Discriminative adaptation based on fast combination of DMAP and DfMLLR

Lukáš Machlica, Zbyněk Zajíc, Luděk Müller

University of West Bohemia in Pilsen,Faculty of Applied Sciences,
Department of Cybernetics, Univerzitní 22, 306 14 Pilsen
machlica@kky.zcu.cz, zzajic@kky.zcu.cz, muller@kky.zcu.cz

Abstract

This paper investigates the combination of discriminative adaptation techniques. The discriminative Maximum A-Posteriori (DMAP) adaptation and discriminative feature Maximum Likelihood Linear Regression (DfMLLR) are examined. Since each of the methods is proposed for distinct amount of adaptation data it is useful to combine them in order to preserve the systems performance in situations with varying amount of adaptation data. Generally, DfMLLR and DMAP are executed subsequently (DMAP preceded by DfMLLR) demanding to approach the data twice. Since both methods address the data through the same statistics an one-pass-combination was proposed in order to decrease the time consumption. The one-pass-combination utilizes the advantage of DfMLLR method to transform directly the feature vectors. However, instead of feature vectors the statistics are transformed, what allows to use already computed statistics for the DMAP pass without the need to process the data once again. All the approaches are compared also to their non-discriminative alternatives.

Index Terms: MAP, fMLLR, DMAP, DfMLLR, MMI, adaptation, speech recognition, combination

1. Introduction

In the field of speech recognition adaptation techniques took an important role significantly increasing the systems performance and robustness. Standard adaptation methods are based on Maximum Likelihood Estimation (MLE) procedure. In order to facilitate MLE of model parameters several assumptions are introduced, which the real data do not fulfill. Still, MLE models behave well, and were successfully applied in praxis. In the past few years, new approaches trying to improve standard MLE methods were presented. One of the most significant approaches is the discriminative training, which in contrast to MLE tries to handle also overlaps between distinct parts (distributions) of MLE models. Loosely speaking, the MLE criterion is adjusted in order to involve and prevent situations when probability distributions of distinct sources (e.g. speech data of different phones) coincide in greater or lesser extent. Further description of discriminative criteria may be found in Section 2.

This paper will focus on well-known MLE adaptation techniques and their discriminative alternatives. Since the functionality of individual adaptation techniques was already verified [1] the paper aims to examine the behavior of their combination. Two adaptation methods were chosen according to their complementarity, Discriminative Maximum A-Posteriory (DMAP) adaptation and Discriminative feature Maximum Likelihood Linear Regression (DfMLLR) adaptation. DfMLLR was proposed to handle the problem with lower amount of adaptation data, whereas DMAP dominates in situations when lots of data are available. However, as described in Section 6, DMAP used in combination with DfMLLR (DfMLLR succeeded by DMAP) may be interpreted as a refinement stage of adaptation. Such a combination significantly improves the systems accuracy, but demands to process the input data twice. Thus, a method avoiding the need of the second pass was proposed in Section 6. To enlighten the procedure description of important parts of MAP and DfMLLR will be given in Sections 4 and 5, respectively. The performance of the system enhanced with adaptation techniques including their combination and analysis of obtained results can be found in Section 7.3.

2. Discriminative adaptation techniques

Standard adaptation methods are mostly based on Maximum Likelihood Estimation (MLE) used in large extent in order to estimate HMM parameters. In MLE the following criterion is maximized

$$\mathcal{F}_{MLE}(\boldsymbol{\lambda}) = p(\boldsymbol{O}|W_{ref}, \boldsymbol{\lambda}), \tag{1}$$

where $\boldsymbol{O} = \{\boldsymbol{o}_1, \boldsymbol{o}_2, \ldots, \boldsymbol{o}_T\}$ represents the sequence of T feature vectors related to one speaker, W_{ref} is the corresponding correct transcription, and $\boldsymbol{\lambda}$ denotes the set of model parameters. Focus will be laid on HMMs with output probabilities of states represented by GMMs, where $\boldsymbol{\lambda}_j = \{\omega_{jm}, \boldsymbol{\mu}_{jm}, \boldsymbol{C}_{jm}\}_{m=1}^{M_j}$ is the set of GMM parameters in the $j-th$ state, where M_j is the number of mixtures, ω_{jm}, $\boldsymbol{\mu}_{jm}$ and $\boldsymbol{C}_{jm}$ are weight, mean and variance of the $m-th$ mixture, respectively. In the scope of this paper only diagonal matrices with $\boldsymbol{\sigma}^2_{jm}$ = diag$(\boldsymbol{C}_{jm})$ are assumed.

