List of Changes

Ref.: Ms. No. JASA-00590
In-Ear Microphone Speech Quality Enhancement Via Adaptive Filtering and Artificial Bandwidth Extension
The Journal of the Acoustical Society of America

Reviewer #1

Suggested Changes	Changes Made
First, the author should mention at some point that the computation was an off-line simulation with a given computer	In second to last paragraph of the intro, authors have added that results are from an off-line simulation.
Abstract: "band-width", remove space.	Changed to bandwidth
Abstract: "pro- posed techniques", please correct.	Changed to proposed
Abstract: It could be interesting to give a result (i.e. number) in the absract, p-value for the "statistically significant improvement" for instance?	Added a phrase indicating the p values for both denoising and bandwidth extension at the end of the abstract
Sec. B. Predicted Quality, the authors write "As shown in previous work, ...". Please add a reference at the end of the sentence.	Added two citations (Bou Serhal et. 2013 and Bouserhal et. al 2015) at the end of the sentence
Sec. IEM Noise Reduction, NLMS Filtering, at least add a general reference to adapttive fitlering or active noise control.	A citation to general adaptive filtering is already given (Manolakis et al. (2005))
Sec. IEM Noise Reduction, NLMS Filtering, the authors mention: "H(z) is the true transfer function of the earplug", please explicitely state the input and output of this single-channel transfer function.	Changed to: Here, $H(z)$ is the transfer function of the earplug expressed as the ratio between the output signal captured by the IEM over the input signal captured by the OEM. $\hat{H}(z)$ is the estimated earplug transfer function.
Sec. IEM Noise Reduction, NLMS Filtering, the authors mention: "The adaptive filter order …", before introducing the order, the fact that a FIR filter is used should de mentionned. Also the selection of the order should be discussed, based on average propagation delay in H(z) multiplied by a given factor? How this was selected?	The following sentence was added before: A finite impulse response (FIR) filter of order 160 is used. The order of 160 is chosen as it is the smallest order than can accurately reproduce the transfer function of the earplug with a delay of 20 ms, which is an undetectable delay for speech as shown by Lezzoum et. al 2016.

Sec. IEM Noise Reduction, NLMS Filtering, equation numbers? Check throughout the paper and check JASA guidelines.	Equation numbers have been added
Sec. IEM Noise Reduction, NLMS Filtering, since some variables are scalars and other are vectors, a notation for vectors should be introduced. Also, the size of each vectors should be explicitly stated.	Vectors in equation 1 were rendered in bold and their dimensions were stated right after.
Sec. IEM Noise Reduction, NLMS Filtering, after the three equations, the authors should clarify the unconventional use of the block diagram. What are the meaning of the dashed lines? Why putting block inside block with dashed lines? Why the arrow entering the double H(z) block is not, as usual, connected to the actual H(z) block? Generally speaking, the reviewer finds confusing and useless the use of dashed lines, if they are relevant and important, the authors should clarify why, otherwise, please simplify to a conventional block diagram representation.	Block Diagram was changed to reflect conventional block diagrams
Fig. 3, the caption should provide more details, what is the meaning of the dashed signal flow? What is the meaning of dashed blocks? Why put thick blocks in dashed blocks, etc.?	Block diagram was changed. No dashed lines present.

