

Research Article

Comparative Evaluation of Single-Channel MMSE-Based Noise Reduction Schemes for Speech Recognition

Emanuele Principi, Simone Cifani, Rudy Rotili, Stefano Squartini, and Francesco Piazza

3MediaLabs, Department of Biomedical Engineering, Electronics and Telecommunications, Polytechnic University of Marche, 60121 Ancona, Italy

Correspondence should be addressed to Emanuele Principi, e.principi@univpm.it

Received 4 September 2009; Accepted 11 February 2010

Academic Editor: Andreas Spanias

Copyright © 2010 Emanuele Principi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One of the big challenges in the field of Automatic Speech Recognition (ASR) consists in developing suitable solutions able to work properly also in adverse acoustic conditions, like in presence of additive noise and/or in reverberant rooms. Recently a certain attention has been paid to deeply integrate the noise suppressor in the feature extraction pipeline. In this paper, different single-channel MMSE-based noise reduction schemes have been implemented both in the frequency and cepstral domains and the related recognition performances evaluated on the AURORA2 and AURORA4 databases, therefore providing a useful reference for the scientific community.

1. Introduction

Automatic Speech Recognition (ASR) is a challenging task largely addressed by the scientific community in the last two decades. A notable interest raised during last years in the study and development of robust solutions in presence of acoustic nonidealities [1], for example, background noise, simultaneous speakers, and reverberation. As result of these efforts, a profuse literature of environment-robust ASR techniques has been registered. The following classification can be proposed therein, as highlighted in [2]: feature-domain (FD) and model-based (MB) algorithms. The latter class encompasses all methodologies aimed to adapt the acoustic model (HMM) parameters in order to maximize the system matching to the distorted environment. Related contributions propose bayesian compensation frameworks [3], and most of them typically rely on Vector Taylor Series approximation for compensating parameter estimation [2].

The FD algorithm class directly enhances the quality of the speech features, trying to make them as close as possible to the clean-speech environment condition. Among the proposed methodologies, we can cite the log-spectral amplitude MMSE suppression rules due to their efficacy to reduce noise at a cost of low distortion level [4, 5]. Recently,

these rules have been implemented in the cepstral domain, so working closely to the backend [6, 7]. Even though the MB approach achieves better performances [2], the FD one still has some advantages. Firstly, independence from the backend: all modifications are accomplished on the feature vectors rather than on the HMM parameters, which has a significant practical mean. Secondly, easy implementation: the algorithm parameterization is extremely simpler with respect to the MB class and no adaptation is requested. This justifies why the FD methodologies are still largely considered and implemented in ASR-based products.

In this paper, the authors have selected some of the most reliable and performing techniques in this field and put them in comparison in terms of recognition performances using the AURORA2 and AURORA4 databases. The noise reduction algorithms involved are all MMSE-based, both in the frequency [4, 5] and in the cepstral [7] domain. Some optimizations have been implemented in frequency-domain [8, 9] and generalized to cepstral domain leading to new and effective algorithmic solutions with respect to what appeared in the literature [7, 10]. Moreover, according to authors' perspective, comparative results of performed simulations provided here would represent a useful reference for further studies.

Here is the outline of the paper. Sections 2 and 3 will be devoted to the description of the noise reduction algorithms in the frequency and cepstral, domains respectively, with the related optimization procedures. In Section 4, results of performed computer simulations are reported and commented. Section 5 concludes the work together with highlights to some future work ideas.

2. Background on Frequenc-Domain MMSE Algorithms

Let $x(n)$ and $d(n)$ denote speech and uncorrelated additive noise signals, respectively, where n is a discrete-time index. The observed noisy signal $y(n)$ is given by $y(n) = x(n) + d(n)$. Let $X(k, l)$, $Y(k, l)$, and $D(k, l)$ denote the short-time Fourier transforms of $x(n)$, $y(n)$, and $d(n)$ respectively, where k is the frequency bin index and l is the time frame index. Our purpose is to find an estimate of the clean speech signal $X(k, l)$, denoted with $\hat{X}(k, l)$, given the noisy signal $Y(k, l)$. In [4], such an estimate is obtained using a log-spectrum MMSE estimator and is given by