MLE criterion takes into account only correct transcriptions, on the other hand discriminative criteria consider in addition incorrect hypotheses. The discriminative training criterion was developed in order to increase the posterior probability of model states corresponding to their adaptation data and decrease probability of confusion data (data incorrectly assigned to HMM states) at the same time. One of the possibilities is to utilize the Maximum Mutual Information (MMI) criterion [2]:

$$\mathcal{F}_{MMI}(\boldsymbol{\lambda}) = \frac{p(\boldsymbol{O}|W_{ref}, \boldsymbol{\lambda})P(W_{ref})}{\sum_W p(\boldsymbol{O}|W, \boldsymbol{\lambda})P(W)}, \tag{2}$$

where W_{ref} is a transcription corresponding to the observation $\boldsymbol{O}$ and W is a transcription with all possible hypothesis.

Another criteria are e.g. Maximum Mutual Information Frame Discrimination (MMI-FD) [3] or Minimum Phone Error (MPE) [4]. The main problem consists in the optimization process, where mainly weak-sense auxiliary function is used [5]. Regrettably, it does not guarantee the convergence of the discriminative criterion. In order to adjust the stability of discriminative criteria a smoothing term is involved, which will be introduced in sequel.

3. Adaptation Statistics

Adaptation techniques do not access the data directly, but only through accumulated statistics, which is the first step preceding the adaptation process. These statistics are

$$\gamma_{jm}(t) = \frac{\omega_{jm} p(\boldsymbol{o}_t|jm)}{\sum_{m=1}^{M} \omega_{jm} p(\boldsymbol{o}_t|jm)} \tag{3}$$

standing for the $m-th$ mixtures' posterior of the $j-th$ state of the HMM,

$$c_{jm} = \sum_{t=1}^{T} \gamma_{jm}(t) \tag{4}$$

representing the soft count of mixture m,

$$\boldsymbol{\varepsilon}_{jm}(\boldsymbol{o}) = \sum_{t=1}^{T} \gamma_{jm}(t) \boldsymbol{o}_t \,, \tag{5}$$

$$\boldsymbol{\varepsilon}_{jm}(\boldsymbol{o}\boldsymbol{o}^{\mathrm{T}}) = \sum_{t=1}^{T} \gamma_{jm}(t) \boldsymbol{o}_t \boldsymbol{o}_t^{\mathrm{T}} \tag{6}$$

denoting the sum of the first and the second moment of features aligned to mixture m in the j-th state of the HMM.

For MMI approach, also denominator statistics $\gamma_{jm}^{den}(t)$, c_{jm}^{den}, $\boldsymbol{\varepsilon}_{jm}^{den}(\boldsymbol{o})$ and $\boldsymbol{\varepsilon}_{jm}^{den}(\boldsymbol{o}\boldsymbol{o}^{\mathrm{T}})$ for confusable states must be accumulated. These are computed in the sense of the denominator in the equation (2).

4. Discriminative Maximum A-posteriori Probability (DMAP) Adaptation

Standard (non-discriminative) MAP is based on the Bayes method for estimation of the acoustic model parameters. MAP demands a huge amount of data, because each of the HMM parameters is adapted separately. In order to demonstrate the differences between MAP and DMAP, adaptation of GMM means will be described, remaining formulas can be found in [6], [7] for MAP, DMAP, respectively. In the case of MAP adaptation means are adapted according to formula

$$\bar{\boldsymbol{\mu}}_{jm} = \frac{\boldsymbol{\varepsilon}_{jm}(\boldsymbol{o}) + \tau_{jm}\boldsymbol{\mu}_{jm}}{c_{jm} + \tau_{jm}} \,, \tag{7}$$

where τ_{jm} is an empirically determined parameter, which controls the balance between old and new parameters.

