PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Noise exposure and auditory thresholds of German airline pilots. A cross sectional study.
AUTHORS	Müller, Reinhard; Schneider, Joachim

VERSION 1 - REVIEW

REVIEWER	Arve Lie STAMI, Oslo, Norway
REVIEW RETURNED	17-Jun-2016

GENERAL COMMENTS	Comments to the paper: General comments: This is a nicely conducted study of hearing in civilian airline pilots. Pilots are highly selected and as expected, hearing of pilots in most studies is found to be similar to or better than hearing of workers not exposed to noise (Lie et al., 2016). The authors find a worse hearing of the left ear compared to the right ear and attributes this to noise from the headset. A better hearing of the right ear is a common finding of population based studies of hearing and a difference of 5 dB in the 3-6 kHz range in men 48-59 was found by Cruickshank in the Beaver Dam study (Cruickshanks et al., 1998). I therefore think that their conclusion should be modified. I, however, do agree with the conclusion that modern headsets with noise reduction function should be preferred for preventive purposes in civil aviation. Specific comments: P1, L 45-49: The conclusion should be modified P2, L 27-30: The left ear is more susceptible than the right one to hearing loss, regardless of noise exposure P4, L 30-37: Manual or automatic testing? P4, L55-59: It is convenient to use ISO 7029 for comparison, but in most studies of occupational groups you find that hearing is a few decibels worse compared to ISO 7029. Therefore I think that ISO 1999 Annex B is more suitable for comparison than ISO 7029 (Engdahl et al., 2005, Lutman & Davis, 1994, Wiley et al., 2001). (I was not aware of the revised unofficial version of the ISO 7029.) P6, L 8-11: You use the term “average audiograms”. Do you mean the average of the left and right ear? P11, L 23: Consider to rephrase to: “More than half of the pilots (53%)” P12, L24-25: Delete the “noise induced hearing loss”. There is a small difference in hearing, but you cannot prove it to be noise induced? P12,L28-32: The reference values are of the better ear. Did you use the better ear from your material? P 13, L24-26: “seems to be more resistant to noise” – should be

	modified to “seems to be more resistant to hearing loss”. P13, L30: Consider to rephrase “noise induced hearing loss” to “hearing loss” References Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E., Klein, R., Mares-Perlman, J. A., & Nondahl, D. M. (1998). Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. The Epidemiology of Hearing Loss Study. Am. J Epidemiol. 148, 879-886. Engdahl, B., Tambs, K., Borchgrevink, H. M., & Hoffman, H. J. (2005). Screened and unscreened hearing threshold levels for the adult population: results from the Nord-Trondelag Hearing Loss Study. Int J Audiol. 44, 213-230. Lie, A., Skogstad, M., Johannessen, H. A., Tynes, T., Mehlum, I. S., Nordby, K. C., Engdahl, B., & Tambs, K. (2016). Occupational noise exposure and hearing: a systematic review. Int Arch Occup Environ Health 89, 351-372, doi:10.1007/s00420-015-1083-5 [doi];10.1007/s00420-015-1083-5 [pii]. Lutman, M. E. & Davis, A. C. (1994). The distribution of hearing threshold levels in the general population aged 18-30 years. Audiology : official organ of the International Society of Audiology 33, 327-350. Wiley, T. L., Torre, P., III, Cruickshanks, K. J., Nondahl, D. M., & Tweed, T. S. (2001). Hearing sensitivity in adults screened for selected risk factors. J Am. Acad. Audiol. 12, 337-347.

REVIEWER	Marion Burgess School of Engineering and Information Technology UNSW AUSTRALIA CANBERRA Australia
REVIEW RETURNED	19-Jun-2016

GENERAL COMMENTS	This paper has the potential to provide a valuable contribution but not in the current form. While it can be understood there are many examples of English expression that need attention and it would benefit from a read by a native English speaker. Example “ pilots of a German airline were analysed” - the pilots were not analysed it was their hearing that was analysed, “all pilots were standardised interviewed” which should be …were interviewed in a standardised manner There are a number of statements that need attention. Example “older pilots have higher threshold levels (presbycusis)” which I think should be older pilots have higher threshold levels than can be explained by presbycusis The Abstract conclusion states the use of headsets with active noise reduction systems will reduce the ….but this has not been demonstrated in the paper – it is mentioned in the final sentence of the conclusion of the paper, and it is mentioned in the discussion but with another option of a communications ear plug. The objective in the abstract does not state the objective – the two

	sentences are statements which I believe no one will disagree, but they are not the objective of the study. Acoustics Measurements – was the 'dummy' a B K dummy head – give details This sentence "By using a middle ear simulator, frequencies above 250 Hz are considered stronger rated than the pure A-rating" needs attention – do you mean that the simulator provides the correct impedance? Cockpit noise – inert reference in list for Hoffman, 2004 The use of impulse response for the ATC signal needs further justification. Do you use it because of the pulses, clicks etc in a comms system. But once you have used it for this you cannot use the AH impulse levels to say there is excess of the workplace exposure level which is based on LAeq,T with just a fast response You could use the peak response value to compare with the LCpeak exposure limit. Is the noise level data provided in Table 2 the equivalent level over the relevant time – ie is the FF the LAeq over the total flight time and is the AH data only over the ATC time? The signal to noise would be more appropriate as the LAeq during the ATC to the FF level at the same time for the sound level in the flight deck changes with the various phases of flight. Also for Signal to Noise it is more appropriate to use the same metric - not different ones. Headset, Fig 3 discusion. The data in Fig 3 does show a worse hearing for left ear use BUT Figure 4 shows that the predominant users of the left ear is a very much smaller sample AND highly skewed >40 years (2 vs 43) – all of the >40 years group has poorer hearing. There is some discussion of this later but it should be mentioned here also. "The free-field sound measurements in Tab. 2 (Hoffmann 2004) in aircraft cockpits show" Im not clear if these authors measured the free field noise levels or if they use the data from Hoffman – which is not in the reference listing The authors do not discuss that the headset provides some noise reduction as it is fixed over the ear. This means that the in ear noise level with no communications will be less than the free field noise level The comments about the use of noise reduction "eliminate the risk" is not justified by the findings in this paper and should be changed to say 'may'

