# The RU Submission to the Evalita'09 "application track" Speaker Recognition Evaluation

Marijn Huijbregts[1] and David van Leeuwen[1,2]
{m.huijbregts,d.vanleeuwen}@let.ru.nl

[1]Centre for Language and Speech Technology, Radboud University, the Netherlands
[2]TNO Human Factors, Soesterberg, The Netherlands

**Abstract.** In this paper the RU submissions for the application track of the Evalita'09 speaker recognition evaluation are presented. The primary submission is a fusion of two independently developed acoustic systems that include channel compensation and full joint factor analysis (JFA), respectively. Both systems are UBM-GMM dotscoring systems, which are characterized by very efficient computation of scores. Of the two systems the JFA approach performs best on its own. All components for this system are trained on the Evalita'09 training data, using the UBM data for UBM, channel and speaker factor traning, Z- and T-norm cohorts, and using the development test trials for calibration. The plain dotscoring system with channel compensation uses an externally trained UBM, and utilizes the Evalita'09 UBM training data for channel compensation training. On the evaluation trials, the fused system obtains equal error rates ranging from 13.45% (TC1, TS1) down to 1.64% (TC6, TS2).

**Key words:** Speaker identification, Evalita benchmark.

## 1 Introduction

In this paper we will describe our work towards the submissions for the application track of the Evalita'09 speaker identification benchmark. We have used the Evalita'09 benchmark to test two new systems. The first is a system is a UBM-GMM linear scoring approach with channel compensation, dubbed "dotscoring." The prior model of this system (the UBM) is trained on NIST data and the Evalita'09 UBM data is only used for channel compensation and T-normalization. The second system is based on Joint Factor Analysis (JFA), also employing linear scoring. It has been developed solely with the use of Evalita'09 data for training and tuning models. We refer to this system as the JFA system.

We submitted four sets of results. The first two submissions represent JFA and dotscoring systems alone, respectively. The third submission is the fusion of the two systems and for the final submission (our primary) we also fused a third system that is based on the JFA system but with the use of Principal

Component Analysis (PCA) on its feature vectors. Especially the fusion of the two baseline systems (dotscoring and JFA) turned out to be profitable.

In the following section we will describe the data sets used for the development of our systems. Next, in section 3, we will describe our four submissions. In section 4 we will list a number of our development experiments and in section 5 we will summarize our evaluation results. We will finish with a short discussion in section 6.

## 2 Data sets

The Evalita'09 training data consists of four sets of data: UBM data, enrollment data and development and evaluation data. Both for enrollment and evaluation, multiple tasks are defined based on channel variation and the amount of available data [2]. In the evaluation protocol it is defined which part of the enrollment and evaluation data could be used for each task.

For development experiments, only a set of 321 target trials were available. We constructed a set of non-target trials for development taking all test segments from the target trial list and combining these with other speaker models of the same gender, resulting in 9951 non-target trials.

The JFA system is developed solely using the Evalita'09 data. The dotscoring system also used a data set of 2257 speakers, collected from NIST SRE 2001, 2003, 2005, Switchboard II phase 2, Fisher English, and NIST LRE 2003 for the development of its UBM.

## 3 System description

We used two baseline systems, the dotscoring and the JFA system. We created a third system with the JFA system by using PCA on the feature vectors. Our primary submission is a fusion of these three systems. Next, these four submissions will be described in detail.

### 3.1 Dotscoring system

The dotscoring system is based on systems submitted by SUNSDV and TNO for the NIST SRE-2008 evaluation. Features, UBM and segment statistics were computed using a virtually identical set-up as for the TNO GMM-SVM NIST submission. However, rather than using an SVM with NAP [4] we implemented SUNSDV's approach for approximating the GMM likelihood function by a first order Taylor series expansion, which results in a scoring function which is the inner product of a model and test supervector.

