A Phase-Averaged Model for the Relationship between Noisy Speech, Clean Speech and Noise in the log-Mel Domain

Friedrich Faubel, John McDonough, Dietrich Klakow

Spoken Language Systems,
Saarland University, D66123 Saarbrücken, Germany
{friedrich.faubel,john.mcdonough,dietrich.klakow}@lsv.uni-saarland.de

Abstract

In this work, we demonstrate that the most widely-used model for the relationship between noisy speech, clean speech and noise in the log-Mel domain is inaccurate due to its disregard of the phase. Moreover, we show how a more exact model can be derived by averaging over the phase in the log-Mel domain, and how this can profitably be applied to particle filter based sequential noise compensation. Experimental results confirm the superiority of the phase-averaged model for both clean speech estimation in general and the particle filter in particular. Reductions in word error rate of up to 17% relative were obtained on a large vocabulary task.

Index Terms: model, relative phase, noise compensation

1. Introduction

In [1] Acero gave a formula for the relationship between noisy speech, clean speech and noise in the log power spectral domain, which would later be used in a large variety of approaches: *parallel model combination* (PMC) [2], *vector Taylor series* (VTS) [3], sequential *expectation maximization* (EM) [4], *interacting multiple models* (IMMs) [5] and, more recently, *particle filters* (PFs) [6, 7]. All of these approaches use an auxiliary clean speech model, either a *Gaussian mixture model* (GMM) [3, 4, 5, 7] or a *hidden Markov Model* (HMM) [2, 6], to map the *probability density function* (pdf) of clean speech to the pdf of noisy speech, based on the formula given in [1], which we from now on refer to as the standard model.

The standard model is deficient in that it does not take account of the relative phase between speech and noise nor its distribution. Usually it is argued that the phase is zero in average and hence the phase term can be neglected. We show that this assumption introduces systematic errors in the log-Mel domain. Admittedly, modeling the phase is not entirely new, as Deng has previously derived a phase-dependent model, which has turned out to be beneficial for speech feature enhancement [8, 9]. The approach proposed here deviates from Deng's approach in two points: first, we do not make the Gaussian assumption, and second, we average the model instead of maintaining a probabilistic phase term.

In addition to demonstrating the superiority of the phase-averaged model over the standard model with respect to clean speech estimation, we show how it can profitably be applied to particle filter based sequential noise compensation.

2. A phase-averaged model

In this section, we derive a phase-averaged model for the relationship between noisy speech, clean speech and noise in the log-Mel domain. This will be done in the following order: first a phase-dependent model is established in Section 2.1. Then the distribution of phase factors is investigated in Section 2.2. Section 2.3 finally shows how the averaging can be performed and how it can be implemented efficiently.

In what follows, all of the vector multiplications are considered to be element-wise, as are the functions of vectors; e.g. for $\mathbf{x} = [x_1 \ldots x_n]^T$: $\cos(\mathbf{x}) = [\cos(x_1) \ldots \cos(x_n)]^T$.

2.1. A phase-dependent model

Denoting magnitude spectra of noisy speech, clean speech and noise by Y, X and N respectively, their relationship in the power spectral domain can be shown [9] to be

$$Y^2 = X^2 + N^2 + 2\cos(\theta)XN, \qquad (1)$$

where θ is the vector of relative phases between X and N. Typically, the phase term $2\cos(\theta)XN$ is omitted based on the argument it is zero on average, such that

$$Y^2 = X^2 + N^2. \qquad (2)$$

This is perfectly reasonable for the power spectral domain. Note, however, that when (2) is translated into the log-Mel domain to obtain the standard model equation given in [1],

$$\mathbf{y} = \log\left(e^{\mathbf{x}} + e^{\mathbf{n}}\right), \qquad (3)$$

a nonlinear transform is applied. The problem with that is that the mean $\mathrm{E}_{p(f(x))}\left[f(x)\right]$ of a nonlinearly transformed pdf is not necessarily equal to the transformed mean $f(\mathrm{E}_{p(x)}\left[x\right])$ of the original pdf. So the effect of the phase term might not be zero on average after taking the logarithm. And indeed it is not, as can be seen in Figure 1, which compares equation (3) – the log of the mean of (1) – to the mean of the log of (1) – the phase-averaged model proposed in this paper.