REVIEWER	Christian Giguère Audiology and SLP Program University of Ottawa Canada
REVIEW RETURNED	04-Jul-2016

GENERAL COMMENTS	Measurements of noise exposure from the use of communication

	headsets are not routinely carrying out to determine if occupational regulatory limits are exceeded, and the possible effects of extended use of headsets on hearing health remain largely undocumented. These are important points to research in the context of the civilian airline industry. A major strength of this study is the very large population of pilots (487) that has been evaluated, both in terms of noise exposure assessment and hearing threshold measurements. The study, however, does not clearly identify a specific research question to address and it is not grounded or framed within the existing body of literature on communication headset exposure. There are also many methodological details that are missing or need clarification. In my opinion, there are two important problems with the way the acoustic measurements were conducted: (1) the unusual selection of sound level meter settings (e.g. mixing "impulse" and "fast" time responses) and unspecified measurement duration or averaging, and (2) the apparent lack of spectral correction applied to the communication signals measured in the ear simulator with the acoustic manikin to transform these data into free or diffuse sound field related levels (as per ISO 11904-2 for example). This step is necessary to be able to compare communication signal levels to the regulatory limit and to compute a meaningful SNR. More details below. All in all, it may be best to streamline the paper and focus solely on the hearing threshold results and perhaps on the free field measurements to characterize the cockpit environment. The measurements under the headsets with the acoustic manikin do not appear to be of sufficient quality to merit publication. The authors also do not address the ethical issues or procedures that have been put in place to utilize data from annual employer heath check-ups (audiometric data) in the context of a research study. Was consent obtained? ABSTRACT Item "Objective" simply lists one fact and one occupational requirement; this does not constitute a research objective. Item "Methods" does not mention that noise exposure assessments were also carried out. Item "Conclusions": The last sentence gives the wrong impression that active noise reduction systems were tested in this study and have been found to reduce exposure to 85 dB(A), but this is not the case. The sentence is better left out or reworded as a recommendation for a possible change in practice. INTRODUCTION Page 3 - The introduction is incomplete and too general. The text does not clearly delineate concepts related to auditory fitness for duty ("responsibility", "decision", "performance") from those related to hearing loss prevention ("health check-up", "consequence for their hearing"). The authors also do not discuss whether hearing thresholds are a good proxy for auditory task performance in noise. Perhaps more importantly, there is no reference to prior studies on

	communication headset exposure (or any other publication) and no specific research question or objective is listed. It only says that the "study is a contribution to supplement existing publications and to uncover additional relationships at crucial points". What existing publications? What sort of supplementary information or additional relationships? What do the authors mean by "crucial points"? STUDY POPULATION AND METHODS Page 4, lines 3-6: I suggest renaming this section as "Methods" and renaming the first subsection "Study Population" instead of "collective. Page 4, line 8: Sentence is unclear. Do you mean "… were administered a standardized interview about …" Page 4, line 12-13: "their questionnaires were imprecise". Please elaborate. Page 4, line 20: "the pilots were divided in two age groups". This is incomplete. Table 1 shows data distributed over four age groups. Page 4, lines 28-38: It will also be important to specify the environment where these measurements took place. In a sound proof room? Page 4, line 44-45: Specify which dummy head was used. It is usually called "acoustic manikin" (see ISO 11904-2 terminology). Page 4, line 45-46: "as it meets the requirement". Which requirement? Also specify which sound level meter was used. What it integrating? Page 4, lines 48-49: Unclear. Were two headsets used, one for the pilot and one for the dummy, or only one headset alternating between the pilot and the dummy? Please give more details on the protocol and indicate what precautions were taken to ensure the measurements on the dummy were representative of the pilot exposure. How long in duration were the acoustic measurements? Page 4, lines 49-50: The sentence starting with "By using …" is unclear and may not be necessary. Page 5, lines 6-7: change "witch" for "which", an "arbitrarily" for "arbitrary". Page 5, lines 10-11: Sentence is unclear. Is it that the focus of the paper was placed on individual left-right differences because they do not require age-correction? Page 5, lines 18-19: "… shown in its essential results as a table". Unclear RESULTS Page 6, line 7: It says "civilian aviation companies", but on page 4 it says "a large German airline. Where pilots coming from several airline companies or just one? Page 6, lines 8-11: These lines describing the octave layout should