**Features.** The feature extraction implementation is a slightly revised version of the TNO NIST 2008 system. PLP features are extracted from the speech signal, using a MEL scale filterbank and a cepstral representation of the features. We use 32 ms frames with a shift of 16 ms, and 12 PLP coefficients plus log energy are augmented with first order derivatives computer over 5 consecutive frames. Silence features are detected as having an energy 30 dB below the maximum frame energy in the segment, and are discarded. Short-time Gaussianization, or feature warping [9], is applied over a window of 255 samples, approximately 4 s.

**UBM.** The UBM is trained on a gender-balanced set of 2257 speakers, collected from NIST SRE 2001, 2003, 2005, Switchboard II phase 2, Fisher English, and NIST LRE 2003. The largest part contains English recordings over the telephone, but 520 speakers from 12 languages in LRE 2003 give a slight international touch to the UBM as a prior distribution of telephone speech.

The UBM is a mixture of 512 diagonal-covariance Gaussian components, which is rather small for a gender-independent model. Also, gender-independence approaches are uncommon to GMM scoring systems, which the dotscoring is. This is reminiscent of the front-end that was used for production of SVM supervectors in the GMM-SVM precursor of the system, and we had kept the UBM the same for reasons of comparison.

We have also experimented with gender-dependent UBMs trained on the EVALITA UBM data alone, but these did not make it into the final submission.

**The linearized likelihood function.** Here, we will describe the mathematics used in plain, uncompensated, dot-scoring in some more detail, because this is not readily found in literature [11, 6]. This approach has been introduced by Niko Brümmer for NIST SRE 2005, and has been revived in NIST SRE 2008 by extending it to include channel compensation techniques, and at the JHU workshop in the summer of 2008 [6], which resulted in the JFA toolbox.

Following SDV's NIST SRE 2008 approach [11], we linearize the log-likelihood function $\log p(\boldsymbol{x}(t)|\boldsymbol{\lambda}_s)$ by seeing the speaker specific model parameters $\boldsymbol{\lambda}_s$ as perturbations of the UBM $\boldsymbol{\lambda}_U$. Here $\boldsymbol{x}(t)$ are the speech features at time $t$, and the UBM is a Gaussian Mixture Model with diagonal covariances. Formally,

$$\log p(\boldsymbol{x}(t)|\boldsymbol{\lambda}_s) \approx \log p(\boldsymbol{x}|\boldsymbol{\lambda}_U) + (\boldsymbol{\lambda}_s - \boldsymbol{\lambda}_U) \cdot \nabla_\lambda \log p(\boldsymbol{x}(t)|\boldsymbol{\lambda}) \Big|^{\boldsymbol{\lambda}_U} . \tag{1}$$

In scoring a test segment $\{\boldsymbol{x}_t\}$ we look at the difference in log likelihoods of UBM and speaker model, so we are left only with the second term in (1). The speaker models are going to be found by MAP adaptation of the UBM means, hence only the mean vectors $\boldsymbol{\mu}_k$ for GMM components $k$ contribute to the perturbation. Writing the components of $\boldsymbol{\mu}_k$ as $\mu_{kd}$ the second term can be written as

$$\sum_{kd} (\mu^s_{kd} - \mu^U_{kd}) \frac{\partial}{\partial \mu_{kd}} \log p(\boldsymbol{x}(t)|\boldsymbol{\lambda}) \Big|^{\boldsymbol{\lambda}_U} . \tag{2}$$

The derivative of the log likelihood function expands to

$$\frac{\partial}{\partial \lambda_{kd}} \log p(\boldsymbol{x}(t)|\boldsymbol{\lambda}) \Big|^{\boldsymbol{\lambda}_U} = \frac{p(\boldsymbol{x}(t), k)}{p(\boldsymbol{x}(t)|\boldsymbol{\lambda})} \frac{x_d(t) - \mu_{kd}^U}{\sigma_{kd}^{U\,2}} \tag{3}$$

$$= P\big(k|\boldsymbol{x}(t)\big) \frac{x_d(t) - \mu_{kd}^U}{\sigma_{kd}^U} \frac{1}{\sigma_{kd}^U} \tag{4}$$

where $p(\boldsymbol{x}, k)$ is the joint probability of $\boldsymbol{x}(t)$ and $k$, or the likelihood of $\boldsymbol{x}$ contributed by Gaussian $k$. In the last line we used the sum and product rules of probability to write the first factor as $P\big(k|\boldsymbol{x}(t)\big)$, the posterior of UBM component $k$ given frame $\boldsymbol{x}(t)$

$$P\big(k|\boldsymbol{x}(t)\big) = \frac{p(\boldsymbol{x}(t), k)}{\sum_{k'} p(\boldsymbol{x}(t), k')}. \tag{5}$$