In the Mel frequency domain equation (1) becomes:

$$\tilde{Y} = \tilde{X} + \tilde{N} + 2\boldsymbol{\alpha}\sqrt{\tilde{X}\tilde{N}}, \qquad (4)$$

where the $\tilde{Y}$, $\tilde{X}$, $\tilde{N}$ are the products of Y^2, X^2, N^2 with the Mel filterbank matrix W and where $\boldsymbol{\alpha}$ is the vector of phase factors

$$\alpha_i = \frac{\sum_k W_{i,k}\cos(\theta_k)X_k N_k}{\sqrt{\tilde{X}_i\tilde{N}_i}}$$

with $-1 \leq \alpha_i \leq 1$ as derived in [9]. Some speech feature enhancement approaches require solving equation (4) for $\tilde{X}$:

$$\tilde{X} = \left(\pm\sqrt{\tilde{Y} + (\boldsymbol{\alpha}^2 - \mathbf{1})\tilde{N}} - \boldsymbol{\alpha}\sqrt{\tilde{N}}\right)^2,$$

This work was supported by the Federal Republic of Germany through the German Research Foundation (DFG), under the research training network IRTG 715 "Language Technology and Cognitive Systems".

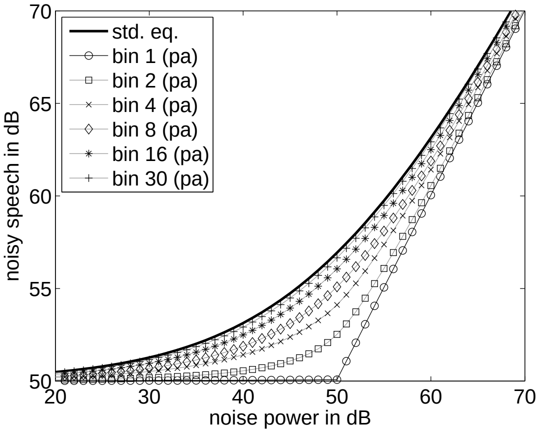

Figure 1: *Noisy speech power y as a function of noise power n in the log-Mel domain calculated with the standard model (3), as well as with the phase-averaged model (8) proposed in this paper. Note that the averages are dependent on the frequency bin.*

where $\mathbf{1}$ denotes a vector of ones. Going to the log-Mel domain by replacing $\tilde{Y}$, $\tilde{X}$ and $\tilde{N}$ with $e^{\mathbf{y}}$, $e^{\mathbf{x}}$ and $e^{\mathbf{n}}$ respectively gives

$$e^{\mathbf{x}} = \left(\pm\sqrt{e^{\mathbf{y}} + (\boldsymbol{\alpha}^2 - \mathbf{1})e^{\mathbf{n}}} - \boldsymbol{\alpha}\sqrt{e^{\mathbf{n}}}\right)^2,$$

which can be solved for $\mathbf{x}$ by pulling $\sqrt{e^{\mathbf{y}}}$ out of the square and then taking the logarithm:

$$\mathbf{x} = \underbrace{\mathbf{y} \pm \log\left(\sqrt{\mathbf{u}} - \mathbf{v}\right)^2}_{\triangleq f_\pm(\mathbf{y},\mathbf{n},\boldsymbol{\alpha})}, \tag{5}$$

where $\mathbf{u} \triangleq \mathbf{1} + (\boldsymbol{\alpha}^2 - \mathbf{1})e^{\mathbf{n}-\mathbf{y}}$ and $\mathbf{v} \triangleq \boldsymbol{\alpha}\sqrt{e^{\mathbf{n}-\mathbf{y}}}$. Note that this solution exists only if $u_i \geq 0$ and $(\sqrt{u_i} - v_i) \neq 0$ for all i. The corresponding Jacobian is

$$\frac{df(\mathbf{y},\mathbf{n},\boldsymbol{\alpha})}{d\mathbf{y}} = \underbrace{\operatorname{diag}\left(\frac{\mathbf{1}}{\mathbf{u} \pm \mathbf{v}\sqrt{\mathbf{u}}}\right)}_{\triangleq f'_\pm(\mathbf{y},\mathbf{n},\boldsymbol{\alpha})}, \tag{6}$$