$$\left| \hat{X}(k, l) \right|^2 = G(k, l) |Y(k, l)|^2. \quad (1)$$

The gain function $G(k, l)$ is computed as

$$G(k, l) = \frac{\xi(k, l)}{1 + \xi(k, l)} \exp\left(\frac{1}{2} \int_{\nu(k, l)}^{\infty} \frac{e^{-t}}{t} dt\right), \quad (2)$$

where $\xi(k, l)$ and $\gamma(k, l)$ are the *a priori* and *a posteriori* SNR, respectively,

$$\xi(k, l) = \frac{\sigma_x^2(k, l)}{\sigma_d^2(k, l)}, \quad \gamma(k, l) = \frac{|Y(k, l)|^2}{\sigma_d^2(k, l)}, \quad (3)$$

and $\nu(k, l) = \gamma(k, l)\xi(k, l)/(1 + \xi(k, l))$.

The noise variance $\sigma_d^2(k, l)$ can be estimated using the well-known minima controlled recursive averaging (MCRA) [5] or the improved MCRA (IMCRA) [11]. Both options have been tested, and recognition results are provided in Section 4. The *a priori* SNR is estimated using the popular *decision-directed* algorithm [4]. In Section 5, this algorithm will be denoted as LSA.

2.1. Gain Modification Based on Soft Decision

2.1.1. Optimally Modified LSA Estimator (OMLSA). In [12], the optimally modified log-MMSE (OMLSA) estimator has been proposed. The optimal spectral gain function is obtained as a weighted geometric mean of the hypothetical gains associated with the speech presence uncertainty. The exponential weight of each hypothetical gain is its corresponding probability, conditional on the observed signal. The probability function is estimated for each frame and each subband via a soft-decision approach, which exploits the strong correlation of speech presence in neighboring frequency bins of consecutive frames. The modified gain function takes the following form:

$$G(k, l) = [G_{H_1}(k, l)]^{p(k, l)} G_{\min}^{1-p(k, l)}, \quad (4)$$

where $G_{H_1}(k, l)$ is the same as (2), $p(k, l)$ is the *speech presence probability* (SPP), and G_{\min} is a lower threshold [12]. The speech presence probability is computed as

$$p(k, l) = \left\{ 1 + \frac{q(k, l)}{1 - q(k, l)} (1 + \xi(k, l)) e^{-\nu(k, l)} \right\}^{-1}, \quad (5)$$

where $q(k, l)$ is the *a priori* speech absence probability estimated using a soft-decision approach [5].

2.1.2. Gain Modification Based on Global SNR (gSNR). In [8], a soft-decision gain modification was introduced to improve the efficacy of Ephraim and Malah algorithm (E&M) when used as a preprocessing stage of ASR. The estimates of $\xi(k, l)$ and $\gamma(k, l)$ are modified as follows:

$$\begin{aligned} \hat{\xi}(k, l) &= \max\left(\eta \frac{|\hat{X}(k, l-1)|^2}{\alpha(l)\hat{\sigma}_d^2(k, l-1)} + (1-\eta)[\tilde{\gamma}(k, l)-1], \beta(l)\right), \\ \tilde{\gamma}(k, l) &= \max\left(\frac{|Y(k, l)|^2}{\alpha(l)|\hat{\sigma}_d^2(k, l)|} - 1, \beta(l)\right) + 1, \end{aligned} \quad (6)$$

where $\alpha(l)$ and $\beta(l)$ are the noise overestimation factor and the spectral floor, respectively, and depend on the global SNR(l) as in [8]. The global SNR can be computed recursively as

$$\begin{aligned} \text{SNR}(l) &= \alpha_{\text{SNR}} 10 \log_{10} \left(\frac{\sum_k |\hat{X}(k, l-1)|^2}{\sum_k \hat{\sigma}_d^2(k, l-1)} \right) \\ &\quad + (1 - \alpha_{\text{SNR}}) \text{SNR}(l-1), \end{aligned} \quad (7)$$

where $\alpha_{\text{SNR}} \in [0, 1]$.

Using (6), the gain function takes greater values when the global SNR is around 20 dB, while it remains unchanged when it is around 0 dB [8]. In [8], this approach was applied to LSA, in this paper we propose its application also to OMLSA.