As noted above, the perturbation components in (1) are formed by only the shift in means of the UBM, i.e., $\mu_{kd}^s - \mu_{kd}^U$. These can be obtained, for a training segment $\{\boldsymbol{y}(t)\}$, using "classical MAP" [5, 10]

$$\mu_{kd}^s - \mu_{kd}^U = \frac{1}{n_k + r} \sum_t P\big(k|\boldsymbol{y}(t)\big)(y_d(t) - \mu_{kd}^U) \tag{6}$$

where $r$ is the *relevance factor*, and $n_k = \sum_t P\big(k|\boldsymbol{y}(t)\big)$ are the *zeroth order statistics* of the training sample $\{\boldsymbol{y}(t)\}$. By scaling down the shift of the means in (6) by the respective standard deviations $\sigma_{kd}$ of the Gaussian component $k$, the terms in the summation become similar in appearance to the RHS of (4), and these are known as the *first order statistics* $f_{kd}^y$ of frame sequence $\{\boldsymbol{y}(t)\}$:

$$f_{kd}^y = \sum_t P\big(k|\boldsymbol{y}(t)\big) \frac{y_d(t) - \mu_{kd}^U}{\sigma_{kd}^U} \tag{7}$$

Combining (4), (6) and our definition of the first order statistics, the second term in (1) integrated over all frames of the test segment $\{\boldsymbol{x}(t)\}$ becomes the speaker recognition score

$$s(\{\boldsymbol{x}\}, \{\boldsymbol{y}\}) = \sum_t \sum_{kd} (\mu_{kd}^s - \mu_{kd}^U) \frac{\partial}{\partial \mu_{kd}} \log p(\boldsymbol{x}(t)|\lambda_U) \Big|^{\lambda_U} \tag{8}$$

$$= \sum_{kd} \Big(\frac{1}{n_k + r} f_{kd}^y\Big) f_{kd}^x \tag{9}$$

$$= \sum_{i=k\otimes d} m_i(\{\boldsymbol{y}\}) f_i(\{\boldsymbol{x}\}) = \boldsymbol{m} \cdot \boldsymbol{f} \tag{10}$$

In the last step, we have defined the *model* supervector $m_{kd} = f_{kd}/(n_k + r)$, and combined the summation over Gaussian component $k$ and the feature dimension $d$ to a single index $i$, so that $m_i$ and $f_i$ form *supervectors*, and scoring has become as simple as taking the inner product of a model supervector and a test segment supervector. Hence the name dotscoring.

**Speech segment statistics.** From the previous section, it may have become clear that a speech segment with features $\{\boldsymbol{x}(t)\}$ is characterized completely by the zeroth and first order statistics computed using $P\big(k|\boldsymbol{x}(t)\big)$, the posterior probabilities of component $k$ given speech frame $\boldsymbol{x}(t)$ and the UBM. After generating these statistics, as has been noted in [11], the UBM does not play a role anymore in scoring or model training. Note that computation of the models and scores can be performed readily in matrix-computation packages such as Octave[1]. There are two ways of computing the outer product of indices $k$ and $d$ to obtain supervector index $i$, and in collaboration with other researchers one typically finds that a different choice has been made. Further, notice that the zeroth order statistics $n_k$ do not depend on the feature dimension $d$, so that for efficient matrix calculation this has to be replicated in the right dimension before flattening the matrix to a supervector.