Sec. Offline Transfer Function Identification, the identification process should be further explained and described. How H(z) is obtained? How many averages? Using windows? Is it based on adaptive system identification using the block diagram shown earlier? Also, the authors should explain why they use white noise for system identification? Indeed, although white noise is often used for system identification in active noise control, it typically performs poorly for audio applications, is this a problem? Why other methods such as MLS (max length sequence) or log-swept sine method have not been used instead? Indeed, white noise system identification suffers from various issues: 1) Any non-linear behavior in the system will appear as spreaded noise in the resulting impulse response, 2) SNR is typically poor in comparison with MLS or log-swept-sine method. Also, if MLS reduces the SNR issue, it also suffers from any non-linearity in the system. The log-swept-sine allows for better SNR and the possible extraction of any higher-order non-linear behavior. See for instance Farina (AES Convention 2000 and 2007) or Muller and Massarani (JAES 2001). If the white noise identification was not, according to the authors, a problem, this should be explicitely stated so in comparison with the other more common methods. Any presence of non-linearity should be discussed briefly. If the authors rely on white noise identification using adaptive system identification using the already implemented adaptive filter, this should be explicitely stated since it would reduce the aforementionned comments. Reference to the used system identification method should be included.	The following sentence was added in Sec. Offline TF: The field microphone-in-real-ear (F-MIRE) technique was used to obtain the transfer function of the earplug as it is not susceptible to disturbances caused by physiological noise as was shown by Voix et. al (2009). Reply to reviewer: No adaptive system identification was used. The F-MIRE approach for earplug assessment is not susceptible to disturbances caused by physiological noise. It is also features a coherence indication that would capture any nonlinearity in the system and ensures proper TF estimation. In the envisioned application, offline identification could be done using ambient broadband noise which is accessible in noisy industrial environments unlike sweep sine.
Sec. The Adaptation Process, the authors mention: "After completion of the two second identification stage the vector of filter weights over the entire index of time", this somehow clarify the previous comments. It suggests that adaptive system identification was used? IF yes, please mention in the previous section.	See above comment

Sec. The Adaptation Process, the max value of a vector is the infinit norm, why not use this notation? Also, add equation number.	Changed and fixed
Sec. The Adaptation Process, how the range 1.01 to 1.2 was decided? The reviewer understand that Tg should be larger than one (meaning amplification of the filter coef from previous step), but how the 1.2 was set or identified?	The following sentence was added to the section: The upper limit of 1.2 was chosen empirically based on results up to 1.5 revealed a local maximum at $T_g=1.06$.
Sec. The Adaptation Process, the reference to Section D should be written Section II.D.	JASA Latex template was used to take care of formatting issues and section naming and numbering.
Sec. The Adaptation Process, when the authors mention w(n-f_s), they should recall that f_s is the sampling frequency. The reviewer is not sure if f_s was defined earlier. Please check.	The following sentence was added after w(n-F_s): where f_s is the sampling frequency.
Sec. The Adaptation Process, the variable k that appears in Fig. should be defined.	The variable k was unnecessary and was removed
Sec. The Adaptation Process, the last paragraph describes the purpose of the next section. It should be placed in the IEM Bandwidth Extension section.	Moved to the next section
Sec. IEM Bandwidth Extension, please correct and reformulate "has been very well studied". JASA does not recommend the use of "very" or "well-known". Also add reference to support such sentence.	Changed to thoroughly and added 3 references on artificial bandwidth extension
Sec. IEM Bandwidth Extension, check for useless line breaks at the begining of section. Also check justification, the text looks centered?	JASA Latex template was used to take care of formatting issues and section naming and numbering.
Sec. IEM Bandwidth Extension, please add a reference to whitening filtering using LPC.	Reference added
Sec. IEM Bandwidth Extension, the purpose of the whitening filter should be explicitely stated. If it is fairly obvious that the x^3 operation will enhance higher-order harmonics, the whitening process might not be obvious for every reader.	The following sentence has been added for clarification: To reach an excitation signal similar to that extracted from a wideband speech signal, the upsampled signal is filtered by the whitening filter using the coefficients of an LPC analysis \citep{Valin2000}