2.2. Gain Function Smoothing (G_{smooth}). In [9], a frequency-domain smoothing of the gain function is proposed to improve the perceived quality of speech signals processed by E&M algorithm. Smoothing is performed by dividing the frequency range in critical bands and computing the median of the gain function in each band.

The smoothed gain function is computed as follows:

$$\tilde{G}(k, l) = \beta_S G(k, l) + (1 - \beta_S) \underset{k \in \langle M(k') \rangle}{\text{med}} (G(k, l)), \quad (8)$$

where “ $\text{med}(f(k))$ ” represents the median of $f(k)$ and $M(k')$ is the number of frequency bins at the k' th critical band. The smoothing factor β_S controls the residual musical noise and the perceived clearness in the output signal [9]. In [9], this approach was applied to LSA, in this paper we propose its application also to OMLSA.

3. Background on Cepstral Domain MMSE Algorithms

The MFCC-MMSE noise reduction algorithm proposed in [7] operates directly on the MFCC coefficients, that is, in the same domain of the decoding algorithm. The followed approach is similar to E&M [4] but differs because the algorithm is applied to the power spectral magnitude of the filter bank's output instead of the DFT spectral amplitude and because the noise variance takes into account the phase difference between the noise and the clean speech.

As reported in [7], the MMSE estimate of the the j th MFCC coefficients is given by the following expression:

$$\hat{c}_x(j, l) = \sum_b a_{j,b} \log(G(b, l)m_y(b, l)), \quad (9)$$

where $G(b, l)$ has the same form of (2) with the frequency bin k replaced with the channel index b . The terms $a_{j,b}$ are the discrete cosine transform coefficients, and $m_y(b, l)$ is the output of the b th Mel-frequency filter at frame l .

The *a priori* and *a posteriori* SNR $\xi(b, l)$ and $\gamma(b, l)$ are given by

$$\begin{aligned} \xi(b, l) &\approx \frac{\sigma_x^2(b, l)}{\sigma_d^2(b, l) + \sigma_\phi^2(b, l)}, \\ \gamma(b, l) &\approx \frac{m_y^2(b, l)}{\sigma_d^2(b, l) + \sigma_\phi^2(b, l)}. \end{aligned} \quad (10)$$

The terms $\sigma_d^2(b, l)$ and $\sigma_x^2(b, l)$ are the variances of the noise and of the speech signal, respectively and $\sigma_\phi^2(b, l)$ is the variance of the phase difference between noise and speech.

As in frequency domain algorithm, the *a priori* SNR has been estimated with the decision-directed approach. In [7], the noise variance has been estimated with MCRA algorithm. In this paper, both the MCRA and IMCRA algorithms have been tested and results are shown in Section 4. The variance $\sigma_\phi^2(b)$ is computed as in [7]. In Section 4, this algorithm will be denoted as C-LSA.

3.1. Cepstral Domain Gain Modification Based on Soft Decision

3.1.1. Optimally Modified LSA Estimator (C-OMLSA). Here, we propose an adaptation of the OMLSA algorithm to the cepstral domain. The speech presence uncertainty can be taken into account in the cepstral domain as well. However, the basic idea of exploiting the strong correlation between neighboring frequency bins does not hold using cepstral coefficients anymore. This problem can be overcome by computing the SPP in the frequency domain and then choosing the median value of all SPP coefficients belonging to a Mel-channel as the probability value of that channel. The MMSE estimate of the MFCC coefficients is then obtained by applying (4).

3.1.2. Gain Modification Based on Global SNR ($gSNR$). Here, we propose an adaptation of the gain modification to the

TABLE 1: Parameters values used in computer simulations.

IMCRA	$\alpha_S = 0.8, \alpha_D = 0.95$
Decision directed	$\alpha = 0.9$
$gSNR$	$\alpha_{SNR} = 0.8$
OMLSA/C-OMLSA	$G_{min} = -10$ dB, $\alpha = 0.98$
G_{smooth} (cepstral domain)	$\beta_S = 0.8, \Delta b = 1$

cepstral domain algorithm. The gain function has been modified by calculating $\xi(b, l)$ and $\gamma(b, l)$ using (6), where $\hat{X}(k, l - 1)$ and $Y(k, l - 1)$ have been replaced by $\hat{m}_x(b, l - 1)$ and $m_y(b, l - 1)$, respectively.