**Channel compensation.** The statistics $f_{kd}$ for each speech segment can be compensated for the channel/session, following [11]. For this we computed a channel compensation matrix on all EVALITA 2009 UBM speech, by finding the directions in which supervector statistics vary most between sessions of the same speaker. We used both male and female speech in the UBM, and used the 40 directions in supervector space to model the variability due to session and channel. These directions were found by doing principal component analysis on the deviations of all segment statistics from the per-speaker-mean. We used the techniques explained in [3] to make the eigenvalue problem feasible.

All training and test segment statistics were compensated by the channel compensation matrix. Models were trained using a relevance factor of 1.

**T-norm.** The UBM speakers were also used to build models for T-norming [1]. We built different T-norm models for GSM and PSTN data for each UBM speaker and included both in the T-norm cohort. We further employed T-norming conditioned on gender. This appears the only point in the dotscoring system where we treat male speech differently from female speech.

Despite various reportings that Z-norm is important in GMM-scoring, including dotscoring, we did not find an improvement in development testing.

### 3.2 The JFA system

Our Joint Factor Analysis (JFA) based submission consists of Speech Activity Detection (SAD), Universal Background Model (UBM) generation, and JFA itself. For SAD and UBM generation we used the Shout toolkit [7] and we used the JFA cookbook[2] developed by Ondrej Glembek at Brno University of Technology and based on [8] for the joint factor analysis. We further applied ZT-normalization.

[1] http://www.gnu.org/software/octave/

[2] http://speech.fit.vutbr.cz/en/software/joint-factor-analysis-matlab-demo

**Speech Activity Detection** is done using a straightforward energy-based approach. Energy is calculated with an interval of 10 ms for windows of 32 ms audio. If 10 or more successive feature frames are above a threshold, those frames are marked as speech. The threshold is set automatically at the mean energy of the recording minus the variance of the energy frames. At least another 10 successive frames must be below the threshold before the frames are marked as silence.

**UBM generation** is performed on features with 12 Mel-Frequency Cepstral Coefficients (MFCC), energy and the deltas of these thirteen coefficients. For each gender, a UBM is trained using all available Evalita'09 UBM data. The UBM (a Gaussian Mixture Model (GMM)) is incrementally trained up to 1024 Gaussians, doubling the number of Gaussians at each iteration. After training the gender dependent UBMs, for both UBMs the zeroth and first order sufficient statistics are calculated for all Evalita'09 audio segments (UBM-, train-, development- and evaluation data). These statistics are used by the JFA component for training of the eigenchannels, eigenvoices and residuals, for enrollment of the speakers (both training and T-norm) and for processing the trials.

**Training the eigenvoices, eigenchannels and residuals.** Similar to UBM generation, JFA is performed gender dependently. The first step in JFA is to determine the eigenvoices, eigenchannels and residual matrices (respectively denoted with $v$, $u$ and $d$). For training of the eigenvoices and the residuals, the Evalita'09 UBM set is split into two. For each gender, the first 25 speakers (with ID 1 to 25) are used for training the eigenvoices and the remaining 5 speakers are used to train the residuals. The eigenchannels are trained on the entire Evalita'09 UBM data set. We used 50 eigenchannel factors and 20 eigenvoice factors. Because of the relatively small UBM data set, using more than 20 eigenvoice factors did not improve the system results on the development trial set.

**Speaker enrollment and ZT-norm.** For T-norm, all available Evalita'09 UBM speakers are used and for Z-norm, all speech segments of all UBM speakers are selected. The JFA cookbook is used to enroll the target speakers and the T-norm speakers (to determine the speaker factors $y$ and $z$). The Z-norm data is used to calculate the distribution of non-target scores for each speaker. For each T-norm speaker, the Z-norm data coming from the speaker himself is not used during this calculation. In contrast to our dotscoring submission, for most test conditions of our development set we did measure a small EER improvement when using ZT-normalization instead of T-normalization.

### 3.3 JFA-PCA system

For this submission, the JFA system is used with an adjustment to the feature extraction. The delta-deltas are added to the feature vectors and Principal Component Analysis (PCA) is performed on the Evalita'09 UBM data (males only, because of time constraints). The first 26 components are used to generate the new features for this system. This system did not outperform our baseline JFA system on our development set, but using it during fusion did improve the results slightly.