DMAP adaptation, according to MMI criterion, uses discriminative statistics mentioned in Section 3, i.e. discriminative statistics are subtracted from MLE statistics to replace them, $c_{jm} := c_{jm} - c_{jm}^{den}$, $\boldsymbol{\varepsilon}_{jm}(\boldsymbol{o}) := \boldsymbol{\varepsilon}_{jm}(\boldsymbol{o}) - \boldsymbol{\varepsilon}_{jm}^{den}(\boldsymbol{o})$ and $\boldsymbol{\varepsilon}_{jm}(\boldsymbol{o}\boldsymbol{o}^{\mathrm{T}}) := \boldsymbol{\varepsilon}_{jm}(\boldsymbol{o}\boldsymbol{o}^{\mathrm{T}}) - \boldsymbol{\varepsilon}_{jm}^{den}(\boldsymbol{o}\boldsymbol{o}^{\mathrm{T}})$, and an additional smoothing term $D_{jm} = f \cdot c_m^{den}$ with a weighting factor f is introduced. Means are now adapted according to formula

$$\bar{\boldsymbol{\mu}}_{jm} = \frac{\boldsymbol{\varepsilon}_{jm}(\boldsymbol{o}) - f\boldsymbol{\varepsilon}_{jm}^{den}(\boldsymbol{o}) + \tau_{jm}\boldsymbol{\mu}_{jm}}{c_{jm} - D_{jm} + \tau_{jm}} \,, \tag{8}$$

As can be seen from equations (7), (8) the only difference between MAP and DMAP consists in shifting of correctly accumulated statistics (data correctly assigned to HMM states) away from denominator statistics (data assigned to incorrect HMM states). This will be the same for the fMLLR vs DfMLLR case.

5. Discriminative Feature Maximum Likelihood Linear Regression (DfMLLR)

DfMLLR technique belongs to the category of Discriminative Linear Transformations (DLTs), another DLT based method is Discriminative Maximum Likelihood Linear Regression (MLLR-DLT). These are the discriminative extensions of Linear Transformations (LTs). Similar model components are clustered into clusters $K_n, n = 1, \ldots, N$ in order to lower the number of adapted parameters [10]. Thus, in contrast to MAP (resp. DMAP) lower amount of adaptation data suffices. fMLLR transforms directly features $\boldsymbol{o}_t$ according to

$$\bar{\boldsymbol{o}}_t = \boldsymbol{A}_{(n)}\boldsymbol{o}_t + \boldsymbol{b}_{(n)} = \boldsymbol{W}_{(n)}\boldsymbol{\xi}(t) \,, \tag{9}$$

where

$$\boldsymbol{W}_{(n)} = [\boldsymbol{A}_{(n)}, \boldsymbol{b}_{(n)}], \tag{10}$$

$\boldsymbol{W}_{(n)}$ represents the transformation matrix corresponding to the $n-th$ cluster K_n and $\boldsymbol{\xi}(t) = [\boldsymbol{o}_t^{\mathrm{T}}, 1]^{\mathrm{T}}$ stands for the extended feature vector.