	be transferred to the Fig.1 caption. They may also be omitted; Fig.1 seems clear enough. Page 6, line 15: I suggest deleting “as expected” in a results section. Interpretation is best left for the discussion. Page 6, lines 17-18: For these individual thresholds, it may be more accurate to assign the largest possible hearing level the audiometer can produce at 16 kHz (or perhaps 5 dB above), instead of considering missing data. It would produce less bias due to equipment limitations. Page 6, line 20: change “cero” for “zero” Page 6, line 21: change “do not” for “does not”. Page 8, line 9: Presumably the Jilek data was only used to correct high frequencies. Which frequencies? Page 8, line 12: change “a less” for “a smaller” Page 9, lines 6-7: “artificial head” is used here while “dummy” is used in page 4. Please use same terminology throughout. Nowadays, “acoustic manikin” is the commonly used terminology (e.g. ISO 11904-2). Page 9: The authors do not detail the communication headsets used. It is important to do so. Are they noise attenuating or not? Are they one-sided or two-sided? Are they circumaural, supra-aural or in-ear devices? Are they the same or different across airplane types? Page 9, lines 9-11: This statement seems incomplete. There is ample data in the literature showing that speech understanding is possible at negative signal to noise ratios. I think what the authors need to say here is that under typical conditions of use in the workplace or airline industry, the speech signals from headsets are significantly above the noise, thereby contributing or becoming the dominant source of sound exposure. A reference would be useful here. Page 9, lines 14-17: “contains impulsive parts of noise”. Unclear statement since the noise is continuous. Are we referring to the speech fluctuations or to noise transients associated with the headsets? Justify the use of an “impulse” time constant to measure speech levels. It is unusual. Is there any reference or standard that can be cited? Page 9, lines 19-41: The reporting of acoustic measurements is unclear. Table 2 presents sound levels using “fast” and “impulse” response settings. Are these values single measurements in time or are they the maximum or average of many such measurements over a substantial period of time on each flight? Why not use an integrating sound level meter? Provide details on the averaging over pilots in each type of airplane? Page 9, lines 43-46: How was this correction factor derived precisely? Why not directly measure headset sound levels in dB(A)f in the first place? It is also not clear if the correction was applied to the number reported in column AH in Table 2. The unit is “dB(A)i”,

	implying these are the actual measurement values and not the corrected values. Please clarify. Page 9, lines 48-51: There are several problems with the calculation of the SNR. As shown in Table 2, the SNR seems to express the difference between a signal level measured with an “impulse” time response setting and a noise level measured in “fast”. It also assumes that the headsets do not provide any noise attenuation. Please specify. Finally, AH speech measurements must first be corrected to represent sound field related levels. The procedure is described in ISO 11904-2 and requires spectral analysis of the signal recorded with the manikin/artificial head. Without this sound field correction, the SNR can be inflated by several dBs. Page 9, lines 56-60: The limit of 85 dB(A) applies to sound field measurements, not manikin/artificial head measurements. The signal level measured in the manikin must first be corrected according to ISO 11904-2 to transform to sound field related levels. Without this correction, the communication sound level can be overestimated by several dBs. Page 11, lines 7-8: This statement seems incomplete. While it is true that the “free” ear will receive a reduce exposure for those pilots listening to only one ear, those who listen with the two ears likely will set the headset volume to a lower value due to binaural effects, leading to a reduced exposure in the two “live” ears compared to the “live” ear of pilots listening to only one ear. The authors mention this aspect themselves on page 13 (lines 36-40). I think it is better to delete “resulting in …. other ear” here and cover all sides of the argument in the discussion. Page 11, line 21: It is suggested to present Fig. 4 before Fig. 3 to improve the logical flow of the paper, i.e. first present the distribution of headset ear preference among pilots, then the audiometric results for each group. DISCUSSION Page 12, line 15: change “round the cero line in” to “around the zero line within” Page 12, lines 30-31: Improve grammar Page 12, line 33: The Hoffmann (2004) reference is not listed on pages 15-16. Also, does is mean that the data dates back from 12 years ago? It is not in itself a problem, provided it is clearly stated in Methods. It is important to do so because the authors refer to “modern headsets” at the end of the paper, so a time frame for the measurements is needed. Page 12, line 38: It should be “none” Page 12, lines 51-55: The authors extracted SNR-50 and SNR-80 values from Fig 2 in Killion et al (1983). This figure is only an illustration of their scoring method, not the results from an actual experiment. The Fig 1 of their paper reports actual data from several studies. The SNR-50 is about 2 dB for normal-hearing individuals, not 7 dB. In any case, the noise used in this (and other) speech test is very different in character then than the noise in an airline cockpit. It may be more relevant to cite studies of communication headset

	exposure in the airline industry or other occupations. Typically the SNR is about 13 dB, which is near the 16 dB found in this study, especially since there is no indication that the data in the current study was corrected according to established methods (ISO 11904-2 or similar), which would lower the SNR. Page 12, lines 58 and rest of paragraph: While it is interesting and relevant to point to ANR systems, why not also point to headsets with passive attenuation properties? Again, the authors have not documented which types of headsets were used by the pilots in the study. Without this information, any discussion on alternative headset solutions becomes rather vague. Page 13, lines 46-47: The authors state “Modern headsets will solve this problem”. Given that the measurements appear to have been made around 2004, are the authors inferring that the headsets used in airlines are different now or are they implying that the commercial airline industry has not caught up with recent headset technology? Please clarify your thoughts. Fig 1 caption: Describe both parts in the same manner. “Part a shows ….. Part b shows”. Table 1: Why are the numbers rounded to the nearest 0.5 dB for the first two age groups and rounded to the nearest 0.1 dB for the two other age groups. Table 2: see above

VERSION 1 – AUTHOR RESPONSE

Comments of Reviewer1: Arve Lie/Responses (yellow)

P1, L 45-49: The conclusion should be modified
Response:
High sound levels of communication with headsets seem to be responsible for the hearing loss at the left ear which is more susceptible to hearing loss.