where $\operatorname{diag}(\cdot)$ denotes the operator that translates a vector to a diagonal matrix, and the fraction denotes a component-wise vector division. As the translation of equation (4) to the log-Mel domain was derived in [9],

$$\mathbf{y} = \underbrace{\mathbf{x} + \log\left(\mathbf{1} + e^{\mathbf{n}-\mathbf{x}} + 2\boldsymbol{\alpha}\sqrt{e^{\mathbf{n}-\mathbf{x}}}\right)}_{\triangleq g(\mathbf{x},\mathbf{n},\boldsymbol{\alpha})}, \tag{7}$$

we now have phase-dependent equations for $\mathbf{x} = f(\mathbf{y},\mathbf{n},\boldsymbol{\alpha})$, $d\mathbf{x}/d\mathbf{y} = f'(\mathbf{y},\mathbf{n},\boldsymbol{\alpha})$ and $\mathbf{y} = g(\mathbf{x},\mathbf{n},\boldsymbol{\alpha})$. The aim is to average these with respect to the phase factors, which requires knowledge of their distribution.

2.2. Distribution of the phase factors

Deng [9] argued that the phase factors follow a zero-mean Gaussian distribution and learned the variance terms accordingly. An alternative is to directly use the empirical distribution, which can be obtained in a controlled experiment, by adding known speech and noise signals and computing the phase factors of the corresponding Mel spectra according to

$$\alpha_i = \frac{\tilde{Y}_i - \tilde{X}_i - \tilde{N}_i}{2\sqrt{\tilde{X}_i\tilde{N}_i}},$$

where i is an index over the Mel frequency bins. Figure 2 shows the resulting distribution for the multi-channel Wall Street Journal audio visual corpus [10] training set and factory noise from the NOISEX-92 database [11].

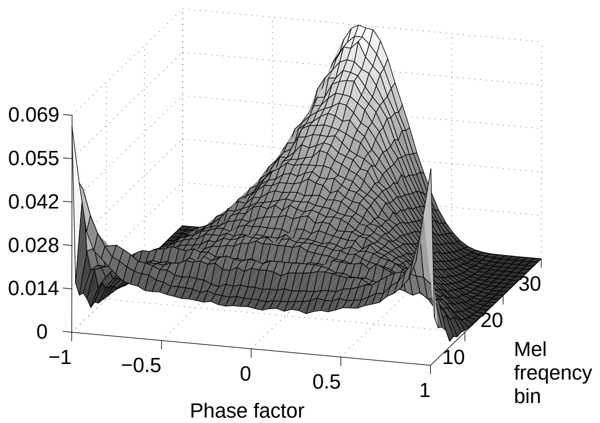

Figure 2: *Empirical distribution of the phase factors α_i.*

Another approach is to approximate the empirical distribution by simulation. Assuming the relative phase is uniformly and independently distributed in the power spectral bins, the vector θ of relative phases can be simulated by drawing samples $\theta^{(j)}$ with $\theta_k^{(j)} \sim \mathcal{U}_{[0,\pi]}$, the uniform distribution on $[0, \pi]$. The corresponding phase factors can be approximated by multiplying the $\theta^{(j)}$ with the Mel filterbank matrix and then taking the cosine of the components, respectively:

$$\alpha_i^{(j)} \approx \cos\left(\sum_k W_{i,k}\theta_k^{(j)}\right).$$

The resulting simulated empirical distribution showed a high degree of similarity to the empirical distributions obtained in experiments. In particular, the empirical distribution did not vary strongly for different noise types. These claims are supported by the plots in figure 3, which show the variance and kurtosis for the simulation as well as for tank (leopard) and factory noise.

Figures 2 and 3 further show that the phase factor distribution is definitely not Gaussian in the lower frequency bins. Note that a Gaussian distribution corresponds to a kurtosis[1] of 3.0, a uniform distribution to a kurtosis of 1.8.