The estimation formulas for rows of $\boldsymbol{W}_{(n)}$ are given as

$$\boldsymbol{w}_{(n)i} = \boldsymbol{G}_{(n)i}^{-1} \left(\frac{\boldsymbol{v}_{(n)i}}{\alpha_{(n)}} + \boldsymbol{k}_{(n)i} \right) \,, \tag{11}$$

where $\boldsymbol{v}_{(n)i}$ is the i-th row vector of cofactors of matrix $\boldsymbol{A}_{(n)}$, $\alpha_{(n)}$ can be found as a solution of a quadratic function defined in [8],

$$\boldsymbol{k}_{(n)i} = \sum_{m \in K_n} \frac{\mu_{mi}\boldsymbol{\varepsilon}_m(\boldsymbol{\xi})}{\sigma_{mi}^2}, \tag{12}$$

$$\boldsymbol{G}_{(n)i} = \sum_{m \in K_n} \frac{\boldsymbol{\varepsilon}_m(\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}})}{\sigma_{mi}^2}, \tag{13}$$

and

$$\boldsymbol{\varepsilon}_m(\boldsymbol{\xi}) = \left[\boldsymbol{\varepsilon}_m^{\mathrm{T}}(\boldsymbol{o}), c_m\right]^{\mathrm{T}} \,, \tag{14}$$

$$\boldsymbol{\varepsilon}_m(\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}) = \begin{bmatrix} \boldsymbol{\varepsilon}_m(\boldsymbol{o}\boldsymbol{o}^{\mathrm{T}}) & \boldsymbol{\varepsilon}_m(\boldsymbol{o}) \\ \boldsymbol{\varepsilon}_m^{\mathrm{T}}(\boldsymbol{o}) & c_m \end{bmatrix} \,. \tag{15}$$

Equation (11) is the solution of minimization problem with auxiliary function given in [8]. Matrices $\boldsymbol{A}_{(n)}$ and $\boldsymbol{b}_{(n)}$ are estimated iteratively, thus they have to be initialized (e.g. randomly).

The discriminative approach (DfMLLR) uses discriminative auxiliary function specified in [4]. The computation differs only in estimation of auxiliary matrices $\boldsymbol{G}_{(n)i}$ and $\boldsymbol{k}_{(n)i}$, where discriminative statistics are subtracted from MLE ones in order to replace them (in analogy with Section 4). Hence, auxiliary matrices are computed according to

$$\boldsymbol{k}_{(n)i} = \sum_{m \in K_n} \frac{\mu_{mi}}{\sigma_{mi}^2} (\boldsymbol{\varepsilon}_m(\boldsymbol{\xi}) - D_m[\boldsymbol{\mu}_m, 1]^{\mathrm{T}}), \tag{16}$$

$$\boldsymbol{G}_{(n)i} = \sum_{m \in K_n} \frac{1}{\sigma_{mi}^2} (\boldsymbol{\varepsilon}_m(\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}) - D_m \boldsymbol{Z}_m), \tag{17}$$

where

$$\boldsymbol{Z}_m = \begin{bmatrix} \boldsymbol{\Sigma}_m + \boldsymbol{\mu}_m\boldsymbol{\mu}_m^{\mathrm{T}} & \boldsymbol{\mu}_m^{\mathrm{T}} \\ \boldsymbol{\mu}_m & 1 \end{bmatrix}, \tag{18}$$

$$D_m = f.c_m^{den} \tag{19}$$

is the smoothing term, and f is a weighting factor.

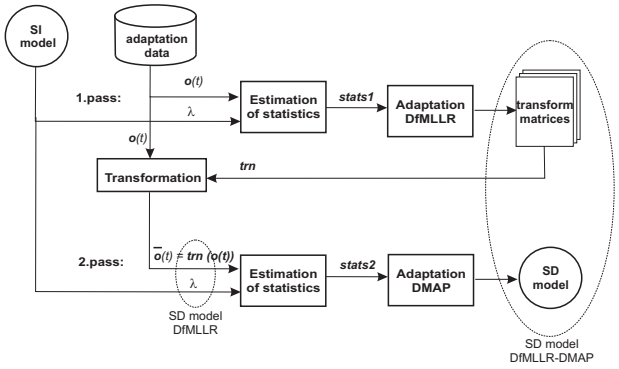

Figure 1: Two-pass-combination of DfMLLR and DMAP adaptation.

