16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

MISSING FEATURE RECONSTRUCTION AND ACOUSTIC MODEL ADAPTATION COMBINED FOR LARGE VOCABULARY CONTINUOUS SPEECH RECOGNITION

Ulpu Remes, Kalle J. Palomäki, and Mikko Kurimo

Adaptive Informatics Research Centre, Helsinki University of Technology
P.O. Box 5400, FIN-02015 HUT, Finland
phone: + (358) 94513276, fax: + (358) 94513277, email: firstname.lastname@hut.fi
web: www.cis.hut.fi/projects/speech/

ABSTRACT

Methods for noise robust speech recognition are often evaluated in small vocabulary speech recognition tasks. In this work, we use missing feature reconstruction for noise compensation in large vocabulary continuous speech recognition task with speech data recorded in noisy environments such as cafeterias. In addition, we combine missing feature reconstruction with constrained maximum likelihood linear regression (CMLLR) acoustic model adaptation and propose a new method for finding noise corrupted speech components for the missing feature approach. Using missing feature reconstruction on noisy speech is found to improve the speech recognition performance significantly. The relative error reduction 36 % compared to the baseline is comparable to error reductions introduced with acoustic model adaptation, and results further improve when reconstruction and adaptation are used in parallel.

1. INTRODUCTION

Large vocabulary continuous speech recognition has reached reasonable performance levels in controlled environments, but real environments with changing and unpredictable noise conditions remain still a challenge even with the current noise compensation methods. It is noteworthy that methods designed to improve statistical robustness in general rather than compensate for noise are often superior in performance. Such methods include e.g. acoustic model adaptation with constrained maximum likelihood linear regression (CMLLR) [1].

In this work, the noise robustness issue is addressed with the missing feature methods as proposed in [2][3]. Applying missing feature methods for noise compensation in automatic speech recognition is based on the observation that additive noise affects some spectrotemporal regions in the speech signal more than others. Thus, while the noise corrupted parts are unreliable and should not be conventionally used in speech recognition, the less affected speech components are somewhat reliable and (i) may be utilised in the usual manner in speech recognition and (ii) provide information about the unreliable (missing) components. Motivation for the missing feature approach originally comes from the human speech perception and auditory scene analysis (ASA) [4][5].

The missing feature methods have performed well under various noise conditions and several different methods have been proposed for handling the unreliable spectrotemporal components [2][3][5]. The methods have mostly been tested with a limited vocabulary, and are yet to become popular in large vocabulary continuous speech recognition (LVCSR) systems. For earlier results on missing feature techniques in LVCSR task, see for example [6], where missing feature reconstruction is used with projected spectral features and evaluated on AURORA 4 database that contains speech with artificially added real-world noises. In this work, we use Finnish large vocabulary speech data recorded in noisy environments such as parks and cafeterias. For noise compensation, we use cluster-based missing feature reconstruction method proposed in [3], which we combine with CMLLR acoustic model adaptation. Reconstruction and adaptation are used in series, with adaptation following reconstruction in the same system, and in parallel, using linear weighting on the system outputs in log-likelihood level as proposed in [7]. According to our knowledge, no results have been published before where a missing feature method would have been used together with acoustic model adaptation.

The performance of any missing feature method depends heavily on the accuracy of the spectrographic mask that partitions the speech signal into the reliable and unreliable regions. We propose a new method for detecting the unreliable speech components. In this method, a noise estimate is calculated from speech pauses detected using a Gaussian mixture based speech/non-speech classifier. The noise estimate is compared to the noisy speech signal to find the unreliable, noise dominated regions. This method is more suitable for on-line applications and changing noise conditions than the commonly used approach where the noise estimate is calculated from a fixed number of frames at the beginning of each audio file, e.g. [2]. Other mask estimation methods suited for on-line applications have been proposed in e.g. [3][6][8].