P2, L 27-30: The left ear is more susceptible than the right one to hearing loss, regardless of noise exposure
Response:
The left ear is more susceptible than the right ear to hearing loss.
We do not agree with the second part oft the sentence. Cruickshanks et al. (1998) did not exclude persons with use of firearms. Many American males like shooting. Prirlä excluded the shooters from his datasets and the differences are smaller as in our pilots.
In our study only pilots which prefered the left ear for communication show remarkable hearing loss. The others show only marginal differences.

P4, L 30-37: Manual or automatic testing?
Response:
Pure tone audiometry was performed by experienced audiologist’s assistants with an audiometer type CA540 from Hortmann GmbH …

P4, L55-59: It is convenient to use ISO 7029 for comparison …
Response:

The values ISO 1999 Annex B in three samples differ a lot and the results depend on the decision for what sample is used. In table 1 the statistical quantils are presented, ready to be compared with the screened dataset of Engdahl et al. (2005).

P6, L 8-11: You use the term “average audiograms”. Do you mean the average of the left and right ear?
Response:
… were presented as average audiograms in age-groups both ears together and the averaged differences between both ears.

P11, L 23: Consider to rephrase to: “More than half of the pilots (53%)”
Response:
Done!

P12, L24-25: Delete the “noise induced hearing loss”. There is a small difference in hearing, but you cannot prove it to be noise induced?
Response:
Done!

P12, L28-32: The reference values are of the better ear. Did you use the better ear from your material?
Response:
No. Thresholds are left and right ear averaged. See caption of Table 1.

Comments of Reviewer2: Mario Burgess/Responses (cyan)

“ pilots of a German airline were analysed” - the pilots were not analysed it was their hearing that was analysed
Response:
The hearing of 487 German pilots was analyzed by audiometry in the frequency range 125 Hz – 16 kHz in age-groups.

“all pilots were standardised interviewed” which should be …were interviewed in a standardised manner
Response:
Done!

Example “older pilots have higher threshold levels (presbycusis)”
Response:
Sentence was deleted.

The objective in the abstract does not state the objective – the two sentences are statements which I believe no one will disagree, but they are not the objective of the study.
Response:
The study examines the hearing thresholds of pilots with respect of ambient noise and communication sound.
The limit of 300 words for the abstract has shortened it with loss of some information.

Acoustics Measurements – was the ‘dummy’ a B K dummy head – give details This sentence “By using a middle ear simulator, frequencies above 250 Hz are considered stronger rated than the pure A-rating” needs attention – do you mean that the simulator provides the correct impedance?

Response:
The measurements were performed with a ½ inch free-field microphone and a head and torso simulator (HATS) Type 4100 with an artificial middle ear Type 4157 of Brüel & Kjær (Denmark). The middle ear simulator conforms to IEC 60318-4, ANSI 3.25 and ITU-T Rec. P.47. The frequency response and impedance is similar to the real human ear.

Cockpit noise – insert reference in list for Hoffman, 2004
Response:
We received the measurement data from Hoffmann in a personal communication without an influence on the measurement setup and performance. Now as reference no. 10 inserted as personal communication.

The use of impulse response for the ATC signal needs further justification. Do you use it because of the pulses, clicks etc in a comms system.
Response:
Table 2 is now distinctly revised. Now the measurements with time constant fast are preferred. Therefore 3 aircrafts must be removed from the table, as the fast measurements were not complete. Only one column with impulse time constant is left in Table 2 to show, that the acoustical signal has impulsive character. The "impulse" and "fast" levels differ up to 6 dB, what is a clear indicator for impulsive sound.

Is the noise level data provided in Table 2 the equivalent level over the relevant time – ie is the FF the LAeq over the total flight time and is the AH data only over the ATC time?
Response:
The column-names were altered to be clearer. The levels are integrated values either to the total flight time or to times with air traffic control ATC.

The signal to noise would be more appropriate as the LAeq during the ATC to the FF level at the same time for the sound level in the flight deck changes with the various phases of flight. Also for Signal to Noise it is more appropriate to use the same metric - not different ones.
Response:
This change was done and the signals from acoustical manikin are spectral corrected according to ISO 11904-2 during times with ATC. The SNR is now between 7 and 11 dB(A).

Headset, Fig 3 discussion. The data in Fig 3 does show a worse hearing for left ear use BUT Figure 4 shows that the predominant users of the left ear is a very much smaller sample AND highly skewed >40 years (2 vs 43) – all of the >40 years group has poorer hearing. There is some discussion of this later but it should be mentioned here also
Response:
Fig. 3 is now Fig. 4 and vice versa. This small group (~20%) is responsible for the difference between left and right ear in the pilots >=40 years (compare Fig. 1b with Fig. 4).

“The free-field sound measurements in Tab. 2 (Hoffmann 2004) in aircraft cockpits show“ I’m not clear if these authors measured the free field noise levels or if they use the data from Hoffman – which is not in the reference listing
Response:
The free-field sound measurement data from Hoffmann in Tab. 2 …
Hoffmann is listed in the Acknowledgements: "Also many thanks to Knut Hoffmann of Lufthansa Technik in Hamburg for the measurement data in jet cockpits". A new reference Hoffmann is inserted in the list as a personal communication.

