models. Then, these models can be transformed according to the noise present during recognition task. This category comprehends: techniques which combine background noise with speech, i.e multi-condition models, or with acoustic models such as parallel model combination (PMC);
- the last approach includes methods which use noisy speech data to adapt acoustic models for a specific background condition by retraining the clean speech models or simply by some transformation as maximum likelihood linear regression (MLLR) or MAP adaptation.

Manuscript received in September 10th, 2013; revised in September 18th, 2013.

C. A. Ynoguti (ynoguti@inatel.br) and T. M. Vital (tati.mv7@gmail.com) are from Instituto Nacional de Telecomunicações - Inatel. Av. João de Camargo, 510 - Santa Rita do Sapucaí - MG - Brazil - 37540-000.

This work is based on [4] approach, that uses a multi-style condition training [5] followed by a Maximum a Posteriori (MAP) adaptation [6]. This method can be summarized as follows:

- in the first stage (multi-style training), a HMM is trained using utterances corrupted by several noise types available on AURORA database [7] at SNRs of 15 dB and 20 dB (this choice of SNRs is based on experimental results [4]). This stage provided a 6.89 % gain in WA (word accuracy) for noisy utterances recognition when compared with a system trained only with clean speech.
- in the second stage, a MAP adaptation was performed to fine tune the system for the actual noise type and SNR that is being experienced by the recognizer. An additional 1.74 % gain in WA was obtained, and thus, the overall gain with these two techniques is 8.63 % over the baseline system.

Although the MAP adaptation gives good results, its performance depends on the correct choice of the so called adaptation coefficients. To date, they have to be found by a scanning process, which is a computacionally inefficient process. Therefore, the main contribution of this work is a method to allow the selection of good adaptation coefficient values in $O(1)$ time, avoiding the scanning process.

The next sections are structured as follows: in Section II, a theoretical framework for the multi-style training and the MAP adaptation is given. Section III presents the proposed method. In Section IV, the experimental setup used for recognition tests is demonstrated and Section V shows the test results. Finally, Section VI brings the final conclusions of the present work.

II. System model.

This work evaluates the combination of multi-style training and MAP estimation, techniques which are presented in the next subsections, to overcome effects caused by different noise types and levels.

A. Multi-Style Trainining

Hidden Markov Models are classifiers. In this sense, their performance heavily rely on the information available in the training stage. If it is desired that they work in several noise conditions, with different channel distortions, reverberation, etc., it is necessary that the training utterances are collected in such environments. Obviosly, the construction of a database that reflects all situations of day to day is infeasible and impractical given the great variability of environmental adverse conditions.

The multi-style or multi-condition training employs utterances artificially corrupted by different noise type and levels in the training stage in order to minimize the performance drop of ASR systems operating in noisy environments [8].

Different approaches can be used in this method: the system can be trained for a particular noise type and level according to environmental condition, or with different levels of a specific noise type, or even, with different noise types and levels. The present work employs the last approach.

B. Maximum a Posteriori Adaptation

In the context of HMMs, the MAP approach consists in the adaptation of the canonical model parameters (mixture weight, mean and variance) using estimated statistics of the background noise only. Usually, it leads to good word accuracy because it provides the modeling of the uncertainty caused by noisy environmental statistics [9].

A canonical model is a HMM generated in the training phase using noisy or clean utterances of several speakers. Then noise statistics from environment are used to adapt these models. The hypothesized speech model is derived by adapting the parameters of canonical model and a form of Bayesian adaptation [6].

The adaptation equations for these parameters are described as follows: given a noise sample and training vectors from the hypothesized speech, $X = x_1, x_2, ..., x_T$, the probabilistic alignment of the noise into the canonical model is given by:

$$\Pr(i|x_t) = \frac{\omega_i p_i(x_t)}{\sum_{j=1}^{M} \omega_j p_j(x_t)} \tag{1}$$

where M is the number of Gaussian densities, ω is the mixture weight and p is the probability density function.