2. METHODS

2.1 Baseline system

Our large vocabulary continuous speech recognition system uses a morph-based growing n-gram language model [9] which is trained on book and newspaper data. The text data contains 145 million words. Since all words and word forms can be represented with the unsupervised morphs, the decoding vocabulary is in practise unlimited [10]. The decoder used is a time-synchronous beam-pruned Viterbi token-pass system [11]. The acoustic models are state-clustered hidden Markov triphone models constructed with a decision-tree method [12]. Each state is modelled with 16 Gaussians, and the states are also associated with gamma probability functions to model the state durations [13]. Speech is represented with 12 MFCC and a log energy feature. Features are used with their first and second order differentials, and they are treated with cepstral mean subtraction (CMS) and max-

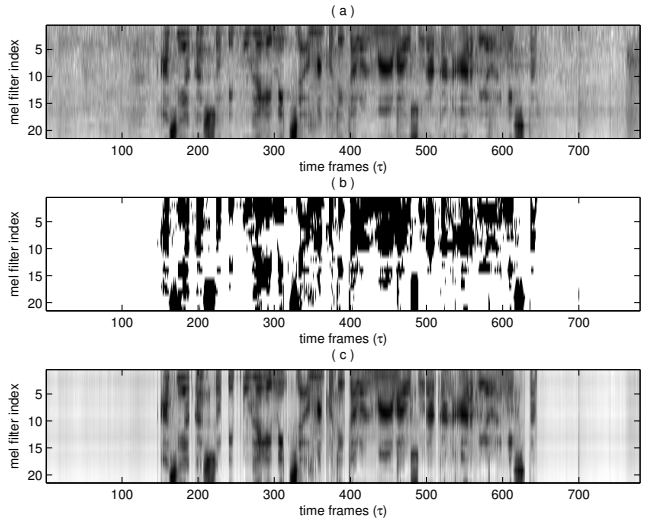

Figure 1: (a) Logarithmic mel spectrogram for an utterance recorded under public place noise conditions and (b) a spectrographic mask that divides the spectrogram to reliable (black) and unreliable (white) regions. (c) An estimate for the clean spectrogram constructed with the cluster-based missing feature reconstruction method [3].

imum likelihood linear transformation (MLLT) [14] learned in training. The training of the acoustic models is described in Section 3.

2.2 Missing feature reconstruction

2.2.1 Noise mask estimation

In most real world acoustic environments, it is reasonable to assume that the most notable noise components originate from sources that are uncorrelated with speech. Uncorrelated noise corrupts speech additively in power spectral domain. Thus, in logarithmic mel spectral domain, when speech dominates over noise, the time-frequency components $Y(\tau,i)$ of noisy speech signal may be considered as reliable estimates of the clean speech values $X(\tau,i)$ that would have been observed if there had not been any noise. The components in the noise dominated regions are, on the other hand, unreliable, and provide only an upper bound for the true values. Labels that divide the data to reliable and unreliable parts are referred to as the spectrographic mask. Partition to reliable and unreliable time-frequency regions is illustrated in Figure 1.

In this work, we propose a new on-line method for spectrographic mask estimation. The method is based on speech pause detection. In this method, the spectrographic mask is constructed based on local signal-to-noise ratio (SNR) estimates. The estimates are derived from noise estimates calculated during speech pauses which are detected using a Gaussian mixture based speech/non-speech classifier. The non-speech frames $N(\tau)$ i.e. the frames $Y(\tau)$ that have been classified as non-speech are collected and temporally smoothed to produce the noise estimate. For non-speech frames, the noise estimate is calculated as

$$N_{ave}(\tau,i) = \sum_k h(k)N(\tau-k,i), \qquad (1)$$

where τ indicates the time-frame and i the frequency channel. The temporal smoothing window is $h(k) = \beta(40-|k|)$, $k = -10\ldots40$, where β is constant chosen so that the window gain is 1. This yields a weighted moving average where the current time instant is given maximal weight while past and future are linearly (in log domain) attenuated. The shape of the window is experimentally chosen. For speech frames, the last noise estimate $N_{ave}(\tau)$ calculated before the speech onset is used. Time-frequency components are taken to be unreliable if their observed value $Y(\tau,i)$ does not exceed the estimated noise power $N_{ave}(\tau,i)$ with minimum of θ dB, where θ is manually optimised SNR threshold. In this work, the threshold is optimised with far recorded parameter optimisation data (see Section 3 for dataset description) and the threshold θ = 3 dB.

2.2.2 Speech pause detection

The speech/non-speech classifier used in this work is a hidden Markov model (HMM) classifier where speech and non-speech are modelled as single states with 24 Gaussian components. Insertion penalties are used in decoding to exclude short speech or non-speech segments. The classifier uses the same features as the speech recognition system described in Section 2.1. The classifier training data contains over 5 hours of television news data from the Finnish Broadcasting Company (YLE). It has precise hand-annotated time marks for speech and other audio. Since the classifier is used for noise mask estimation, it is crucial that it does not misclassify any speech segments as non-speech. To avoid this, the speech model probabilities are multiplied by two.