A similar approach has been proposed in [10]: in that paper, the gain value is directly modified on the basis of the estimated noise power σ_d^2 , while here a smoothed SNR is used.

3.2. Cepstral Domain Gain Function Smoothing (G_{smooth}). Here, we propose the smoothing of gain function in the cepstral domain. A similar approach for the MFCC-MMSE algorithm has been proposed in [10]: in that paper, the smoothed gain was calculated as a weighted average between the gain value on the actual and on the previous frame. Differently, in this paper, smoothing is computed as in (8), by calculating the median of the gain value in each critical band. Equation (8) must be modified substituting the frequency index k with the channel index b and calculating the median in the interval $b \pm \Delta b$. The low degree of correlation between neighboring Mel-coefficients brought us to choose $\Delta b = 1$.

4. Computer Simulations

Evaluation has been conducted on the AURORA2 [13] and AURORA4 [14] databases. AURORA2 consists of a subset of TIDigits utterances downsampled to 8 kHz and with added noise and channel distortion. AURORA4 consists of the Wall Street Journal database to which various noise types have been added at different SNR levels. Experiments on AURORA4 have been conducted on the 8 kHz version of the official selection test set composed of 166 utterances.

Recognition has been performed using the Hidden Markov Model Toolkit [15]. Acoustic models structure and recognition parameters are the same as in [13] for AURORA2 tests, and as in [14] for AURORA4. Feature vectors are composed of 13 MFCCs (with C0 and without energy) and their first and second derivatives. Pre-emphasis and cepstral mean normalization have been included in the feature extraction pipeline.

Recognition results are expressed as word accuracy percentage. On AURORA2 results, averages are computed on the 0–20 dB SNR range [13]; on AURORA4 results, averages are computed on noisy test sets relative to the Sennheiser microphone (sets 2–7) and on test sets relative to the second microphone (sets 9–14) [14]. For both databases, clean and multicondition acoustic models have been created using the provided training sets. Results obtained with clean and multicondition acoustic models are reported in column denoted with “C” and “M”, respectively.

TABLE 2: Results of frequency-domain algorithms on AURORA2.

	Test A		Test B		Test C		A-B-C AVG	
	C	M	C	M	C	M	C	M
LSA	76.83	87.02	77.06	85.24	78.97	88.48	77.62	86.91
LSA + gSNR	78.69	89.45	79.95	88.33	79.80	89.95	79.48	89.24
LSA + G_{smooth}	76.44	87.12	77.34	85.97	77.61	87.66	77.13	86.92
OMLSA	76.81	89.13	78.15	88.22	78.34	90.20	77.77	89.18
OMLSA + gSNR	77.28	89.88	79.00	89.13	78.10	90.23	78.13	89.75
OMLSA + G_{smooth}	75.49	88.88	77.77	88.59	76.49	89.56	76.58	89.01
baseline	65.00	88.11	69.89	89.09	65.87	89.18	66.92	88.79

TABLE 3: Results of cepstral domain algorithms on AURORA2.

	Test A		Test B		Test C		A-B-C AVG	
	C	M	C	M	C	M	C	M
C-LSA	75.68	87.81	77.06	86.85	76.94	89.25	76.56	87.97
C-LSA + gSNR	75.17	89.42	77.44	88.95	75.82	90.04	76.14	89.47
C-LSA + G_{smooth}	76.37	88.23	77.74	87.06	77.30	89.38	77.14	88.22
C-OMLSA	72.69	89.00	75.85	88.60	73.62	90.28	74.05	89.29
C-OMLSA + gSNR	70.67	89.72	74.19	89.61	71.26	90.64	72.04	89.99
C-OMLSA + G_{smooth}	72.86	89.09	76.06	88.71	73.69	90.42	74.20	89.41

TABLE 4: Results of frequency-domain algorithms on AURORA4.