### 3.4 Fusing the systems, the primary submission

Our primary submission is a fusion of the three systems described above. For fusion we used Niko Brümmers FoCal package[3]. We tuned the fusion on the Evalita'09 development set. Note that because we almost didn't finish the JFA-PCA system in time, we also submitted a fusion of our first two systems.

## 4 Development experiments

In this section we will describe a number of our development experiments. First we will list some experiments we performed to determine the optimal dimensionality of the feature vectors for our JFA system. Next, we will describe experiments on the use of T-norm, Z-norm and ZT-normalization.

### 4.1 Principal component analysis

Our JFA system uses features with a dimensionality of 26 (12 MFCCs, energy and the delta's). We found in an orienting experiment that adding the double delta's (dimensionality of 39) did not improve the system. This lack of improvement might be because the Evalita'09 data set is not big enough to train models for high dimensionality features. To test this hypothesis we performed PCA on the UBM data in order to reduce the dimensionality.

We created two systems to perform these experiments. The first system reduced the 39 dimensions back to 26 (PCA-26) and the second system reduced the dimensionality further to 20 (PCA-20). The results of the experiments are listed in table 1. Unfortunately, the two PCA based systems did not outperform our JFA system. As can be seen in the next section, it did help to use the PCA-26 system for fusion.

### 4.2 Normalization

We tested T-norm, Z-norm and ZT-norm on both the dotscoring and the JFA system. For the JFA system, all types of normalization improved the performance. On the TC6-TS2 task (the easiest task), without normalization the EER is 2.49%. With Z-norm the EER is 2.14%, with T-norm 1.55% and with ZT-norm 1.24%. On the other train- and test conditions, the results were similar. In contrast with these findings for the JFA system, as noted in section 3.1, for the dotscoring system, Z-norm does not improve the results at all.

[3] http://niko.brummer.googlepages.com/focalbinary

**Table 1.** The equal error rates on our development set of the JFA system and the two PCA systems.

| Train | Test | %EER JFA | %EER PCA-26 | %EER PCA-20 |
|---|---|---|---|---|
| TC1 | TS1 | 13.40 | 16.13 | 25.23 |
| TC2 | | 16.49 | 19.31 | 26.17 |
| TC3 | | 10.22 | 13.08 | 20.49 |
| TC4 | | 13.40 | 14.95 | 21.81 |
| TC5 | | 8.74 | 11.21 | 19.64 |
| TC6 | | 5.30 | 7.14 | 14.95 |
| TC1 | TS2 | 8.72 | 11.21 | 16.76 |
| TC2 | | 8.72 | 11.84 | 15.26 |
| TC3 | | 7.79 | 8.99 | 13.71 |
| TC4 | | 5.24 | 8.41 | 10.59 |
| TC5 | | 2.78 | 5.03 | 10.65 |
| TC6 | | 1.25 | 2.23 | 6.53 |

### 4.3 NIST or Evalita'09 UBM

For the dotscoring system, we carried out a small test on the type of UBM trained. We compared the baseline UBM trained for NIST SRE 2008 with a UBM trained on Evalita'09 UBM data alone, both gender-dependent and gender-independent, keeping all other components constant (except using a gender-dependent channel compensation for gender-dependent UBM training). In table 2 the development test results for the easiest task, TC6-TS2 are shown, separated for male and female trials as well as overall.

**Table 2.** Comparison of UBM trained on NIST and Evalita'09 data, for the dotscoring system on the devolpment set. Numbers represent EER, in %

| UBM training | all | female | male |
|---|---|---|---|
| NIST (system 2) | 3.17 | 2.48 | 3.12 |
| Evalita'09 (gender independent) | 4.97 | 4.97 | 4.31 |
| Evalita'09 female | | 1.95 | 11.88 |
| Evalita'09 male | | 8.57 | 2.54 |

## 5 Evaluation results

Figure 1 contains the DET plots averaged over all train and test conditions of our JFA submission (system 1), our dotscoring submission (system 2) and our primary fused submission (system 4). Fusion of the two systems was successful for all test and training conditions. The fusion might work that well because the systems were developed completely independently. The systems do not share a single component.