2.2.3 Cluster-based reconstruction

The missing feature methods used in automatic speech recognition are commonly divided in two categories, classifier modification and data imputation approaches. Bounded marginalisation, for example, has been found efficient when tested with a limited vocabulary (e.g. in a connected digit recognition task), but as a classifier-modification method it limits the speech recogniser to use spectral features [2]. In HMM-based speech recognition systems, cepstral features are preferred [15]. In data imputation methods, the unreliable components are replaced with estimates that correspond to clean speech so that the reconstructed spectral features may be further processed as usual and the recogniser needs not be modified. This is especially important in large vocabulary continuous speech recognition where the state-of-the-art systems typically use various normalisation methods and feature transformations.

In this work, we use data imputation with the cluster-based feature reconstruction method proposed in [3]. Here, the log spectral feature vectors $Y(\tau)$ are assumed independent and identically distributed, and the unreliable spectral components in $Y(\tau)$ are reconstructed based on their statistical relation to the other components in the same vector. The clean log spectral features $X(\tau)$ are assumed to originate from a Gaussian distribution, but not necessarily all from the same one; the feature distribution model is estimated from clean training data so that the data is divided to clusters, and each cluster is modelled with a Gaussian distribution. For a description on how the unreliable features are reconstructed with the distribution model, see [3].

The feature distribution model used in this work is a 5-

component GMM trained with 96 minutes of clean speech extracted from the SPEECON training set described in Section 3. Using the whole training set did not improve the reconstruction results in preliminary tests conducted with our parameter optimisation dataset. The clusters and distribution parameters were jointly estimated using the expectation-maximisation (EM) algorithm from the GMMBAYES Matlab Toolbox [16].

2.3 Adaptation and reconstruction combined

Regression based model adaptation methods are a common choice for speaker as well as environmental adaptation when hidden Markov model (HMM) based acoustic models with states modelled as Gaussian mixture distributions are used. Variation in environmental conditions affects the feature distribution, so with a new speaker or in a new environment, the distribution becomes mismatched with the acoustic models unless the models are adapted. Now, although the constrained maximum likelihood linear regression (CMLLR) is essentially a model adaptation method, it can be formulated as a linear transformation on features [1]. The transformation parameters are estimated from adaptation data so that the transformed models maximise the likelihood of the data.

In this work, since the noise corrupted features are reconstructed to appear as equivalent to the clean speech features, CMLLR transformations can be estimated based on the reconstructed features and applied to the features in the usual manner. This will be referred to as using the methods in series. The methods are also used in parallel, embedded in separate systems that use the same acoustic models and receive the same input. The system outputs are combined using linear weighting as proposed in [7]. Acoustic model outputs are the feature log-likelihoods $P(\mathbf{O}|S)$, where $\mathbf{O} = \{\mathbf{o}(\tau)\}$ is the feature representation for the input speech and $S = \{s\}$ denotes the states in the HMM-based acoustic model. The combined output log-likelihoods are calculated as

$$P = \alpha P(\widehat{\mathbf{O}}|S) + (1-\alpha) P(\widetilde{\mathbf{O}}|S) \qquad (2)$$

where $\widehat{\mathbf{O}}$ are the adapted features and $\widetilde{\mathbf{O}}$ the reconstructed features, and $\alpha \in [0,1]$ is the weight that determines whether the system with adaptation or the system with reconstruction should be emphasised in the log-likelihood calculation. In this work, the weight parameter is manually optimised using far recorded parameter optimisation data (see Section 3 for dataset description) and $\alpha = 0.7$. The log-likelihoods are sent to the decoder where they are combined with language model probabilities to produce the final hypothesis (speech recognition result).

3. DATA

The acoustic models are trained with data selected from the Finnish SPEECON database [17]. The training dataset contains 26 hours of clean speech recorded with a close-talk microphone. The data is collected from 208 speakers with both male and female speakers. Among utterances are words, sentences and free speech.

Parameter optimisation and evaluation data are also from the SPEECON database. The 72-minute parameter optimisation set has speech from 23 speakers, and the 101-minute evaluation set has been collected from 32 speakers. The evaluation set does not share speakers with the parameter optimisation set or the acoustic model training data. In addition, 105 sentences selected from the parameter optimisation set were hand-annotated for evaluating the speech/non-speech classifier performance. The utterances used for parameter optimisation and evaluation are all read sentences recorded in public places both indoors and outdoors where speech, footsteps, unspecified clatter etc. appears in the background. The sentences are excerpts from Internet texts and occupy a large (unlimited) vocabulary.