2.3. Averaging over the phase factors

Making use of the phase factor distribution we can now average (5), (6) and (7) with respect to $\boldsymbol{\alpha}$. In the following we show how these averages can be computed with Monte Carlo integration and how they can be stored in a table to obtain an efficient runtime implementation through table lookups. Defining

$$\begin{aligned}\tilde{f}(\mathbf{z},\boldsymbol{\alpha}) &\triangleq \log\left(\sqrt{\mathbf{1} + (\boldsymbol{\alpha}^2 - \mathbf{1})e^{\mathbf{z}}} - \boldsymbol{\alpha}\sqrt{e^{\mathbf{z}}}\right)^2 \\ \tilde{g}(\mathbf{z},\boldsymbol{\alpha}) &\triangleq \mathit{log}\left(\mathbf{1} + e^{\mathbf{z}} + 2\boldsymbol{\alpha}\sqrt{e^{\mathbf{z}}}\right)\end{aligned}$$

[1] We compute the kurtosis as the fourth central moment divided by the variance square.

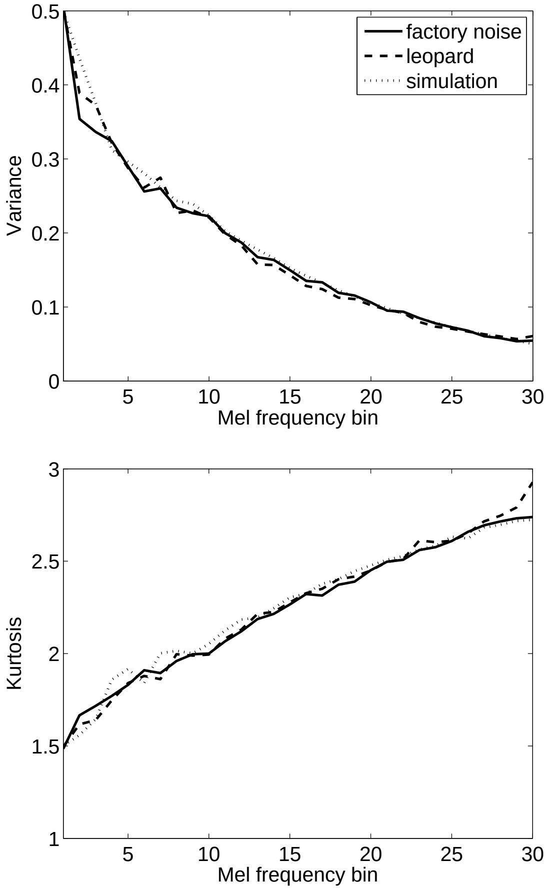

Figure 3: *Variance and kurtosis of the distribution of phase factors for the simulation, tank (leopard) and factory noise*

f, f' and g can be written

$$\begin{aligned} f_\pm(\mathbf{y},\mathbf{n},\boldsymbol{\alpha}) &= \mathbf{y} \pm \tilde{f}(\mathbf{n}-\mathbf{y},\boldsymbol{\alpha}) \\ f'_\pm(\mathbf{y},\mathbf{n},\boldsymbol{\alpha}) &= \tilde{f}'_\pm(\mathbf{n}-\mathbf{y},\boldsymbol{\alpha}) \\ g(\mathbf{x},\mathbf{n},\boldsymbol{\alpha}) &= \mathbf{x} + \tilde{g}(\mathbf{n}-\mathbf{x},\boldsymbol{\alpha}) \end{aligned}$$

where $\tilde{f}'_\pm(\mathbf{z},\boldsymbol{\alpha}) \triangleq f'_\pm(\mathbf{y},\mathbf{n},\boldsymbol{\alpha})$ results from equation (6) by replacing $(\mathbf{n}-\mathbf{y})$ by $\mathbf{z}$ in the $\mathbf{u}$ and $\mathbf{v}$ terms. Now $f_\pm$, $f'_\pm$ and g can be averaged with respect to $\boldsymbol{\alpha}$ by averaging the $\tilde{f}(\mathbf{z},\boldsymbol{\alpha})$, $\tilde{f}'_\pm(\mathbf{z},\boldsymbol{\alpha})$ and $\tilde{g}(\mathbf{z},\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$. This can be done by Monte Carlo integration, i.e. by using a set $\mathcal{M}$ of samples from the empirical distribution of phase factors. Assuming the phase factors are stochastically independent the average of $\tilde{g}(\mathbf{z},\boldsymbol{\alpha})$ can be approximated as

$$\mathrm{E}[\tilde{g}(z_i,\alpha_i)] \approx \frac{1}{|\mathcal{M}|} \sum_{\boldsymbol{\alpha}\in\mathcal{M}} \tilde{g}(z_i,\alpha_i). \tag{8}$$