Then, $Pr(i|x_t)$ and x_t are used to determinate the noisy statistical parameters weight (n_i), mean ($E_i(x)$) and variance ($E_i(x^2)$), as described below:

$$n_i = \sum_{t=1}^{T} Pr(i|x_t) \tag{2}$$

$$E_i(x) = \frac{1}{n_i} \sum_{t=1}^{T} Pr(i|x_t) x_t \tag{3}$$

$$E_i(x^2) = \frac{1}{n_i} \sum_{t=1}^{T} Pr(i|x_t) x_t^2 \tag{4}$$

Finally, these estimated statistics of background noise are used to adapt the canonical models generating a new model. The adaptation equations for these parameters are:

$$\hat{\omega}_i = \left[\alpha_i^{\omega} n_i / T + (1 - \alpha_i^{\omega}) \omega_i\right] \gamma \tag{5}$$

$$\hat{\mu}_i = \alpha_i^{m} E_i(x) + (1 - \alpha_i^{m}) \mu_i \tag{6}$$

$$\hat{\sigma}_i^2 = \alpha_i^{\nu} E_i(x^2) + (1 - \alpha_i^{\nu})(\sigma_i^2 + \mu_i^2) - \mu_i^2 \tag{7}$$

where:

- ω_i, μ_i and σ_i^2 are the mixture weights, means and variances of the multi-style trained system;
- $\hat{\omega}_i$, $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ are the mixture weights, means and variances after adaptation and
- n_i, $E_i(x)$ and $E_i(x^2)$ are the noise statistics.

The adaptation coefficients α_i^{ω}, α_i^{m} and α_i^{ν} can assume values in the $[0, 1]$ interval and control the balance between old and new estimates for the weights, means and variances, respectively.

It seems reasonable that a good choice of these parameters depends on the noise type and its intensity, since higher values favor the noise estimates whereas lower values tend to preserve the original ones.

It is possible to employ different coefficient values to adapt weights, means and variances. However, this approach provides a small gain compared to use solely a single value (α_i^{ω}=α_i^{m}=α_i^{ν}) when adapting them [6]. Therefore, in this work, a single adaptation coefficient for all parameter is used.

As mentioned before, the performance of this technique depends on a good choice of this adaptation coefficient. Unfortunately, the process reported in the literature to find such optimal value is based on a scanning process: varying the value of α in the $(0, 1)$ range, the optimal value is the one that leads to the best recognition performance. Clearly this approach has a big computational cost, and in the next section the method we devised to avoid such scanning process is presented in details.

III. Algorithm to find the α Coefficient

As aforementioned, the proposed method tries to find an α value that leads to a performance improvement when comparing to a non-adapted system, for the noise being experimented by the recognition system in a given moment. In other words, the goal is not to find an α that leads to the maximum system performance but to provide a value that returns a gain compared to the baseline.

The algorithm is outlied below:

- for each value of SNR, perform a grid search for the best values of α in the (0,1) interval and record the ones that lead to a performance improvement (in our experiments, we used 0 dB, 5 dB, 10 dB, 15 dB and 20 dB).
- For a given SNR, there can be several values of α that lead to a performance improvement, but we keep only one. After several tests, we decided to choose a weighted average, given by:

$$\alpha' = \frac{\displaystyle\sum_i WA(i) \times \alpha(i)}{\displaystyle\sum_i WA(i)} \tag{8}$$

where $WA(i)$ is the word accuracy obtained by using the value $\alpha(i)$, and α' is the chosen value for the α parameter.

Therefore, after this step, there is a single α' value for each value of SNR.

- By looking at the curves generated by several tests, we observed that they resembled the format of a logistic curve. Thus, as a final step, a logistic curve with three free parameters was adjusted to the experimental points [10].

$$f(x) = \frac{1}{1 + e^{b-ax}} - c \tag{9}$$

where x is the noise level and $f(x)$ is the adaptation coefficient. The configuration parameters a, b and c can be obtained by curve fitting techniques. These parameters can be interpreted as follows:

  - a parameter determines the slope of the logistic curve. The smaller its value, the steeper is the curve;
  - b parameter controls the horizontal offset. If its value decreases, the curve is shifted to the right. Otherwise, it is shifted to left and
  - c parameter is the offset, allowing the vertical adjustment. If its value increases, the curve is moved down. Otherwise, it is shifted up.

Figures 1, 2 and 3 show the behavior of a, b and c parameters, respectively.