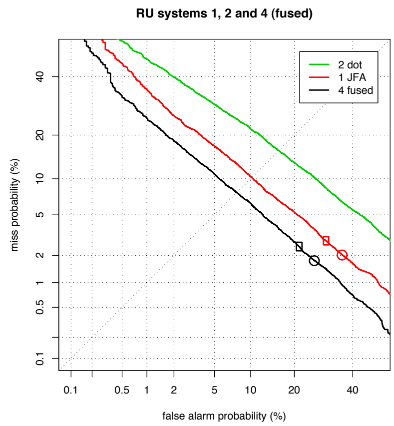


**Fig. 1.** The DET curves of the three submissions, averaged over all train and test conditions.

Table 3 shows the most important evaluation metrics on the primary submission (the fused system) for all train and test conditions and figure 2 contains the DET plots of the primary submission. The graph to the left contains all trials of the TS1 test condition and the graph to the right contains all trials of the TS2 test condition.

**Table 3.** The results of the primary system (system 4) for all train and test conditions.

| Train | Test | Cllr | Cllr.min | %EER | Cdet | Cdet.min |
|---|---|---|---|---|---|---|
| all | all | 0.290 | 0.2713 | 7.74 | 0.2144 | 0.03459 |
| TC1 | TS1 | 0.456 | 0.4339 | 13.45 | 0.4404 | 0.05126 |
| TC2 | | 0.468 | 0.4321 | 13.54 | 0.3870 | 0.05657 |
| TC3 | | 0.348 | 0.3316 | 10.10 | 0.3930 | 0.03919 |
| TC4 | | 0.375 | 0.3368 | 10.06 | 0.2974 | 0.03935 |
| TC5 | | 0.332 | 0.2499 | 7.27 | 0.1273 | 0.03511 |
| TC6 | | 0.242 | 0.1868 | 5.73 | 0.0933 | 0.02595 |
| TC1 | TS2 | 0.266 | 0.2503 | 7.15 | 0.2729 | 0.03101 |
| TC2 | | 0.306 | 0.2502 | 7.28 | 0.1456 | 0.03264 |
| TC3 | | 0.183 | 0.1653 | 4.54 | 0.2174 | 0.02029 |
| TC4 | | 0.220 | 0.1879 | 5.38 | 0.1246 | 0.02498 |
| TC5 | | 0.153 | 0.1099 | 3.15 | 0.0559 | 0.01565 |
| TC6 | | 0.129 | 0.0616 | 1.64 | 0.0180 | 0.00953 |

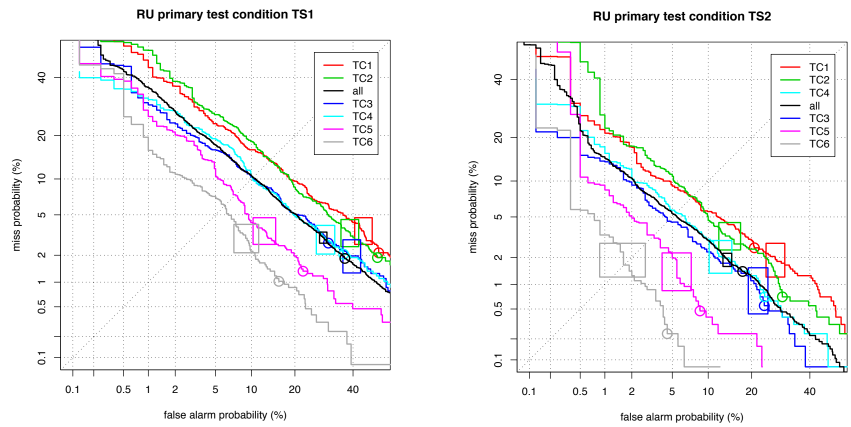


**Fig. 2.** The DET curves of our primary submission on the TS1 (left) and TS2 (right) evaluation sets.

## 6 Discussion

In this paper we have described our work towards the submissions for the application track of the Evalita'09 speaker identification benchmark. We submitted one system that was developed using mainly NIST data (dotscoring system) while the other system was developed using solely the data provided for the Evalita'09 evaluation.