The authors do not discuss that the headset provides some noise reduction as it is fixed over the ear.

This means that the in ear noise level with no communications will be less than the free field noise level
Response:
The pilots in the measurement setup used no circum-aural but supra-aural headphones with only marginal noise attenuation. Otherwise the pilots often use to put off the headset in times without communication.

The comments about the use of noise reduction “eliminate the risk” is not justified by the findings in this paper and should be changed to say ‘may’
Response:
The use of headsets with active or passive noise reduction at both ears can solve this last problem and may eliminate any risk for hearing loss in pilots during their normal occupational activity. The hint to pilots to allways use both ears for communication and never use only the left ear may also be helpful.

Comments of Rev3: Christian Giguère/Responses (grey)

The authors also do not address the ethical issues or procedures that have been put in place to utilize data from annual employer heath check-ups (audiometric data) in the context of a research study. Was consent obtained?
Response:
Ethics statement:
The audiometric measurements were carried out as part of the annual health checkups and personal questions answered pilots voluntarily with consent to publish the data anonymously.

Item “Objective” simply lists one fact and one occupational requirement; this does not constitute a research objective.
Response:
The study examines the hearing thresholds of pilots with respect of ambient noise and communication sound.

Item “Methods” does not mention that noise exposure assessments were also carried out.
Response:
The hearing of 487 German pilots was analyzed by audiometry in the frequency range 125 Hz – 16 kHz in age-groups. Cockpit noise (free-field) and communication sound (acoustic manikin) measurements were edited.

Page 3 - The introduction is incomplete and too general.
Response:
The introduction was extended and relates to to literature.

Page 4, lines 3-6: I suggest renaming this section as “Methods” and renaming the first subsection “Study Population” instead of “collective.
Response:
Done!

Page 4, line 8: Sentence is unclear. Do you mean “… were administered a standardized interview about …”
Response:
All pilots were interviewed in a standardized manner about their professional and leisure -related noise exposures.

Page 4, line 12-13: “their questionnaires were imprecise”. Please elaborate.
Response:
… because their questionnaires were lost or incomplete.

Page 4, line 20: “the pilots were divided in two age groups”. This is incomplete. Table 1 shows data distributed over four age groups.
Response:
Four age groups with ten year range were pooled for statistical characteristics (percentiles).

Page 4, lines 28-38: It will also be important to specify the environment where these measurements took place. In a sound proof room?
Response:
Pure tone audiometry was performed by experienced audiologist’s assistants in a sound proof room of the medical center of the airline company.

Page 4, line 44-45: Specify which dummy head was used. It is usually called “acoustic manikin” (see ISO 11904-2 terminology).
Response:
…and an acoustical manikin Type 4100 with an artificial middle ear Type 4157 of Brüel & Kjær (Denmark).

Page 4, line 45-46: “as it meets the requirement”. Which requirement? Also specify which sound level meter was used. What it integrating?
Response:
In all sound measurements the integrating function and an A-filter were used, as it corresponds to the regulations in the EU DIRECTIVE 2003/10/EC.

Page 4, lines 48-49: Unclear. Were two headsets used, one for the pilot and one for the dummy, or only one headset alternating between the pilot and the dummy? Please give more details on the protocol and indicate what precautions were taken to ensure the measurements on the dummy were representative of the pilot exposure. How long in duration were the acoustic measurements?
Response:
The setup is described in the Methods/Acoustic Measurements chapter:
The free-field microphone was placed besides the pilot near the ear. The acoustic manikin was placed on a seat just behind the pilot wearing a headset in the same way as the pilot receiving the same signal.

Page 4, lines 49-50: The sentence starting with “By using …” is unclear and may not be necessary.
Response:
Deleted!

Page 5, lines 6-7: change “witch” for “which”, an “arbitrarily” for “arbitrary”.
Response:
...of which version is used and become arbitrarily.

Page 5, lines 10-11: Sentence is unclear. Is it that the focus of the paper was placed on individual left-right differences because they do not require age-correction?
Response:
The sentence: "For further analysis differences between both ears were used with the advantage to eliminate the aging effects on hearing thresholds“ is replaced with:
The focus of the paper was placed on individual left-right threshold differences because they do not

require age-correction.

Page 5, lines 18-19: “… shown in its essential results as a table”. Unclear
Response:
Deleted!

Page 6, line 7: It says “civilian aviation companies”, but on page 4 it says “a large German airline. Where pilots coming from several airline companies or just one?
Response:
…of jet pilots from a German airline company were presented…

Page 6, lines 8-11: These lines describing the octave layout should be transferred to the Fig. 1 caption. They may also be omitted; Fig.1 seems clear enough.
Response:
Deleted!

Page 6, line 15: I suggest deleting “as expected” in a results section. Interpretation is best left for the discussion.
Response:
Deleted!

Page 6, lines 17-18: For these individual thresholds, it may be more accurate to assign the largest possible hearing level the audiometer can produce at 16 kHz (or perhaps 5 dB above), instead of considering missing data. It would produce less bias due to equipment limitations.
Response:
The expenditure is too big for the minimum advantage of this method at a frequency which is not relevant for calculations!

Page 6, line 20: change “cero” for “zero”
Response:
Done!