	Set 01		Set 02–07		Set 08		Set 09–14		Average	
	C	M	C	M	C	M	C	M	C	M
LSA	87.92	74.48	52.07	64.48	77.08	74.99	39.00	55.15	50.82	62.16
LSA + gSNR	88.73	77.78	60.71	72.19	81.36	78.20	49.17	65.48	59.24	70.14
LSA + G_{smooth}	88.14	77.69	54.11	65.38	76.88	74.22	40.73	57.51	52.43	63.52
OMLSA	88.32	79.08	56.31	69.27	77.20	76.24	43.60	61.59	54.64	67.18
OMLSA + gSNR	88.99	79.74	61.04	72.66	79.77	78.82	49.06	65.93	59.24	70.72
OMLSA + G_{smooth}	88.62	79.12	57.60	70.52	75.42	75.47	44.67	63.22	55.55	68.36
Baseline	89.80	84.53	54.57	74.57	77.05	81.62	42.82	67.68	53.65	72.83

Table 1 shows the values of the parameters used in computer simulations. For the sake of conciseness, only the parameters which values differ from the ones reported in references papers are reported. Unless otherwise stated, the same values are used both in frequency and cepstral domain.

4.1. Comparison of MCRA and IMCRA Algorithms. In order to evaluate the impact of the noise estimator algorithm, the LSA and C-LSA algorithms have been tested both with the MCRA and the IMCRA estimators on the AURORA2 database. The overall average accuracy for LSA algorithm is 80.21% using MCRA, and 82.27% using IMCRA; as for the C-LSA algorithm, accuracy is 79.15% using MCRA and 82.27% using IMCRA.

Results clearly show that the accuracy of noise variance estimation is crucial for both the frequency and cepstral domain algorithms. In the following experiments, the

IMCRA estimator has been used on all noise reduction algorithms.

4.2. Results and Discussion. Tables 2 and 3 show results on AURORA2 database for frequency and cepstral domain algorithms, respectively, while Tables 4 and 5 show results on AURORA4. Tables 2 and 4 also show results obtained without noise reduction algorithms (baseline). Higher accuracy values are shown in **bold**.

Frequency domain results on AURORA2 show that LSA algorithm produces a remarkable improvement of recognition accuracy, and that the gSNR modification gives a further increase of about 2% on average. On the contrary, the gain smoothing does not produce significant results, and OMMLSA does not give any advantage over the LSA counterpart.

Cepstral domain LSA results are comparable to frequency domain ones, and the introduced modifications produce

TABLE 5: Results of cepstral domain algorithms on AURORA4.

	Set 01		Set 02–07		Set 08		Set 09–14		Average	
	C	M	C	M	C	M	C	M	C	M
C-LSA	86.59	76.17	52.95	67.86	71.95	72.67	39.06	58.76	50.75	64.90
C-LSA + gSNR	88.91	80.66	58.09	73.58	79.40	78.16	46.07	66.18	56.66	71.24
C-LSA + G_{smooth}	86.63	76.65	54.26	68.72	72.96	73.52	40.63	59.88	52.07	65.84
C-OMLSA	89.58	83.35	55.64	74.86	77.13	80.74	43.51	67.77	54.40	72.84
C-OMLSA + gSNR	89.72	84.35	55.16	74.82	77.24	81.03	43.17	67.83	54.07	72.95
C-OMLSA + G_{smooth}	89.65	83.54	55.74	74.97	76.91	80.44	43.58	67.79	54.46	72.90

similar accuracy values. Cepstral domain OMLSA slightly degrades accuracy with the clean acoustic model, while it slightly improves it with the multicondition one.

On AURORA4, frequency domain results show that LSA algorithm alone degrades recognition performance, and only introducing the gSNR modification one can obtain accuracy increase. OMLSA gives a performance improvement over LSA and show a similar behavior using gSNR. Similar considerations can be made on cepstral domain algorithms: differently from AURORA2 results, the gSNR modification increases accuracy of about 4% on average over C-OMLSA.

Both in AURORA2 and AURORA4 results, cepstral domain algorithms perform better than frequency domain ones when multicondition acoustic model is used. This is consistent with the results obtained in [7] on the AURORA3 database, in which the training set is always multicondition.