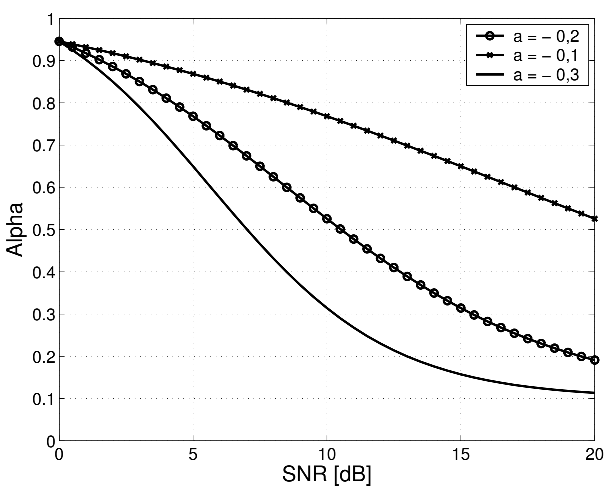

Fig. 1. Example of different values for ‘a’ parameter

This is an important step because it allows the choice of α values for SNRs different from the ones used to generate the curve.

In the next section the experimental setup used in the recognition tests is presented.

IV. Experimental Setup

In this section, the database and speech recognition engine ised for the experiments are described.

A. Database

Experiments were performed using a 40 speakers (20 male and 20 female) clean speech database [11]. Each speaker recorded 40 phonetically balanced utterances in Brazilian

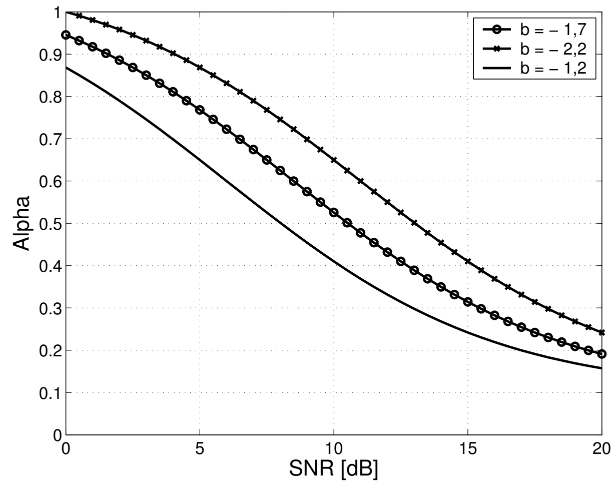

Fig. 2. Example of different values for ‘b’ parameter

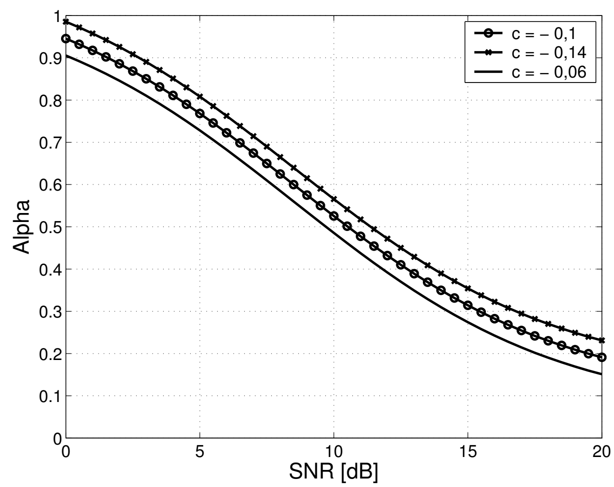

Fig. 3. Example of different values for ‘c’ parameter

Portuguese which were drawn from [12]. The corpus has 1600 sentences comprising 694 different words and it was divided in two groups: training corpus (1200 utterances) and testing corpus (400 utterances).

The recordings were performed in a low noise environment at 11,025 kHz sample rate and 16-bit coded. The sampling frequency was lowered to 8 kHz to make them compatible with the AURORA database.

The original database was artificially corrupted by noises (airport, babble, car, exhibition, restaurant, street, subway and train) from the Aurora Project database [7]. For the training material, for each clean sentence, 2 new versions were created combining each noise type at levels 15 and 20 dB as proposed by [4]. For each clean utterance of testing corpus, 5 new versions were created adding each noise type at levels 0, 5, 10, 15 and 20 dB. Therefore, training and testing database have now 19200 and 16000 corrupted utterances, respectively.

B. Speech recognition engine

To test our ideas, a continuous density HMM based speech recognition engine developed by [11] was used. It uses context independent phones as fundamental units, with each of them modeled as a 3 state Markov chain, as shown in Figure 4 and the One Pass search algorithm [13]. A mixture of 10 multi-dimensional Gaussian distributions with diagonal covariance matrix was used in each state.