For $\tilde{f}_+(\mathbf{z},\boldsymbol{\alpha})$ and $\tilde{f}_-(\mathbf{z},\boldsymbol{\alpha})$ we must consider only those $\boldsymbol{\alpha} \in \mathcal{M}$ that result in valid solutions, characterized by $u_i \geq 0$ and $(\sqrt{u_i} - v_i) \neq 0$. This implies that for each Mel frequency bin i we work on a subset $\mathcal{M}_i \subseteq \mathcal{M}$ of samples, and the average of $\tilde{f}_\pm(\mathbf{z},\boldsymbol{\alpha})$ can be approximated as

$$\mathrm{E}[\tilde{f}_\pm(z_i,\alpha_i)] \approx \frac{1}{2|\mathcal{M}_i|} \sum_{\alpha_i\in\mathcal{M}_i} \tilde{f}_+(z_i,\alpha_i) + \tilde{f}_-(z_i,\alpha_i) \tag{9}$$

if we average the two solutions, $\tilde{f}_+(\mathbf{z},\boldsymbol{\alpha})$ and $\tilde{f}_-(\mathbf{z},\boldsymbol{\alpha})$. The phase-averaged Jacobian can be obtained in the same way, i.e. by averaging $\tilde{f}'_\pm(\mathbf{z},\boldsymbol{\alpha})$ for those $\boldsymbol{\alpha}$ that result in solutions with respect to $\tilde{f}_\pm(\mathbf{z},\boldsymbol{\alpha})$:

$$\mathrm{E}[\tilde{f}'_\pm(z_i,\alpha_i)] \approx \frac{1}{2|\mathcal{M}_i|} \sum_{\alpha_i\in\mathcal{M}_i} \tilde{f}'_+(z_i,\alpha_i) + \tilde{f}'_-(z_i,\alpha_i). \tag{10}$$

As on-line Monte Carlo integration is computationally expensive we store precomputed averages in a table. Then, at runtime, the phase-averaged model can be implemented through table lookups:

$$\begin{aligned} f(y_i,n_i) &\approx y_i + \mathrm{E}[\tilde{f}(n_i-y_i,\alpha_i)] \qquad (11)\\ df(y_i,n_i)/dy_i &\approx \mathrm{E}[\tilde{f}'(n_i-y_i,\alpha_i)] \qquad (12)\\ g(x_i,n_i) &\approx x_i + \mathrm{E}[\tilde{g}(n_i-x_i,\alpha_i)] \qquad (13) \end{aligned}$$

In addition to the averages we store the probabilities

$$c_i(z_i) \triangleq \frac{|\mathcal{M}_i|}{|\mathcal{M}|} \tag{14}$$

of $\tilde{f}_\pm(z_i,\alpha_i)$ being a valid solution, which will be needed in the upcoming section.

3. Application to particle filter based sequential noise compensation

Recently particle filters [7, 12, 13] have been used to track the noise portion of noisy speech features in order to compensate for distortions introduced by the noise. The particle filter approach can be described as keeping a set of noise hypotheses that are propagated according to a dynamical system model. The dynamical system model can be *autoregressive* as in [7, 13], *dynamical autoregressive* as described in [14], or a transition based on Polyac averaging and feedback [12]. In addition to being propagated forward in time, the noise hypotheses are pruned at each time instant t by multiplying hypotheses that have a high relative likelihood and removing hypotheses that have a low relative likelihood. Thereby the likelihood is evaluated as

$$p_y(\mathbf{y}_t|\mathbf{n}_t^{(j)}) = p_x(f(\mathbf{y}_t,\mathbf{n}_t)) \left|\det\left(\frac{df(\mathbf{y}_t,\mathbf{n}_t)}{d\mathbf{y}_t}\right)\right|, \tag{15}$$

where p_x is an auxiliary clean speech Gaussian mixture model and where $f(\mathbf{y}_t,\mathbf{n}_t) = \log(e^{\mathbf{y}_t} - e^{\mathbf{n}_t})$ is the standard model equation solved for $\mathbf{x}$. The multiplication by the absolute Jacobian determinant is due to the transformation of the probability density from p_y to p_x.