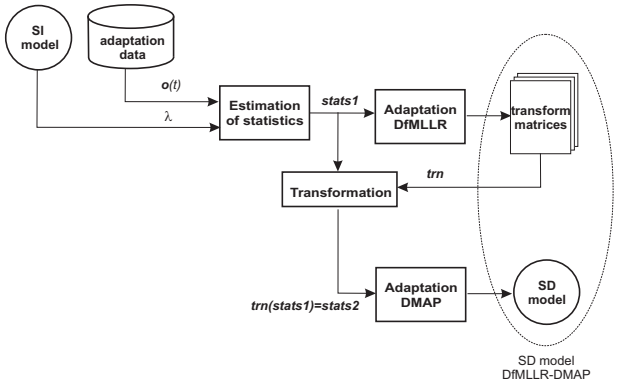

Figure 2: One-pass-combination of DfMLLR and DMAP adaptation.

6. Combination of DMAP and DfMLLR

Several efforts in combination of non-discriminative fMLLR and MAP methods were already examined in [11]. These two techniques are well suited for combination, mainly in order fMLLR succeeded by MAP.

The motivation can be stated as follows. Imagine a data stream continuously submitting feature vectors to the system. After a short while, when only a few vectors (several seconds of speech) were introduced, the fMLLR adaptation is performed. The feature vectors are continually submitted to the system. Hence, subsequently accumulated statistics should be more precise. After another while, when the amount of data has reasonably increased, fMLLR is applied again, and so on. This process continues till enough data were accumulated for MAP adaptation to be effective. Now, rather than apply the fMLLR adaptation MAP is utilized instead (MAP works well when lots of data are available).

Let us consider another situation when both adaptations are performed subsequently utilizing the same amount of data. The fMLLR pass shifts the whole model in the direction of acoustic space formed by adaptation data at once (even if lower amount of data is present). Thus, the data statistics for MAP pass become more precise. The statistics should now more properly describe the part of the acoustic space where the input data live. The next step, MAP adaptation, can be viewed as a refinement pass correcting parameters of mixtures (in relation to the amount of data aligned to these mixtures) utilizing more relevant (more properly aligned - fMLLR) statistics.

These are very natural behaviors of a combination of adaptation techniques. However, in order to perform both passes the time consumption increases significantly since the data statistics defined in Section 3 have to be accumulated twice (one per each pass). Note that experiments will be focused on the latter case (both adaptations performed subsequently utilizing the same amount of data).

Problem of time consumption is even more evident in case of discriminative methods where in addition the denominator statistics $\gamma_{jm}^{den}(t)$, c_{jm}^{den}, $\boldsymbol{\varepsilon}_{jm}^{den}(\boldsymbol{o})$, $\boldsymbol{\varepsilon}_{jm}^{den}(\boldsymbol{o}\boldsymbol{o}^{\mathrm{T}})$ have to be accumulated. The two-pass procedure is as follows (a more detailed schema can be found in Figure 1):

$$\begin{aligned} &1: \text{SI} \rightarrow stats_1 \text{ for SI} \overset{\text{DfMLLR}}{\Longrightarrow} \text{SD}_{\text{DfMLLR}} \\ &2: \text{SD}_{\text{DfMLLR}} \rightarrow stats_2 \text{ for SD}_{\text{DfMLLR}} \overset{\text{DMAP}}{\Longrightarrow} \text{SD}_{\text{DfMLLR+DMAP}} , \end{aligned} \tag{20}$$

where SI denotes the Speaker Independent model.

The evidence of an improvement in the systems performance can be found in Table 1. In order to join both adaptation passes into one, re-utilization of statistics computed in the first pass was proposed. Since DfMLLR is in use rather than accumulate statistics once again the already computed statistics are transformed to match the new feature space. Assuming feature transformation $\bar{\boldsymbol{o}}_t = \boldsymbol{A}_{(n)}\boldsymbol{o}_t + \boldsymbol{b}_{(n)}$ specified in Section 5 the statistics can be transformed as