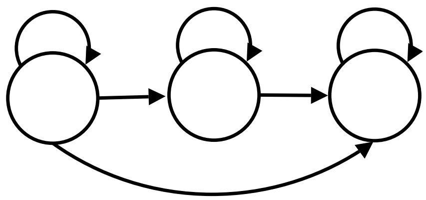

Fig. 4. Markov chain for each phone model

As acoustic parameters, 12 mel-cepstral coefficients together with their first and second derivatives were used, leading to feature vectors of dimension 36. Finally, to improve the system performance, a bigram language model was applied.

V. Experimental Results

To verify that the proposed method leads to a performance improvement over the baseline system, the following tests were performed:

A. Baseline

The baseline performance is obtained from an ASR system trained using the multi-style approach. For training, all noise types available in the AURORA database at SNR levels of 15 dB and 20 dB were used.

The results of the tests with this system are shwon in the column named "Reference WA" (second column) of Tables I to VIII.

TABLE I
Airport test results.

SNR (dB)	Reference WA (%)	Maximum WA (%)	α for maximum WA	weighted α	WA for weighted α (%)
0	3.2	7.3	0.6500	0.3387	6.3
5	25.4	32.1	0.4500	0.2779	31.1
10	62.9	66.2	0.1500	0.1950	65.6
15	78.0	77.8	0.0100	–	–
20	78.0	77.8	0.0100	–	–

TABLE II
Babble test results.

SNR (dB)	Reference WA (%)	Maximum WA (%)	α for maximum WA	weighted α	WA for weighted α (%)
0	4.6	5.9	0.1500	0.2172	5.8
5	32.2	37.4	0.2000	0.2139	37.1
10	66.3	66.7	0.1500	0.1500	66.2
15	77.1	76.9	0.0100	–	–
20	77.1	76.9	0.0100	–	–

TABLE III
Car test results.

SNR (dB)	Reference WA (%)	Maximum WA (%)	α for maximum WA	weighted α	WA for weighted α (%)
0	5.3	7.1	0.3000	0.2167	6.0
5	31.7	37.3	0.3500	0.2195	36.3
10	65.8	67.7	0.1500	0.1017	67.5
15	76.5	76.6	0.0400	0.0350	76.3
20	76.5	76.6	0.0400	0.0350	76.3

TABLE IV
Exhibition test results.

SNR (dB)	Reference WA (%)	Maximum WA (%)	α for maximum WA	weighted α	WA for weighted α (%)
0	0.2	1.8	0.4000	0.2310	1.0
5	16.4	23.5	0.0900	0.2283	23.1
10	58.2	58.7	0.1500	0.1267	58.0
15	75.5	75.5	0.0200	–	–
20	75.5	75.5	0.0200	–	–

TABLE V
Restaurant test results.

SNR (dB)	Reference WA (%)	Maximum WA (%)	α for maximum WA	weighted α	WA for weighted α (%)
0	13.8	15.4	0.2500	0.1751	14.8
5	4.4	9.1	0.3500	0.2371	8.5
10	35.5	41.6	0.3500	0.1976	41.1
15	69.2	69.7	0.0300	0.0325	69.7
20	69.2	69.7	0.0300	0.0325	69.7

TABLE VI
Street test results.

SNR (dB)	Reference WA (%)	Maximum WA (%)	α for maximum WA	weighted α	WA for weighted α (%)
0	15.6	18.6	0.2000	0.1043	18.4
5	56.1	57.1	0.0700	0.0600	56.3
10	73.6	73.6	0.0200	–	–
15	75.8	77.1	0.0200	0.0250	76.8
20	75.8	77.1	0.0200	0.0250	76.8

TABLE VII
Subway test results.

SNR (dB)	Reference WA (%)	Maximum WA (%)	α for maximum WA	weighted α	WA for weighted α (%)
0	−0.3	3.4	0.4000	0.3186	2.7
5	18.4	25.8	0.3000	0.2594	25.6
10	57.9	61.8	0.0500	0.1193	61.7
15	73.0	73.4	0.0800	0.0650	72.9
20	73.0	73.4	0.0800	0.0650	72.9

TABLE VIII
TRAIN TEST RESULTS.