The proposed methods are tested under two different conditions: we use data recorded with the close-talk microphone and data recorded from 0.5 m–1 m distance. The recordings have been made simultaneously so they have the same speech contents. SNR values estimated with the recording platform are on average 24 dB in the close-talk data and around 9 dB in the far recorded data. The close-talk data almost corresponds to clean speech, whereas in the far recorded data, both environmental noise and reverberation (in some environments) affect the speech signal and decrease speech recognition performance. The data is recorded in sessions with one speaker in certain environment. Each session is around 3 minutes. When CMLLR is applied, the transformations are estimated based on the sessions, and the system is adapted to both the speaker and the environment. In this work, we use offline adaptation with all available data first used as adaptation data and then recognised.

Since Finnish speech data is used, speech recognition performance is measured primarily with the letter error rate (LER). For other languages, the word error rate (WER) is more common, but it is not well applicable to Finnish where words are long. Finnish words are often concatenations of several morphemes and correspond to more than one English words like the word 'kahvin+juoja+lle+kin' which translates to 'also for a coffee drinker'. In this work, both error rates are reported, but system comparisons are based on the letter error rates alone.

4. RESULTS

4.1 Speech/non-speech classifier

To evaluate the speech/non-speech classifier, we compared the classification results to hand-annotated speech and non-speech regions. The frame classification accuracy was 93 % for the close-talk data and 92 % for the far recorded data. In addition, we tested how speech/non-speech classification affects the speech recognition performance when the classification results are used for spectrographic mask estimation for the cluster-based missing feature reconstruction. The spectrographic mask estimates were constructed either based on automatically classified or manually set speech/non-speech segments. SNR threshold θ was optimised for both segmentations with the far recorded data, and in this experiment, the threshold $\theta = 3$ dB for manually segmented data and $\theta = 5$ dB for automatically segmented data. The difference in threshold values suggests that the speech/non-speech classifier has a tendency to label noise as speech rather than opposite. The speech recognition results are given in Table 1.

4.2 Missing feature reconstruction

Speech recognition performance under environmental noise is evaluated with (a) the baseline system and when (b) cluster-based missing feature reconstruction or (c) acoustic model adaptation with the constrained maximum likelihood linear regression (CMLLR) or (d)–(e) both reconstruc-

Table 1: Speech recognition results over the 105-utterance speech/non-speech classifier evaluation data when the speech and non-speech boundaries for spectrographic mask estimation are automatically detected with the classifier or manually set. The close-talk data is almost clean speech with average SNR 24 dB and the far recorded data is noisy speech with average SNR 9 dB.

	Close WER	Close LER	Far WER	Far LER
Autom. segmentation	14.4	4.0	64.6	37.1
Manual segmentation	14.4	4.1	61.7	34.6

Table 2: Speech recognition results over the full evaluation dataset when missing feature reconstruction and constrained maximum likelihood linear regression (CMLLR) are used in a large vocabulary continuous speech recognition system. The close-talk data is almost clean speech with average SNR 24 dB and the far recorded data is noisy speech with average SNR 9 dB. The best results obtained with the data are underlined.

	Close WER	Close LER	Far WER	Far LER
(a) Baseline	13.4	3.4	62.0	37.2
(b) Reconstr.	13.3	3.3	47.0	23.8
(c) CMLLR	<u>11.6</u>	<u>2.7</u>	43.9	22.9
(d) Reconstr. $\rightarrow$ CMLLR	12.3	2.9	45.0	23.2
(e) Reconstr. + CMLLR	11.9	2.8	<u>41.9</u>	<u>22.2</u>

tion and adaptation are used for noise compensation. The reconstruction and adaptation are tested (d) in series and (e) in parallel as described in Section 2.3. The results are given in Table 2.

Missing feature reconstruction (b) does not significantly improve the speech recognition performance on the close-talk data, but with the far recorded data, the relative error reduction from missing feature reconstruction is 36 % compared to the baseline (a). The percentage of unreliable (missing) components in the frames classified as speech is on average 43 % among the utterances recorded with the close-talk microphone, and on average 63 % among the corresponding far recorded utterances. Speaker and environmental adaptation with CMLLR (c) improves the speech recognition results from the baseline (a) so that the relative error reduction is 21 % with the close-talk data and 38 % with the far recorded data. Thus, with the far recorded data, missing feature reconstruction (b) results in performance comparable to the CMLLR performance (c). The difference between the results (b) and (c) on the far recorded data is not statistically significant according to Wilcoxon signed rank test (see Table 3). With the close-talk data that almost corresponds to clean speech, speaker adaptation (c) improves the results significantly more than missing feature reconstruction (b).