As one might expect because the two systems are so different, fusion of the systems worked really well. The gain of fusing these two systems was much bigger than the modest improvement we got from adding the third system (based on the JFA system, JFA-PCA) to the fusion.

It is also interesting to note that even though the Evalita'09 data set is relatively small, it is possible to build a well performing system with it. This suggests it is better to use a well balanced data set that is as similar to the evaluation data as possible, than to use a huge data set that does not resemble the evaluation that well.

We compared the Evalita'09 UBM training with NIST UBM training for the simpler dotscoring system, see table 2. There we found that a gender-dependent Evalita'09 UBM performed somewhat better than the the gender-independent NIST UBM, but still not as well as the JFA system. Making the Evalita'09 gender-independent makes it perform a bit worse than the NIST UBM—but this may be an effect of tuning which had happened for the NIST UBM system.

There are several differences between the JFA and dotscoring systems, among which are speaker factors (missing from the dotscoring system), a different training of channel factors, Z-norming (not used in dotscoring), gender-dependence in the design, and UBM training. In order to do a full comparison of all system components, more research needs to be carried out, but one of the more interesting experiments would be to evaluate the JFA system with an UBM trained on the NIST data similar to the dotscoring system.

**Acknowledgments.** This paper is based on research carried out for the BATS project within the research program IM-Pact funded by IBBT and ICTRegie. We would like to thank Ondrej Glembek for his great JFA cookbook and Niko Brümmer for his calibration and fusion toolkit FoCal.

## References

1. Auckenthaler, R., Carey, M., Lloyd-Thomas, H.: Score normalization for text-independent speaker verification systems. Digital Signal Processing, vol. 10, pp. 42–54 (2000)
2. Aversano, G.: Speaker Identity Verification - Application Track Task Guidelines, http://evalita.fbk.eu/speaker.html
3. Brümmer, N., Burget, L., Černocký, J., Glembek, O., Grezl, F., Karafiát, M., Matějka, P., van Leeuwen, D.A., Schwarz, P., Strassheim, A.: Fusion of heterogeneous speaker recognition systems in the STBU submission for the nist speaker recognition evaluation 2006. IEEE Transactions on Speech, Audio and Language Processing, vol. 15, issue 7, pp. 2072–2084 (2007)
4. Campbell, W., Sturim, D., Reynolds, D., Solomonoff, A.: SVM based speaker verification using a GMM supervector kernel and NAP variability compensation. In: Proceedings of ICASSP, pp. 97–100 (2006)
5. Gauvain, J.L., Lamel, L.F., Adda, G., Adda-Decker, M.: Speaker-independent continuous speech dictation. Speech Communication archive, vol. 15, pp. 21–37 (1994)
6. Glembek, O., Burget, L., Dehak, N., Brümmer, N., Kenny, P.: Comparison of scoring methods used in speaker recognition with joint factor analysis. In: Proceedings of ICASSP 2009. Taipei, Taiwan (2009)
7. Huijbregts, M.: Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled. PhD thesis, University of Twente, Enschede (2008)
8. Kenny, P., Ouellet, P., Dehak, N., Gupta, V., Dumouchel, P.: A study of inter-speaker variability in speaker verification. IEEE Transactions on Audio, Speech and Language Processing, vol. 16, issue 5 (2008)
9. Jason Pelecanos and Sridha Sridharan. Feature warping for robust speaker verification. In: Proceedings of Speaker Odyssey, pp. 213–218. Crete, Greece (2001)
10. 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)
11. Albert Strasheim and Niko Brümmer. Sunsdv system description: NIST sre 2008. In: Proceedings of NIST SRE workshop (2008)