Indeed, cepstral domain algorithms advantage over frequency domain ones is the reduced computational complexity. Typically, the number of frequency bins is around 256 in frequency-domain noise reduction algorithms, while 23 is the number of Mel-channels used in many ASR systems. This results in a complexity reduced of about one order of magnitude and makes cepstral domain algorithms the best choice for applications with limited computational resources or with strict real-time constraints. In [7], experiments were conducted on AURORA3 database: this paper completes the results with AURORA2 and AURORA4 experiments and proposed different solutions to the algorithms improvements presented in [10].

5. Conclusions

In this paper a comparative evaluation of single-channel MMSE-based feature enhancement schemes in terms of speech recognition performances has been carried out. As aforementioned, such an approach has some advantages with respect to the MB techniques for robust ASR and hence a certain appealing for practical implementations. Different algorithms have been considered, both in the frequency and in the cepstral domain, including some interesting innovative improvements to existing approaches, in this sense the contribution wants to be a useful reference for researchers in the field. The computer simulations have been performed by using HTK on the AURORA2 and AURORA4 databases, completing the results in [7]. Obtained results allowed

us concluding that cepstral domain solutions have similar performances to their log-spectral domain counterpart with clean condition acoustic models, superior performance with multicondition ones, while attaining a significative computational burden reduction. Future efforts will be oriented to analyze the impact of further optimizations on the addressed algorithms and the extension to the multichannel case study.

References

- [1] A. Peinado and J. Segura, *Speech Recognition over Digital Channels: Robustness and Standards*, John Wiley & Sons, West Sussex, UK, 2006.
- [2] J. Li, L. Deng, D. Yu, Y. Gong, and A. Acero, “A unified framework of HMM adaptation with joint compensation of additive and convolutive distortions,” *Computer Speech and Language*, vol. 23, no. 3, pp. 389–405, 2009.
- [3] X. Wang and D. O’Shaughnessy, “Environmental independent ASR model adaptation/compensation by bayesian parametric representation,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 4, pp. 1204–1217, 2007.
- [4] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error-log-spectral amplitude estimator,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 443–445, 1985.
- [5] I. Cohen, “Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator,” *IEEE Signal Processing Letters*, vol. 9, no. 4, pp. 113–116, 2002.
- [6] K. M. Indrebo, R. J. Povinelli, and M. T. Johnson, “Minimum mean-squared error estimation of mel-frequency cepstral coefficients using a novel distortion model,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 8, pp. 1654–1661, 2008.
- [7] D. Yu, L. Deng, J. Droppo, J. Wu, Y. Gong, and A. Acero, “Robust speech recognition using a cepstral minimum-mean-square-error-motivated noise suppressor,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 5, pp. 1061–1070, 2008.
- [8] R. Gemello, F. Mana, and R. De Mori, “Automatic speech recognition with a modified Ephraim-Malah rule,” *IEEE Signal Processing Letters*, vol. 13, no. 1, pp. 56–59, 2006.
- [9] M.-S. Choi and H.-G. Kang, “An improved estimation of a priori speech absence probability for speech enhancement: in perspective of speech perception,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’05)*, vol. 1, pp. 1117–1120, 2005.
- [10] D. Yu, L. Deng, J. Wu, Y. Gong, and A. Acero, “Improvements on Mel-frequency cepstrum minimum-mean-square-error noise suppressor for robust speech recognition,” in *Proceedings*

- of the 6th International Symposium on Chinese Spoken Language Processing (ISCSLP '08)*, pp. 69–72, December 2008.
- [11] I. Cohen, “Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging,” *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 466–475, 2003.
 - [12] I. Cohen and B. Berdugo, “Speech enhancement for non-stationary noise environments,” *Signal Processing*, vol. 81, no. 11, pp. 2403–2418, 2001.
 - [13] H.-G. Hirsch and D. Pearce, “The Aurora experimental framework for the performance evaluation of speech recognition systems under noise conditions,” in *Proceedings of the Automatic Speech Recognition: Challenges for the New Millennium (ISCA ITRW ASR '00)*, Paris, France, September 2000.
 - [14] N. Parihar, J. Picone, D. Pearce, and H.-G. Hirsch, “Performance analysis of the Aurora large vocabulary baseline system,” in *Proceedings of the European Signal Processing Conference*, 2004.
 - [15] S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, *The HTK Book. V2.2*, 1999.