SNR (dB)	Reference WA (%)	Maximum WA (%)	α for maximum WA	weighted α	WA for weighted α (%)
0	7.6	12.6	0.4500	0.3174	12.1
5	34.0	40.7	0.4500	0.2398	38.8
10	69.0	70.0	0.0400	0.0749	69.9
15	78.1	78.1	0.0100	–	–
20	78.1	78.1	0.0100	–	–

B. Results with adaptation using optimum α values

To evaluate the maximum improvement in the recognition performance using multi-condition training and MAP adaptation combined together, the HMM obtained in the baseline system was adapted for each noise type, generating new 8 models.

In this test, the value of the adaptation parameter α was chosen by a scanning process, with the selection criterion being the highest WA.

The recognition results and the chosen value for α are shown in the third ("Maximum WA") and fourth ("α for maximum WA") columns of Tables I to VIII.

It can be seen that the adition of the adaptation stage provides an improvement of the recognition performance when comparing with the results obtained by the baseline system, for almost all noise types and levels. In a few situation, this technique led to no gain or a little drop in the WA.

C. Results with adaptation using α values computed by the proposed method

To avoid the scanning process, the adaptation coefficients α were calculated using the algortihm shown in Section III. The coefficient values for each noise type are shown in Table IX, and the resulting curves are shown in Figures 5 to Fig 8.

With this settings, a new bunch of tests was performed, and the resulting α values, together with the recognition results can be viewed in the two last columns ("Weighted α") and ("WA for weighted α") of Tables I to VIII.

These results show that, as expected, this technique provides α values that lead to a performance improvement, but not to the best possible improvement. Also, for some SNRs it doesn't exist a weighted α value that lead to a performance improvement. For these cases, MAP adaptation introduced a little performance drop or did not provide gain comparing to the baseline, and therefore these results are not shown in the Tables.

D. Extending the results for other SNRs

A final question to be answered is: are the α values obtained by this method adequate for other SNRs that are different from the ones used to obtain the logistic curve?

To validate the proposed logistic functions, recognition tests were performed for different preselected SNR levels using the canonical model adapted which their correspondent adaptation factors from respective curves. Experimental results showed