$$\bar{\boldsymbol{\varepsilon}}_{jm}(\boldsymbol{o}) = \frac{\sum_{t=1}^{T}\gamma_{jm}(t)\bar{\boldsymbol{o}}_t}{\sum_{t=1}^{T}\gamma_{jm}(t)} = \boldsymbol{A}_{(n)}\boldsymbol{\varepsilon}_{jm} + \boldsymbol{b}_{(n)} , \tag{21}$$

$$\begin{aligned} &\bar{\boldsymbol{\varepsilon}}_{jm}(\boldsymbol{o}\boldsymbol{o}^{\mathrm{T}}) = \frac{\sum_{t=1}^{T}\gamma_{jm}(t)\bar{\boldsymbol{o}}_t\bar{\boldsymbol{o}}_t^{\mathrm{T}}}{\sum_{t=1}^{T}\gamma_{jm}(t)} = \\ &= \boldsymbol{A}_{(n)}\boldsymbol{\varepsilon}_{jm}(\boldsymbol{o}\boldsymbol{o}^{\mathrm{T}})\boldsymbol{A}_{(n)}^{\mathrm{T}} + 2\boldsymbol{A}_{(n)}\boldsymbol{\varepsilon}_{jm}(\boldsymbol{o})\boldsymbol{b}_{(n)}^{\mathrm{T}} + \boldsymbol{b}_{(n)}\boldsymbol{b}_{(n)}^{\mathrm{T}} , \end{aligned} \tag{22}$$

The transformed statistics are then utilized in the second DMAP pass (equation (8)). The only difference in one and two-pass approach consists in the use of SI mixtures' posterior $\gamma_{jm}(t)$, which remained untransformed. Thus, the one-pass combination can be expressed as (the schema is depicted in Figure 2)

$$\begin{aligned} &\text{SI} \rightarrow stats_1 \text{ for SI} \overset{\text{DfMLLR}}{\Longrightarrow} \text{SD}_{\text{DfMLLR}} \rightarrow \\ &\rightarrow \text{transform } stats_1 \overset{\text{DMAP}}{\Longrightarrow} \text{SD}_{\text{DfMLLR+DMAP}} . \end{aligned} \tag{23}$$

Hence, there is no need to see the adaptation data twice. Even if $\gamma_{jm}(t)$ remained unchanged the transformed statistic do not suffer from apparent inaccuracies as proved the experiments.

7. Experiments

7.1. Test Data

Experiments were carried out on the SpeechDat-East [12] corpus, which contains telephone speech in 5 languages Czech, Polish, Slovak, Hungarian, and Russian. For experiments only the Czech part of SD-E was used. The acoustic HMM was trained on 700 speakers with 50 sentences for each speaker (cca 5 sec. for sentence). For testing purposes 150 speakers were chosen with 50 sentences for each speaker, 10 sentences (cca 50 sec.) of each speaker were used for adaptation and the rest for testing (cca 200 sec.). The vocabulary consisted of 7000

	Acc[%]
SI model	55.85
MAP	62.42
fMLLR	65.03
DMAP	64.05
DfMLLR	65.81
two-pass-combination	
fMLLR+MAP	65.49
DfMLLR+DMAP	66.60
one-pass-combination	
fMLLR+MAP	65.30
DfMLLR+DMAP	66.38

Table 1: Accuracy (Acc)[%] of system performance for each type of adaptation and their combinations.

words. No OOV words were present. Triphones were modeled using 3 state HMM with 8 gaussian mixtures (diagonal covariances) in each of the states. For the recognition a language model based on zerograms was considered. In order to extract the features Perceptual Linear Prediction (PLP) was utilized [9], 12 dimensional feature vectors were extracted each 10 ms utilizing a 32 ms hamming window, Cepstral Mean Normalization (CMN) was applied, and Δ, $\Delta\Delta$ coefficients were added.

7.2. Adaptation Settings

In previous sections four adaptation methods were introduced with several parameters to be set. In the case of MAP and DMAP adaptation τ_{jm} was set for each mixture component to $\tau = 16$, for DMAP $f = 1$ was set in addition. In both methods adaptation of $\omega_{jm}, \boldsymbol{\mu}_{jm}$ and $\boldsymbol{C}_{jm}$ was assumed. In the case of fMLLR only one global transformation matrix for each speaker was utilized, and for DfMLLR f occuring in the smoothing term was set to 1. One iteration of all methods was performed.