Sec. IEM Bandwidth Extension, the double use of filtering at 1.8 kHz to combine the 160-1800 range with the Bandwidth Extension operate as a crossover. However, the use of Butterworth filter is not recommended for crossover (such as for bass management). The authors should discuss this. Normally, the use of Butterworth filter for a crossover will induce a bumb in the summed response at the crossover frequency. Phase can also be altered in the cross-over region. Normally, cascaded Butterworth filters are used for crossover in order to obatain Linkwitz-Riley filter (see Lr4 based on 2-nd order Butterworh filters) (https://en.wikipedia.org/wiki/Linkwitz%E2%80%93Riley_filter). This should be further motivated, why the authors did not rely on conventional cross-over filtering? Since this can be easily implemented with cascaded biquad, this should not be an implementation issue? Is the bump caused by the use of Butterworth is not a problem for audio quality, if this is the case. Please mention.	The authors noticed some errors in this part of the manuscript that have been corrected and will answer reviewer concerns. The block diagram as well as the text describing it have been updated. In fact the choice of an L-R4 was done empirically through optimization of the POLQA value as well as the perceived quality of the signals.
Sec. Performance Evaluation. The use of an on-line test with limited control on the playback devices, levels, SNR, on the listener side should be further motivated. Also, how many participants were included in the test, it appears latter, but it could mentionned at this point. Does the tests included some verification of listener performance? Such as repeatability check and repeatability-based rejection of participants? What is the approximate duration of the listening test? Too long listening tests are typically not recommended, above 60 minutes.	Following sentence was added: Instructions emphasized that a comfortable volume should be set at the beginning of the test and should not be changed throughout. No assumptions were made on the participants' hearing abilities and no repeatability checks were performed. A total of 42 participants took part of the test that should have taken less than 30 minutes. After review of the results two subjects were rejected from the pool as it appeared they did not understand the nature of the test (low scores for hidden reference).
Fig. 9, please correct the wide ticks for the x and y axis.	Corrected

Sec. Performance Evaluation, the authors mention "The cumulative distribution of the difference between the clean and the noisy POLQA MOS-LQO scores of the IEM signals as well as the difference between the clean and the noisy POLQA MOS-LQO scores of the OEM signals are shown in Fig. 10." Please provide an interpretation of Fig. 10. Otherwise, if not discussed, please remove Fig. 10.	Response to Reviewer: An accidental start of new paragraph may have caused this confusion. The discussion of Fig 10 is actually the sentence that follows.
Sec. Performance Evaluation, the authors mention: "The completely degraded signal captured by the OEM in noisy conditions can be utilized to denoise the relatively superior quality speech signal captured by the IEM, as described in Section C." please correct for Section II.C. (if this is case).	Corrected.
Sec. IEM Noise Reduction, please write "threshold chosen as T_g = 1.06, the ...", i.e. remove the first "," before "T_g".	Corrected.
Sec. IEM Noise Reduction, correct "band-width" for "bandwidth" or "band-width", but remove space. Check throughout the paper.	Corrected.
Sec. IEM Noise REduction, the selection of mu and epsilon seems rather arbitrary. Any preliminary tests were performed? Convergence or divergence was observed? Please give more details.	Sentence was modified for clarity to: Since the effect of μ and ε showed no major changes in performance within a specific window ($0.4 \leq \mu \leq 0.8$ and $0.0001 \leq \varepsilon \leq 0.01$), for the denoising phase $\mu = 0.7$ and $\varepsilon = 0.001$ were chosen empirically.
Fig. 14, the figure should include a color range with a mention of the amplitude units, dB? dB ref 1? Does all the four spectrograms share the same color range? This should be corrected.	The figure has been changed to colour with the heat map next to each of the figures and clarification that the magnitude is in dB FS was mentioned in the caption
Tab. 2, check "be- tween", correct.	Corrected.
Tab. 2, the meaning of N, NS, BWE, etc. should be reminded in the caption. As it is currently done in Tab. 1.	description added