Table 3: Test statistics from pairwise system comparisons. Systems are compared based on the letter error rate, and using the Wilcoxon signed rank test. The difference between two systems is statistically significant ($p < 0.05$) if the test statistic $|Z| > 1.98$. Test statistics suggesting statistical significance are underlined. The systems compared are (a) baseline, (b) reconstruction, (c) CMLLR, and (d) reconstruction and CMLLR in series and (e) reconstruction and CMLLR in parallel. See Table 2 for the system average performance rates.

Close-talk data

	(a)	(b)	(c)	(d)
(b)	0.24			
(c)	<u>-4.38</u>	<u>-3.98</u>		
(d)	<u>-2.75</u>	<u>-2.71</u>	-1.27	
(e)	<u>-4.38</u>	<u>-3.77</u>	-1.57	-0.95

Far recorded data

	(a)	(b)	(c)	(d)
(b)	<u>-4.94</u>			
(c)	<u>-4.94</u>	-1.33		
(d)	<u>-4.94</u>	-1.10	-0.42	
(e)	<u>-4.94</u>	<u>-2.42</u>	<u>-2.37</u>	-1.89

Adaptation improves the results on close-talk data also when applied together—either in series or in parallel—with missing feature reconstruction, but with the far recorded data, using reconstruction and adaptation in series (d) does not improve the results significantly compared to the results (b). Using reconstruction and adaptation in parallel does, and the relative error reduction is 40 % compared to the baseline (a). The difference in results (c) and (e) in close-talk data is not statistically significant.

5. CONCLUSIONS AND DISCUSSION

In this work, we used cluster-based missing feature reconstruction [3] in a LVCSR system trained with clean speech. We tested the method with noisy speech data recorded in real public environments, and used it with constrained maximum likelihood linear regression (CMLLR) [1]. In addition, we presented a new method for finding the noise corrupted speech signal components. The results indicate that missing feature reconstruction can significantly improve noise robustness under changing and unpredictable noise conditions. With the far recorded noisy speech data, the improvements from cluster-based missing feature reconstruction were comparable to the improvements from CMLLR speaker and environmental adaptation. With the close-talk data that almost corresponds to clean speech, the improvements were minor and not statistically significant.

CMLLR and the other methods from the linear transformation family are amongst the most popular and efficient methods for improving robustness in automatic speech

recognition. They have but one constraint: the methods need to be provided with enough (minimum 1000–1500 frames) adaptation data from the target speaker or environment in order for the estimated transformations to be reliable [18]. Since 2–3 utterances are sufficient, adaptation data is seldom a problem in continuous speech recognition tasks, unless that is, the data has not been organised in sessions (e.g. broadcast news, meeting room recordings). In such case, methods like missing feature reconstruction that do not need adaptation data have a clear advantage.

Missing feature methods do not use any information about the current speaker, and cannot compensate for speaker variation. LVCSR systems usually need some speaker compensation method in order to reach good performance, so missing feature reconstruction needs to be combined with e.g. CMLLR. In the reported experiments, CMLLR did not improve the speech recognition performance when the transformations were estimated from reconstructed features. When the methods were used in parallel, the results were marginally better than the results obtained with the methods used in series. With the parallel system, the relative error reduction was 40 % compared to the baseline. The parallel system results could be further improved if, for example, an adaptive weight $\alpha = \alpha(\tau)$ was used. The parallel approach, however, requires more computation than using the methods in series because state probabilities need to be calculated for two feature streams.

What should be determined now, is why CMLLR did not improve the speech recognition results when applied on the reconstructed features. It is possible that under noisy conditions, CMLLR focuses on modelling the environmental noise rather than the speaker, and effectively becomes a noise compensation method. This should, however, mostly affect the parallel system performance. When CMLLR is applied on cleaned speech i.e. after reconstruction, it should again compensate for speaker variation, but it is possible that (i) because the same speaker-independent model is used to reconstruct all features, missing feature reconstruction may have resulted in removing or smoothing speaker-specific characteristics, or (ii) missing feature reconstruction may have caused unusual variation in the features, in which case adaptation has sought to compensate for this rather than speaker variation.