The phase-averaged model can be integrated by replacing the standard model equation and its Jacobian by (11) and (12). However, we further have to multiply (15) by $\prod_i c_i(n_{t,i} - y_{t,i})$ to compensate for the fact that different noise hypotheses result in a different number of solutions of (5) with respect to the phase factor $\boldsymbol{\alpha}$. Clean speech estimation has to be modified accordingly if the *straight-forward approach* [13] is used:

$$\mathrm{E}\left[\mathbf{x}_t|\mathbf{y}_{1:t}\right] \approx \sum_j f(\mathbf{y}_t,\mathbf{n}_t^{(j)}) p_y(\mathbf{y}_t|\mathbf{n}_t^{(j)}). \tag{16}$$

4. Experiments

In order to compare the performance of the phase-averaged and the standard model, we performed a simulation in which known clean speech and noise signals were mixed at 0 dB.

Clean speech came from the test set of the *multi-channel Wall Street Journal audio visual* (MC-WSJ-AV) corpus [10]. As a disturbance, we chose the factory1 noise from the NOISEX-92 database [11]. The resulting noisy speech features were enhanced with either the standard equation or the phase averaged model, using the perfectly known noise sequence. As the true clean speech features were known, the mean squared error of the estimated clean speech features could be computed. It was 477.4 per frame for the standard equation, but only 153.9 for the phase-averaged model, clearly showing the superiority of the phase-averaged model.

In the experiments reported below, the feature extraction of our ASR system was based on *Mel frequency cepstral coefficients* (MFCC)s, where a triangular Mel filterbank was used. After *cepstral mean subtraction* (CMS) with variance normalization, 15 consecutive frames of 13-coefficient MFCCs were concatenated and subsequently reduced by *linear discriminant analysis* (LDA) to obtain the final 42-dimensional feature. The decoder used in the experiments is based on the fast on-the-fly composition of weighted finite-state transducers (WFSTs), as described in [14, §8]. It produces word lattices which are then optimized with WFST operations as described in [15]. The triphone acoustic model was trained with 30 hours WSJ0 and 12 hours WSJCAM0 data, resulting in 1,743 fully continuous codebooks with a total of 70,308 Gaussians. The auxiliary clean speech GMM with 128 mixture components was trained on the same data set.

We evaluated the phase-averaged particle filter (pa-PF) described in Section 3 through a series of automatic speech recognition experiments. These experiments were conducted for speakers 16-25 of the MC-WSJ-AV corpus. The corresponding 352 utterances were artificially contaminated by adding various noise-types from the NOISEX-92 database at different *signal-to-noise ratios* (SNR)s. Table 1 shows the results in comparison to the baseline (no PF) as well as to the particle filter described in [13] (std-PF), for a first, unadapted speech recognition pass as well as for an adapted pass using *maximum likelihood linear regression* (MLLR) [16] *feature space adaptation*. In all cases investigated, the pa-PF performed best. With factory2 noise it achieved a relative improvement of 15.7%, 17.2% and 7.0% over the baseline on the unadapted pass, which reduced to 9.3%, 5.4% and 3.1% on the adapted pass. The std-PF failed to improve over the baseline in more than half of all cases, which was especially severe on the adapted pass. It performed even worse when the *fast acceptance test* [13] was not used. This is due to the problems with sample attrition and dropouts reported in [13]. The pa-PF completely overcomes those problems without having to resort to an acceptance test.

