March 19, 2013

Dear Prof. Ono,

please find attached the revised manuscript T-ASL-03898-2012 "Regularization for Partial Multichannel Equalization for Speech Dereverberation" by Ina Kodrasi, Stefan Goetze, and Simon Doclo.

We believe that we have addressed all remaining concerns of Reviewer 3 by implementing the following changes to the manuscript:

1. *Page 8, left column, last Sentence: "From Fig. 4b ... outperforms all other regularized techniques ..." should be changed to "From Fig. 4b ... outperforms all other investigated regularized techniques ...". The original claim would require a mathematical proof to be justified. Similarly for page 9, left column, line 35-36.*
   Changed in the revised manuscript.

2. *Page 10, right column, line 34: Please briefly comment why the error is significantly larger with 450ms compared to the other scenarios.*

   We believe that the actual value of the regularization parameter is irrelevant, as long as it yields a high perceptual speech quality. In the considered simulation, while the normalized mean square error between the optimal and automatic regularization parameter is 0.64, the average performance degradation in terms of PESQ is only 0.11, which was already mentioned in the manuscript. Hence, **we have not added any further remarks about this issue in the revised manuscript.**

   The remaining comments were considered optional by the reviewer. However, we have still implemented some changes in the manuscript, addressing comments 3 and 4.

3. *Thank you for adding additional comments about how the PESQ score is calculated. Now I understand why the score is increasing with increasing L_d. From the reviewer's point of view, the fact that the PESQ score is monotonically increasing with increasing L_d (even for the unprocessed microphone signal) also reveals the perceptual limitation of the way PESQ is used here (Now I understand why you do not want to increase above $50ms$). This means that the increase in PESQ score for regularized P-MINT with increasing L_d has two reasons:*
   a) with increasing L_d the reference signal becomes more and more similar to the unprocessed microphone signal
   b) increased robustness against channel estimation errors (and background noise).
   It would be good if you could add a brief discussion so that these two effect are clear to the readers. Alternatively, you could consider keeping L_d fixed (e.g. $L_d = 50ms$) for calculating the PESQ-reference signal while changing L_d for the equalization algorithms. In this way, effect a) could be avoided.

   In order to clarify in more detail how the PESQ score is calculated, **we have reformulated the paragraph where PESQ is introduced (page 7)** to the following:

   The perceptual speech quality of the output signal $\hat{s}(n)$ is evaluated using the objective speech quality measure PESQ [33], which generates a similarity score between the output signal and a reference signal in the range of 1 to 4.5. It has been shown in [34] that measures relying on auditory models such as PESQ exhibit the highest correlation with

subjective listening tests when evaluating the quality of dereverberated speech. The reference signal employed here is $s(n) * h_1^{\mathrm{d}}(n)$, i.e. the clean speech signal convolved with the first part of the true first RIR (which is different for each value of the desired window length L_d). It should be noted that with increasing L_d, the reference signal becomes more similar to the unprocessed microphone signal.

4. *From the reviewer's point of view, the benefit of the regularization for both channel estimation errors and background noise would be clearer if background noise was included in equations (35)-(37). Confusion among the readers could be avoided by adding a sentence that explains that the design of the filter is done for ideal conditions (i.e. e=0 and v=0) but it has to be made robust against channel estimation errors and background noise. With the noise included in the equations, referring to (35)-(37) in Section V.D would become clearer for the readers. Therefore, I recommend that the authors consider including the noise term into these equations.*

   Equations (35)-(37) have been changed in the revised manuscript by also including the noise term. **The following sentence has been added on page** 5 in order to clarify that the noise term is disregarded when designing reshaping filters:

   As previously mentioned, acoustic multichannel equalization techniques generally design reshaping filters disregarding the presence of noise, hence in the following it is again assumed that $v_m(n) = 0$.

5. *The newly introduced discussion about how the channel estimation error is simulated clarifies the choice of the method used. However, the values $E_m = -33dB$ and $E_m = -15dB$ used in the simulations still appear arbitrary. It is still unclear to the reader whether these are values that are typically encountered in realistic applications. Since these values are very important for assessing whether the methods are useful in practice, it would be very beneficial for the readers if the authors could comment on how realistic these values are. This could be done by citing BSI papers where similar values have been achieved.*

   We have not implemented any changes with respect to this issue in the revised manuscript.

6. *Setting the delay $\tau = 0$ does not seem to be the best choice. It has been shown in [15] that increasing τ to values larger equal than the delay of the true impulse responses can help to reduce the filter norm. Therefore, (if not already done by the authors) it could be interesting to increase τ in future simulations.*

   We agree with the reviewer that incorporating a delay is beneficial for acoustic multichannel equalization techniques. The same simulations were conducted for $\tau > 0$, and while marginal improvements in the performance of individual techniques are obtained, the conclusions comparing the different techniques remain the same as the ones derived in the manuscript. Due to space constraints, **we have not implemented any changes** with respect to this issue in the revised manuscript.

Best regards,
Ina Kodrasi