Page 6, line 21: change “do not” for “does not”.
Response:
Done!

Page 8, line 9: Presumably the Jilek data was only used to correct high frequencies. Which frequencies?
Response:
The age-correction by Jilek is deleted! Also the corresponding reference. Fig. 2a was limited to frequencies up to 8 kHz as the ISO 7029 2nd edition provides.

Page 8, line 12: change “a less” for “a smaller”
Response:
Done!

Page 9, lines 6-7: “artificial head” is used here while “dummy” is used in page 4. Please use same terminology throughout. Nowadays, “acoustic manikin” is the commonly used terminology (e.g. ISO 11904-2).
Response:
Done!

Page 9: The authors do not detail the communication headsets used. It is important to do so. Are they noise attenuating or not? Are they one-sided or two-sided? Are they circumaural, supra-aural or in-ear devices? Are they the same or different across airplane types?
Response:
The headset was a two-sided supra-aural headphone without active noise attenuation. The passive noise attenuation was marginal.

Page 9, lines 9-11: This statement seems incomplete. There is ample data in the literature showing that speech understanding is possible at negative signal to noise ratios. I think what the authors need to say here is that under typical conditions of use in the workplace or airline industry, the speech signals from headsets are significantly above the noise, thereby contributing or becoming the dominant source of sound exposure. A reference would be useful here.
Response:
The sound pressure levels for communication are higher than the ambient noise for a clear understanding of the messages. These sound pressure levels were measured with an acoustic manikin under the headset …
We didn't find a reference especially for speech understanding in noise under headsets. But we agree with you, that it would be useful.

Page 9, lines 14-17: "contains impulsive parts of noise". Unclear statement since the noise is continuous. Are we referring to the speech fluctuations or to noise transients associated with the headsets? Justify the use of an "impulse" time constant to measure speech levels. It is unusual. Is there any reference or standard that can be cited?
Response:
… contains impulsive parts of sound. Sorry it is clear that speech is no noise in this context. The effect on hearing is the same if noise or sound. The fast pressure levels are now added in Table 2 and you can see, that the difference between the LAIeq and LAeq is between 5 and 6 dB, what means the speech signal is impulsive and not continuous like the ambient noise.

Page 9, lines 19-41: The reporting of acoustic measurements is unclear. Table 2 presents sound levels using "fast" and "impulse" response settings. Are these values single measurements in time or are they the maximum or average of many such measurements over a substantial period of time on each flight? Why not use an integrating sound level meter? Provide details on the averaging over pilots in each type of airplane?
Response:
All measurement data are integrated levels during the total flight time. The part of communication is calculated on the basis of the ATC (air traffic control) time.
The sound measurements are independent of the examined pilots and no individual noise dose can be calculated for the pilots.

Page 9, lines 43-46: How was this correction factor derived precisely? Why not directly measure headset sound levels in dB(A)f in the first place? It is also not clear if the correction was applied to the number reported in column AH in Table 2. The unit is "dB(A)i", implying these are the actual measurement values and not the corrected values. Please clarify.
Response:
Now in Table 2 the column with fast values is placed before the impulse values. Three aircrafts had no values for "fast" measurement at acoustical manikin and must be excluded. The correction factor for impulsive sound is the difference between the "impulse" and "fast" levels and can now be directly calculated in Table 2.

Page 9, lines 48-51: There are several problems with the calculation of the SNR. As shown in Table 2, the SNR seems to express the difference between a signal level measured with an "impulse" time

response setting and a noise level measured in “fast”. It also assumes that the headsets do not provide any noise attenuation. Please specify. Finally, AH speech measurements must first be corrected to represent sound field related levels. The procedure is described in ISO 11904-2 and requires spectral analysis of the signal recorded with the manikin/artificial head. Without this sound field correction, the SNR can be inflated by several dBs.
Response:
The SNR is now calculated on the basis of the “fast” levels through the ATC time and corrected by using the ISO 11904-2. The SNR values are now between 7 dB and 11 dB not 15.7 dB as Hoffmann calculated.
Lower and Bagshaw (1996) reports ambient noise levels between 72 and 79 dB(A)f and corrected measurements with acoustical manikin (Kemar) between 80 and 88 dB(A)f. The SNRs can be calculated as the difference between this both levels and are between 6 and 13 dB with supra-aural headphones. The span of their SNR data is 7 dB and in our data 4 dB. The average is the same: 10 dB.

Page 9, lines 56-60: The limit of 85 dB(A) applies to sound field measurements, not manikin/artificial head measurements. The signal level measured in the manikin must first be corrected according to ISO 11904-2 to transform to sound field related levels. Without this correction, the communication sound level can be overestimated by several dBs.
Response:
The corrected sound pressure levels of communication sound (ATC) exceeds in 6 of 9 cases the upper exposure action value of the directive of 85 dB(A).

Page 11, lines 7-8: This statement seems incomplete. While it is true that the “free” ear will receive a reduce exposure for those pilots listening to only one ear, those who listen with the two ears likely will set the headset volume to a lower value due to binaural effects, leading to a reduced exposure in the two “live” ears compared to the “live” ear of pilots listening to only one ear. The authors mention this aspect themselves on page 13 (lines 36-40). I think it is better to delete “resulting in …. other ear” here and cover all sides of the argument in the discussion.
Response:
The statement is here deleted as it has to be placed into the discussion.