		10 dB		15 dB		20 dB	
noise	PF	1st	2nd.	1st	2nd	1st	2nd
fac	none	63.7	34.5	55.2	25.8	47.4	22.4
	std	65.8	44.3	52.0	32.1	46.8	27.4
	pa	53.7	31.1	45.7	24.4	44.1	21.7
ops	none	66.3	46.7	48.2	32.1	42.6	23.0
	std	64.0	47.3	48.9	30.7	41.8	23.9
	pa	63.2	44.3	46.7	29.8	40.4	22.1
eng	none	81.0	56.0	70.3	36.9	64.1	27.3
	std	84.6	71.7	72.3	51.5	64.0	36.6
	pa	73.3	48.4	63.1	33.0	57.0	24.2

Table 1: *Word error rates* (WER)s for the particle filter used in [13] (std), the particle filter with the phase-averaged model (pa) and the baseline (none), each for the unadapted (1st) and adapted (2nd) pass. For clean speech the WER was 41.9% and 20.5% respectively. The added noise was either factory2 (fac), destroyer-ops (ops) or destroyer-eng (eng) noise.

5. Conclusions

The experiments confirm that the proposed model is considerably superior to the standard model. For clean speech estimation the mean squared error of the phase-averaged model has been shown to be a third that of the standard model. Moreover, it has been shown to improve particle filter based speech feature enhancement. Maybe other approaches can benefit from that too.

6. References

[1] A. Acero, *Acoustical and Environmental Robustness in Automatic Speech Recognition*. Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, 1990.

[2] M. Gales and S. Young, "Robust continuous speech recognition using parallel model combination," *IEEE Transactions on Speech and Audio Processing*, vol. 4, pp. 352–359, Sep. 1996.

[3] P. Moreno, B. Raj, and R. Stern, "A vector Taylor series approach for environment-independent speech recognition," *Proc. ICASSP*, vol. 2, pp. 733–736, May 1996.

[4] N. S. Kim, "Nonstationary environment compensation based on sequential estimation," *IEEE Signal Processing Letters*, vol. 5, no. 3, pp. 57–59, Mar. 1998.

[5] ——, "IMM-based estimation for slowly evolving environments," *IEEE Signal Processing Letters*, vol. 5, no. 6, pp. 146–149, Jun. 1998.

[6] K. Yao and S. Nakamura, "Sequential noise compensation by sequential Monte Carlo method," *Advances in Neural Information Processing Systems*, vol. 14, 2002.

[7] B. Raj, R. Singh, and R. Stern, "On tracking noise with linear dynamical system models," *Proc. ICASSP*, May 2004.

[8] L. Deng, J. Droppo, and A. Acero, "A Bayesian approach to speech feature enhancement using the dynamic cepstral prior," *Proc. ICASSP*, vol. 1, pp. 829–1 – 829–32, Mar. 2002.

[9] ——, "Enhancement of log mel power spectra of speech using a phase-sensitive model of the acoustic environment and sequential estimation of the corrupting noise," *IEEE Transactions on Speech and Audio Processing*, vol. 12, pp. 133–143, Mar. 2004.

[10] M. Lincoln, I. McCowan, J. Vepa, and H. Maganti, "The multi-channel wall street journal audio visual corpus (mc-wsj-av): specification and initial experiments," *Proc. ASRU*, Dec. 2005.

[11] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, pp. 247–251, 1993.

[12] M. Fujimoto and S. Nakamura, "Particle filtering and Polyak averaging-based non-stationary noise tracking for ASR in noise," *Proc. ASRU*, Dec. 2005.

[13] F. Faubel and M. Wölfel, "Overcoming the vector Taylor series approximation in speech feature enhancement - a particle filter approach," *Proc. ICASSP*, Apr. 2007.

[14] M. Wölfel and J. McDonough, *Distant Speech Recognition*. New York: John Wiley & Sons, 2008.

[15] A. Ljolje, F. Pereira, and M. Riley, "Efficient general lattice generation and rescoring," *Proc. Eurospeech*, pp. 1251–1254, Sep. 1999.

[16] C. Leggetter and P. Woodland, "Maximum likelihood linear regression for speaker adaptation of the parameters of continuous density hidden markov models," *Computer Speech and Language*, vol. 9, pp. 171–185, 1995.