If missing feature reconstruction degrades speaker adaptation performance as suggested above, adaptation and reconstruction should be applied in a different order: adaptation before reconstruction. The reversal is not straightforward as missing feature reconstruction operates on spectral features while adaptation is used only after cepstral and the other feature transformations. Note that speaker normalisation methods applied in time or spectral domain come naturally before missing feature reconstruction, so there should be no similar problems in combining missing feature reconstruction and speaker normalisation.

6. ACKNOWLEDGMENTS

This work was supported by the Academy of Finland in the projects *Auditory approaches to automatic speech recognition*, *New adaptive and learning methods in speech recognition*, and *Adaptive Informatics* and by the Helsinki Graduate School in Computer Science.

REFERENCES

[1] M. J. F. Gales, "Maximum likelihood linear transformations for HMM–based speech recognition," *Computer Speech and Language*, vol. 12, pp. 75–98, Apr. 1998.

[2] M. Cooke, P. Green, L. Josifovski, and A. Vizinho, "Robust automatic speech recognition with missing and unreliable acoustic data," *Speech Communication*, vol. 34, pp. 267–285, Jun. 2001.

[3] B. Raj, M. L. Seltzer, and R. M. Stern, "Reconstruction of missing features for robust speech recognition," *Speech Communication*, vol. 43, pp. 275–296, Sep. 2004.

[4] A. S. Bregman, *Auditory Scene Analysis*. Address: MIT Press, 1990.

[5] J. Barker, "Robust automatic speech recognition," in D. L. Wang and G. J. Brown, editors, *Computational Auditory Scene Analysis*. Address: Wiley-Interscience, 2006.

[6] M. Van Segbroeck and H. Van hamme, "Vector-quantization based mask estimation for missing data automatic speech recognition," in *Proc. INTERSPEECH 2007*, Antwerp, Belgium, August 27-31. 2007, pp. 910–913.

[7] K. J. Palomäki, G. J. Brown, J. P. Barker, "Recognition of reverberant speech using full cepstral features and spectral missing data," in *Proc. ICASSP 2006*, Toulouse, France, May 15-19. 2006, pp. 289–292.

[8] M. L. Seltzer, B. Raj, and R. M. Stern, "A bayesian framework for spectrographic mask estimation for missing feature speech recognition," *Speech Communication*, vol. 43, pp. 379–393, Sep. 2004.

[9] V. Siivola and B. Pellom, "Growing an n-gram language model," in *Proc. INTERSPEECH 2005*, Lisbon, Portugal, September 4-8. 2005, pp. 1309–1312.

[10] T. Hirsimäki, M. Creutz, V. Siivola, M. Kurimo, S. Virpioja, and J. Pylkkönen, "Unlimited vocabulary speech recognition with morph language models applied to Finnish," *Computer Speech and Language*, vol. 20, pp. 515–541, Oct. 2006.

[11] J. Pylkkönen, "An efficient one-pass decoder for Finnish large vocabulary continuous speech recognition," in *Proc. 2nd Baltic Conference on Human Language Technologies*, Tallinn, Estonia, April 4-5. 2005, pp. 167–172.

[12] J. J. Odell, *The use of context in large vocabulary speech recognition*. Ph.D. thesis, University of Cambridge, 1995.

[13] J. Pylkkönen and M. Kurimo, "Duration modeling techniques for continuous speech recognition," in *Proc. INTERSPEECH 2004*, Jeju Island, Korea, October 4-8. 2004, pages 385–388.

[14] M. J. F. Gales, "Semi-tied covariance matrices for hidden Markov models," *IEEE Trans. SAP*, vol. 7, pp. 272–281, May. 1999.

[15] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. ASSP*,vol. 28, pp. 357–366, Aug. 1980.

[16] *GMMBAYES*. [http://www.it.lut.fi/project/gmmbayes/].

[17] D. Iskra, B. Grosskopf, K. Marasek, H. van den Heuvel, F. Diehl, and A. Kiessling, "SPEECON - speech databases for consumer devices: Database specification and validation," in *Proc. LREC*, Las Palmas, Canary Islands, Spain, May 29-31. 2002, pp. 329–333.

[18] P. C. Woodland, "Speaker adaptation for continuous density HMMs: A review," in *Proc. ISCA Tutorial and Research Workshop on Adaptation Methods for Speech Recognition*, Sophia Antipolis, France, August 29-30. 2001, pp. 11–19.