Page 11, line 21: It is suggested to present Fig. 4 before Fig. 3 to improve the logical flow of the paper, i.e. first present the distribution of headset ear preference among pilots, then the audiometric results for each group.
Response:
Done!

Page 12, line 15: change “round the cero line in” to “around the zero line within”
Response:
… the difference values oscillate around the zero line within a ± 1 dB range.

Page 12, lines 30-31: Improve grammar
Response:
… the percentiles of our data are lower on an average of 4.5 dB and the 80 % span in the dataset is smaller on an average of 9 dB

Page 12, line 33: The Hoffmann (2004) reference is not listed on pages 15-16. Also, does is mean that the data dates back from 12 years ago? It is not in itself a problem, provided it is clearly stated in Methods. It is important to do so because the authors refer to “modern headsets” at the end of the paper, so a time frame for the measurements is needed.
Response:

Hoffmann is listed in the Acknowledgements and as reference 10. The data are unpublished. We got them as a personal communication to use them in our study with pilots.

Page 12, line 38: It should be “none”
Response:
Done!

Page 12, lines 51-55: The authors extracted SNR-50 and SNR-80 values from Fig 2 in Killion et al (1983). This figure is only an illustration of their scoring method, not the results from an actual experiment. The Fig 1 of their paper reports actual data from several studies. The SNR-50 is about 2 dB for normal-hearing individuals, not 7 dB. In any case, the noise used in this (and other) speech test is very different in character then than the noise in an airline cockpit. It may be more relevant to cite studies of communication headset exposure in the airline industry or other occupations. Typically the SNR is about 13 dB, which is near the 16 dB found in this study, especially since there is no indication that the data in the current study was corrected according to established methods (ISO 11904-2 or similar), which would lower the SNR.
Response:
Now the SNR has a range from 7 dB to 11 dB using the “fast” levels with a mean of about 10 dB. We deleted the SNR-80 and set the SNR-50 to 2 dB as a result in Killion et al. (2004).

Page 12, lines 58 and rest of paragraph: While it is interesting and relevant to point to ANR systems, why not also point to headsets with passive attenuation properties? Again, the authors have not documented which types of headsets were used by the pilots in the study. Without this information, any discussion on alternative headset solutions becomes rather vague.
Response:
In chapter Methods/Acoustic Measurements the following sentence is inserted:
The headset was a two-sided supra-aural headphone without active noise attenuation. At this open types of headsets the passive attenuation is only marginal.

Page 13, lines 46-47: The authors state “Modern headsets will solve this problem”. Given that the measurements appear to have been made around 2004, are the authors inferring that the headsets used in airlines are different now or are they implying that the commercial airline industry has not caught up with recent headset technology? Please clarify your thoughts.
Response:
Circum-aural headsets with passive sound attenuation can be helpful to reduce the communication sound levels, but they impede the communication between the crew as the attenuation at high frequencies is much better than at low frequencies in that earphones. Headsets with active noise reduction (ANR) systems are now commonly installed.

Fig 1 caption: Describe both parts in the same manner. “Part a shows ….. Part b shows”.
Response:
Part a shows hearing thresholds of civilian airline pilots in two age groups at both ears averaged from 125 Hz up to 16 kHz. Part b shows …

Table 1: Why are the numbers rounded to the nearest 0.5 dB for the first two age groups and rounded to the nearest 0.1 dB for the two other age groups.
Response:
The data are not rounded to 0.5 dB but averaged of both ears. So we got 0.5 values. For medians we need odd numbers to get those exact values otherwise the Median is the average of two numbers. For quantils the procedure uses linear interpolation between values and so we get dezimal numbers rounded to 0.1.
Young pilots do not differ much in their threshold values and their quantiles seem to be rounded to 1

or 0.5.

Table 2: see above
Response:
In Table 2 all dB values are rounded to 0.1 dB. Minutes are rounded to 1.

VERSION 2 – REVIEW

REVIEWER	Arve Lie STAMI, Norway
REVIEW RETURNED	23-Sep-2016

GENERAL COMMENTS	No 13: Checklist? I could not find any

REVIEWER	Marion Burgess School Engineering and Information Technology UNSW Australia Canberra
REVIEW RETURNED	29-Sep-2016

GENERAL COMMENTS	BMJ review #2 The changes in this paper have made a considerable improvement. There are still some aspects that need attention – and many grammatical corrections. The entire paper needs a careful review, preferably by a native English speaker, to attend to the many grammatical/expression errors in the sentences. The following are just some I have identified on page 2 alone – but the entire paper needs careful check. P2 L12 the sentence “Cockpit noise (free-field) and communication sound (acoustic manikin) measurements were edited. “ In this context the word ‘edited’ is incorrect and I think that you mean something like “were obtained” P2 L13/14 this is a statement implying that all cockpit noise and all under headset noise is within the values quoted – and im sure that is not correct! Suggest that you ammend to say something like “in cockpits were found to be between...” and similarly “headset were found to be…” P2 L22 replace “only at the right have” with “only on the right ear”