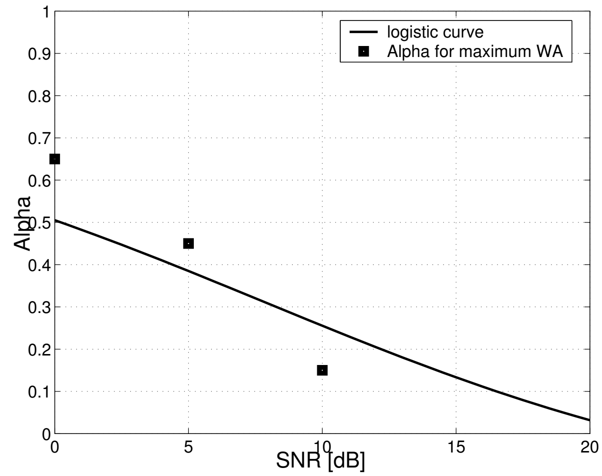

Fig. 5. Logistic curve for recognition using airport noise

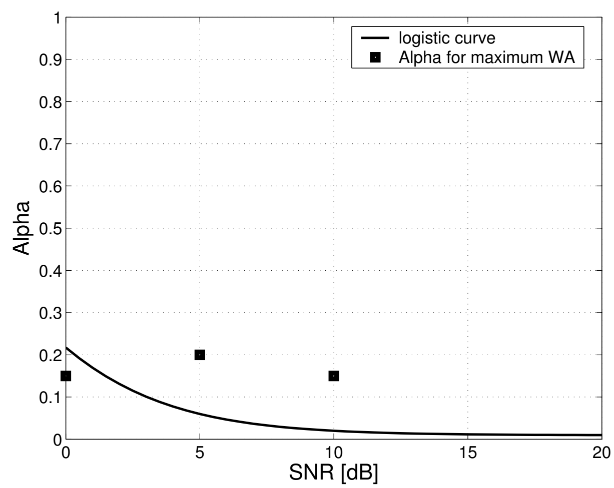

Fig. 6. Logistic curve for recognition using babble noise

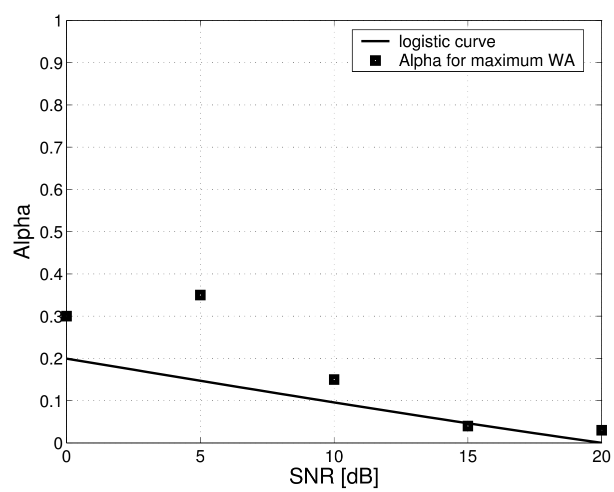

Fig. 7. Logistic curve for recognition using car noise

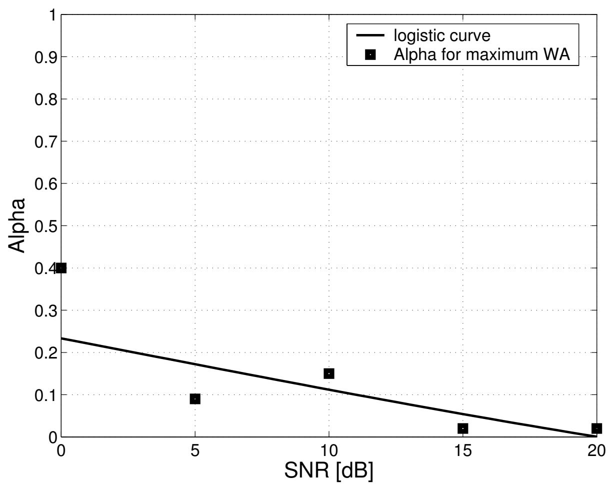

Fig. 8. Logistic curve for recognition using exhibition noise

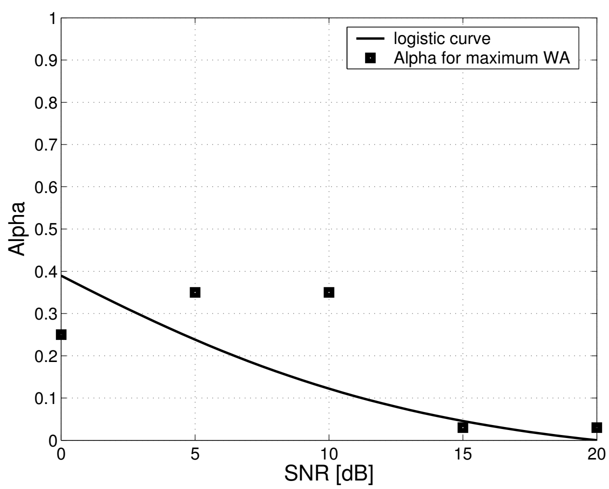

Fig. 9. Logistic curve for recognition using restaurant noise

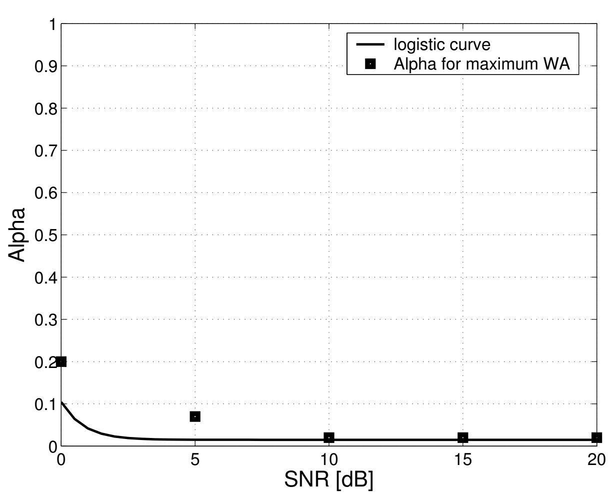

Fig. 10. Logistic curve for recognition using street noise

TABLE IX

COEFFICIENTS FOR LOGISTIC CURVE.