7.3. Results

To demonstrate the influence of both passes (DfMLLR, DMAP) only smaller amount of data was used for adaptation (cca 50 sec. - see Section 7.1). The results of experiments are shown in Table 1. In the upper part of the table results obtained for the baseline system using SI model and non-combined adaptation methods can be found. The two-pass-combination of non-discriminative and discriminative techniques are located in the middle part of the table, and at the end results of the proposed on-pass-combination techniques can be examined. The results show obvious improvement of system performance when combining the MAP and fMLLR approach. Additional increase in accuracy is obtained for the discriminative alternative. The proposed one-pass-combination gives very similar results to the two-pass-combination with lower time consumption demands.

8. Conclusions

In this paper combination of discriminative adaptation techniques was examined. Comparison of non-discriminative and discriminative methods was carried out. MAP, DMAP, fMLLR, DfMLLR techniques were investigated. In order to decrease the time consumption of their combination an one-pass approach was proposed, where the accumulated statistics from the DfMLLR pass are transformed and handed to the DMAP pass. It was demonstrated that combination of adaptation techniques brings significant improvements into the speech recognition task. The discriminative techniques outperformed the non-discriminative ones, and the one-pass-combination shows to be most effective in the sense of time consumption preserving the systems accuracy (in relation to the two-pass-combination).

9. Acknowledgements

This research was supported by the Ministry of Education of the Czech Republic, project No. 2C06020 and project No. MŠMT LC536, and by the grant of The University of West Bohemia, project No. SGS-2010-054.

10. References

[1] Vank, J., Psutka, J., Zelinka, J., Prak, A., Psutka, J.: Discriminative training of gender-dependent acoustic models. In: Text, Speech and Dialogue, pp. 331-338, Berlin (2009).

[2] Chow,Y.L.:Maximum mutual information estimation of HMM parameters for continuous speech recognition using the N-best algorithm. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 701-704, Albuquerque (1990).

[3] Povey, D., Woodland, P.C.: Frame Discrimination Training Of HMMs For Large Vocabulary Speech Recognition. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 333-336 (1999).

[4] Wang, L., Woodland P.C.:Discriminative adaptive training using the MPE criterion. In: IEEE Automatic Speech Recognition and Understanding, pp. 279-284 (2003).

[5] Kai Yu: PhDThesis Adaptive Training for Large Vocabulary Continuous Speech Recognition. Hughes Hall College and Cambridge University Engineering Department (2006).

[6] Gauvain, L., Lee, C.H.: Maximum A-Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains. In: IEEE Transactions SAP, pp. 2:291–298 (1994).

[7] Gao, Y., Ramabhadran, B., Picheny, M.: New Adaptation Techniques for Large Vocabulary Continuous Speech Recognition. In:ICSA ITRW ASR, pp. 107-111, Paris (2000)

[8] Povey, D., Saon, G.: Feature and Model Space Speaker Adaptation with Full Covariance Gaussians. In: Interspeech, paper 2050-Tue2BuP.14 (2006).

[9] Psutka, J.: Robust PLP-Based Parameterization for ASR Systems. In: SPECOM 2007 Proceedings, pp. 509-515, Moscow (2007).

[10] Gales, M.J.F.: The Generation and use of Regression class Trees for MLLR Adaptation, Cambridge University Engineering Department (1996).

[11] Zajíc, Z., Machlica, L., Müller, L.: Refinement approach for adaptation based on combination of MAP and fMLLR. In: TSD, pp. 274-281, Pilsen (2009).

[12] Pollak P., et al.:,SpeechDat(E) - Eastern European Telephone Speech Databases, XLDB - Very Large Telephone Speech Databases, European Language Recources Association (ELRA), Paris (2000).