Sec. Subjective Evaluation. Before provinding a concluding affirmation such as "The subjective results from the MUSHRA listening test confirm the objective trends found using POLQA.", the authors should neutraly introduce the results and let the reader creates its understanding of the results. An then state "The subjective results from the MUSHRA listening test confirm the objective trends found using POLQA." Please rewrite this section accordingly. For instance, give the number of participant before stating such conclusion, etc.	Paragraph was changed by putting the first sentence at the end of the paragraph.
Fig. 18 should also include a description of the * and ** and accolades in the caption. Explanation and discussion of these *, ** and {} should also appear in the text. Ref to tab. 4. could also be used in the Fig. 18 caption.	The ** were removed from the figure for simplicity and the reader is directed to Table 4 for statistical significance
Sec. Subjective Evaluation, please write "log-spectral" in place of "log- spectral".	Corrected.
Sec. Subjective Evaluation, please provide more details on the log-spectral distance computation, it was based on a short signal? How the freq domains signals s_i(w) are obtained? Welch, FFT size, etc., averaging? Sample length, etc. Does the LSD was averaged between FFT frames, or the LSD is based on the Welch average spectrum, etc.?	The following sentence was added for clarity. The LSD was calculated for 25\,ms long, hamming windowed, frames with a 15ms overlap.
Sec. Subjective Evaluation, the LSD involves an integration over w from -w to w? Please correct the interval, from -pi to pi normalized freq.?	Corrected to -pi to pi.
Sec. Discussion, please correct "in-dependent".	Corrected.
Sec. Discussion, please replace "well studied" by a more appropriate sentence.	well studied was removed.
Sec. Discussion, please rewrite "it is relevant as well" more approriately.	Changed to: “…it is also relevant..”

Reviewer #3

Suggested Changes	Changes Made
On pdf-page 21, you state that significance was tested using ANOVA. Did you check if the data is normally distributed or not? please include this in the manuscript.	Yes, data is normally distributed. the sentence was added to the objective evaluation portion of the results: All results were verified to be normally distributed before the ANOVA tests.
Often in the text, variables are introduced in commas, like "tissue conduction, s(n), with minimal effects of noise". Omit the commas. There are several instances of this issue in the text.	Instances were found and corrected.
The Oxford or serial comma is required according to the JASA style guide, stating that "a JASA manuscript would refer to the "theory of Rayleigh, Helmholtz, and Kirchhoff" rather than to the "theory of Rayleigh, Helmholtz and Kirchhoff"". Please revise!	Instances were found and corrected.
several commas missing, to separate introductory clause from the main clause...	Instances were found and corrected.
Treat formulas like "words" in the text, meaning that punctuation should be taken into account. For example, on pdf-page 16, a comma is missing after the equation: "Therefore once [formula], ..."	Corrected.
The equation numbers are missing?!...or is this just an issue of this review-template?	Equation numbers were missing, but have now been added.
Numbers and their units should be separated with a half-space (latex: \,), like 20\,kHz, 10\,dB, or 90\,\%. Please revise.	Switched to Latex and applied reviewer's comment to all numbers with units.
In the text, please refer to Figures and Tables like in their captions, i.e., Figure X and not Fig. X. There are several instances of this issue in the text. According to the JASA-style guide, tables should be references in the text using capital letters, like TABLE III.	JASA Latex template was used to take care of formatting issues and figure naming and referencing.

Suggested Changes	Changes Made
All in all, 17 Figures are in this article despite the style guide states that no more than 12 figures should be used! I think it is possible to reduce their amount as some figures seem to provide no valuable extra-information or are needed to support the claims in the text, like Figures 5, 11, or 14.	Figures 5 and 11 were removed based on reviewer recommendation. However, authors feel figure 14 is relevant and have kept it in the manuscript. Current JASA guide for authors limits to 20 figures and not 12.
Some formulas in the figures seem quite small, like d(n)=s(n)+n(n)+[i can't read this](n) in Figure 3(b). Figure 13 provides a lot of info. I fear they are unreadable in the two-column version?!	Figur 3(b) was changed.
Some figures are in color which is not necessary. The 3 lines in Figure 2 can be set as dotted, dashed and solid lines. Change the red color in Figure 4. Figure 11 would be sufficient in grey scale.	All figures can be printed in black and white and still be read clearly. However, authors chose to keep colour figures for online readers of JASA.
Several figures are "pixelated"	No figures should be pixelated in the reviewed manuscript.
The font size within the figures is inconsistent. please revise	All matlab figures have the same font size, however, block diagrams cannot be completely matched.
Table design is not adequate. Omit vertical lines, just use a top rule, mid rule and bottom rule. Refer to the booktabs-Latex package on how to design nice tables.	All tables have been modified according to reviewer comments