Noise	a	b	c
airport	-0.104175	-0.810155	0.186922
babble	-0.319246	1.338508	-0.009345
car	-0.442008	1.445377	-0.025966
exhibition	-0.124371	1.049212	0.028005
restaurant	-0.148612	1.701653	-0.020750
street	-1.264809	2.320370	-0.014900
subway	-0.131225	1.000045	-0.049691
train	-0.299248	0.882365	0.000910

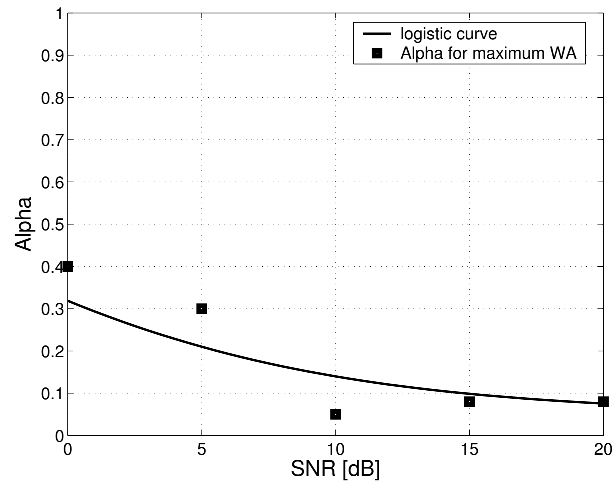

Fig. 11. Logistic curve for recognition using subway noise

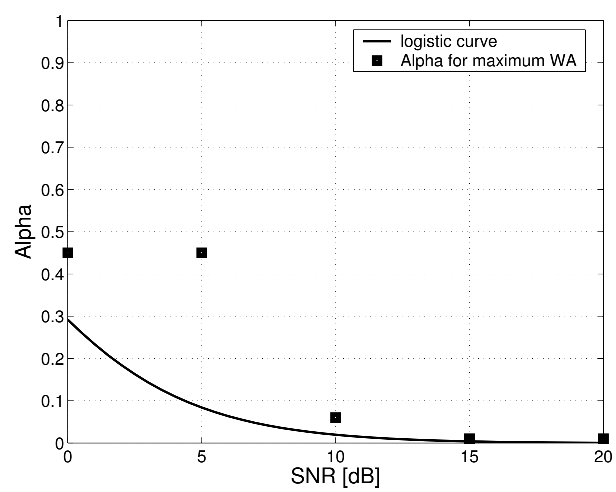

Fig. 12. Logistic curve for recognition using train noise

that proposed algorithm led to a good adaptation coefficient value improving the system performance as can be observed in Tables X, XI and XII.

VI. CONCLUSIONS

In this paper we showed that a combination of multi-style training together with the MAP adaptation leads to a performance improvement of an ASR system operating in noisy conditions.

TABLE X

WORD ACCURACY FROM LOGISTIC CURVE FOR SNR = 2 dB.

Noise	α from Logistic Curve	Reference WA (%)	WA using Logistic Curve (%)	ΔWA (%)
airport	0.4591	7.9	12.8	4.9
babble	0.1310	10.5	15.4	4.9
car	0.1147	12.5	15.9	3.4
exhibition	0.1865	2.4	6.9	4.5
restaurant	0.1401	6.1	7.5	1.4
street	0.0227	30.3	31.5	1.2
subway	0.2702	2.3	9.0	6.7
train	0.1844	12.3	15.6	3.3

TABLE XI

WORD ACCURACY FROM LOGISTIC CURVE FOR SNR = 7 dB.

Noise	α from Logistic Curve	Reference WA (%)	WA using Logistic Curve (%)	ΔWA (%)
airport	0.3333	39.8	47.3	7.5
babble	0.0366	48.7	50.3	1.6
car	0.0365	49.8	50.8	1.0
exhibition	0.0999	31.3	38.4	7.1
restaurant	0.0813	11.4	15.5	4.1
street	0.0149	65.9	65.9	0.0
subway	0.1777	33.9	40.8	6.9
train	0.0476	51.6	53.4	1.8

TABLE XII

WORD ACCURACY FROM LOGISTIC CURVE FOR SNR = 12 dB.

Noise	α from Logistic Curve	Reference WA (%)	WA using Logistic Curve (%)	ΔWA (%)
airport	0.2048	71.7	72.4	0.7
babble	0.0150	73.7	74.2	0.5
car	0.0271	73.9	73.9	0.0
exhibition	0.0450	69.1	69.6	0.5
restaurant	0.0505	53.2	54.9	1.7
street	0.0149	76.0	76.0	0.0
subway	0.1205	66.8	67.3	0.5
train	0.0104	74.6	74.6	0.0

After the multi-style training, the system was presented to different noise types and levels. However, it is very difficult that, in real operating conditions, the same noise type and level will be present in the utterance to be recognized. The MAP adaptation stage is responsible to further tune the system models to the actual noise type and level of the utterance being processed. However, the adaptation stage needs the value of the adaptation coefficient α, which is usually calculated by a costly scanning procedure.

The main contributions of this work are: i) the modelling of the relationship between the adaptation coefficient α versus the noise type and level, as a logistic curve, and ii) using this information, establishing a method to calculate the adaptation coefficient α for a given noise type and level in $O(1)$ time.

The proposed strategy finds an α value for a given type and noise level that ensures an improvement when compared to the baseline system. Although it does not give the α value that leads to the best possible performance, in practice it gives an improvement of approximately 3% on system performance compared to the reference value (multi-style trained system).

REFERENCES

[1] L. Lu, A. Ghoshal and S. Renals. Maximum a Posteriori Adaptation of Subspace Gaussian Mixture Models for Cross-Lingual Speech Recognition. IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4877-4880, March 2012.

[2] S. Furui. *50 years of progress in speech recognition technology: Where we are, and where we should go? From a poor dog to a super cat*. Keynote Presentation, ICASSP 2007.

[3] J. Rajnoha. Multi-Condition Trainin for unknown environment adaptation in robust ASR under real conditions. Acta Polytecnica, Vol. 49, No. 2-3, 2009.

[4] T. A. F. Valério. Treinamento multi-estilo e adaptação de modelos via MAP para reconhecimento de fala em ambientes ruidosos. Dissertação de Mestrado. Inatel. Santa Rita do Sapucaí, 2009.

[5] Huang, X., Acero, A. and Hon, H.W. Spoken Language Processing: A Guide to Theory, Algorithm and System Development. Prentice Hall. 2001.

[6] D. A. Reynolds, T. F. Quatieri and R. B. Dunn. Speaker Verification using Adapted Gaussian Mixture Models. Digital Signal Processing, vol. 10, pp.19-41, 2000.

[7] H. G. Hirsch and D. Pearce. The Aurora experimental framework for the evaluation of speech recognition systems under noisy conditions. In Proc. ASR-2000, pp 181-188,September 2000.

[8] R. Lippman, E. Martin and D. Paul. Multi-style training for robust isolated-word speech recognition. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, vol. 12, pp. 705-708, 1987.

[9] L. Buera, E. Lleida, A. Miguel, A. Ortega and O. Saz. Cepstral Vector Normalization Based on Stereo Data for Robust Speech Recognition. IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, pp. 1098-1113, 2007.

[10] A. G. Chiovato. Avaliação da Relação entre Qualidade Perceptual da Fala e Taxa de Acerto de Sistemas de Reconhecimento de Fala em Ambientes Ruidosos. Dissertação de Mestrado. Inatel. Santa Rita do Sapucaí, 2005.

[11] C. A. Ynoguti and F. Violaro. Desenvolvimento de um Conjunto de Ferramentas para Pesquisa em Reconhecimento de Fala. Revista Telecomunicações, vol. 4(2), pp.36-43, 2001.

[12] A. Alcaim, J. A. Solewicz and J. A. Moraes, Frequência de Ocorrência dos Fones e Listas de Frases Foneticamente Balanceadas no Português Falado no Rio de Janeiro. Revista da Sociedade Brasileira de Telecomunicações, vol. 7(1), pp.23-41, 1992.

[13] H. Ney, The use of a one-stage dynamic programming algorithm for conected word recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, ASSSP-32(2), 1984.

Tatiane Melo Vital was born in Volta Redonda, RJ, in December 1982. She received the BsC. Degree in Electrical Engineering from Inatel in 2006 and the MsC. degree in Telecommunications Engineering also from Inatel in September 2013. She worked five years at the Centre for Research and Development at Ericsson Telecommunications, developing software for wireline core.

Carlos Alberto Ynoguti was born in São Paulo, SP, in May 1967. He received the BsC. and MsC. degrees in Electrical Engineering from EESC-USP in 1991 and 1994, respectively. In 2000 he received the Doctor in Electrical Engineering from FEEC-Unicamp, where he stayed for 2 more years in a post-doctorate program in Automatic Speech Recognition.

Since December 2000 he is professor at Inatel, and his current research interests are in Digital Signal Processing, Speech Processing and Artificial Intelligence.

From 2005 to 2013 he was editor of Revista Telecomunicações, and General Chair of the International Workshop of Telecommunications in the 2007, 2009, 2011 and 2013 editions.