	P2 L 25 should be “higher than” not “higher as” P2 L29 “. allow a more relaxed working for pilots” should be something like “. allow for a more relaxed working environment for pilots” P2 L29 active noise reduction systems don’t of themselves reduce the sound levels of communication – I think you are trying to say that they reduce the noise from outside the communication headset and allow for a reduced level for the communication signal Abstract – Results – the levels quoted here should be rounded to the nearest dB as they are in the main part of the manuscript Introduction – the expanded new paragraph in the introduction makes reference to “raised flight amount” assuming this is meant to be the increased number of flights it is important to separate the over flight noise issue from the internal aircraft noise aspect. It is important to make it clear that [3]and [4] do not specifically relate to airline pilots. Table 1 – there is no reference to this table in the text AND there is an error as the same frequency is listed for the three main rows. Page 8/9 the sentences referencing the EU Action levels are incorrect. The action and upper action levels are in terms of $L_{Aeq,8hr}$ – while communication sound level may be 83dBA or 88 dBA it is unlikely that there would be continuous communication over an 8 hour period !! P11 the instroduction of an ‘impulse correction factor’ is not appropriate in this discussion - it is not identified where the ‘ between 5 and 6 dB” has been determined but I suspect it is from environmental assessment where the factor is intended to indicate the additional annoyance and not workplace noise exposure and hearing loss. I am not have not seen a workplace noise assessment

	that requires the adoption of such an impulse correction factor! P12 the new sentences including SNR-50 and SNR-100 do not make sense

VERSION 2 – AUTHOR RESPONSE

Comments of Reviewer1: Arve Lie	**Responses**
No 13: Checklist? I could not find any	Checklist at the end of document!

Comments of Reviewer2: Mario Burgess	**Responses (cyan)**
P2 L12 the sentence “Cockpit noise (free-field) and communication sound (acoustic manikin) measurements were edited.“ In this context the word ‘edited’ is incorrect and I think that you mean something like “were obtained”	Cockpit noise (free-field) data and communication sound (acoustic manikin) measurements were evaluated.
P2 L13/14 this is a statement implying that all cockpit noise and all under headset noise is within the values quoted – and im sure that is not correct! Suggest that you ammend to say something like “in cockpits were found to be between...” and similarly “headset were found to be …”	The ambient noise levels in cockpits were found to be between 74 dB(A) and 80 dB(A) and the sound pressure levels under the headset were found to be between 84 dB(A) and 88 dB(A).
P2 L22 replace “only at the right have” with “only on the right ear”	Done!
P2 L 25 should be “higher than” not “higher as”	Done!
P2 L29 “. allow a more relaxed working for pilots” should be something like “. allow for a more relaxed working environment for pilots”	… and allow for a more relaxed working environment for pilots.
P2 L29 active noise reduction systems don’t of themselves reduce the sound levels of communication – I think you are trying to say that they reduce the noise from outside the communication headset and allow for a reduced level for the communication signal.	Active noise reduction systems allow for a reduced sound level for the communication signal below the upper exposure action value …

Abstract – Results – the levels quoted here should be rounded to the nearest dB as they are in the main part of the manuscript.	Done!
Introduction – the expanded new paragraph in the introduction makes reference to "raised flight amount" assuming this is meant to be the increased number of flights it is important to separate the over flight noise issue from the internal aircraft noise aspect.	This results in reduced noise exposure in the flight cabin and less annoyance for the affected population. However, the reduced annoyance per flight will be overcompensated by a higher flight frequency.
It is important to make it clear that [3]and [4] do not specifically relate to airline pilots.	… which are independent of the occupation [3, 4].
Table 1 – there is no reference to this table in the text AND there is an error as the same frequency is listed for the three main rows.	The reference is just below the table. The main rows are corrected (3 4 6 kHz).
Page 8/9 the sentences referencing the EU Action levels are incorrect. The action and upper action levels are in terms of LAeq,8hr – while communication sound level may be 83dBA or 88 dBA it is unlikely that there would be continuous communication over an 8 hour period !!	The free field measured ambient noise in Airline cockpits does not reach the lower exposure action values of 80 dB(A) of the EU DIRECTIVE 2003/10/EC [7] if the flight time is below 8 hours. The corrected sound pressure levels of communication sound $AMc_{(ATC)}$ exceeds the upper exposure action value of the directive of 85 dB(A) in 6 cases for a flight times of 8 hours and more. The minimum communication sound level was calculated to 83.5 dB(A) in the Airbus A320-200, and the maximum level to 88.1 dB(A) in the Airbus A310-300. Only in intercontinental flights the flight time reaches or exceeds 8 hours.
P11 the introduction of an 'impulse correction factor' is not appropriate in this discussion - it is not identified where the ' between 5 and 6 dB" has been determined but I suspect it is from environmental assessment where the factor is intended to indicate the additional annoyance and not workplace noise exposure and hearing loss. I am not have not seen a workplace noise assessment that requires the adoption of such an impulse correction factor!	In Germany such a "impulse correction factor" was used for many years to consider the increased danger of impulses compared to continuous noise. If the difference between the "impulse" and "fast" levels is about 5 dB and more, I think this indicates clearly the impulsive character of the communication signal and should be addressed in this context.
P12 the new sentences including SNR-50 and SNR-100 do not make sense	Killion et. al (2004) used this notation, but we can avoid this usage.

VERSION 3 – REVIEW

REVIEWER	Marion Burgess Acoustics and Vibration Unit School Engineering and Information Technology UNSW CANBERRA
REVIEW RETURNED	05-Jan-2017

GENERAL COMMENTS	Page 12 line 48 ACT should be ATC Page 12 I am still not confident that the reference to the "impulse correction factor" in relation to workplace noise exposure is appropriate in this paper - it is not the way impulse noise is dealt with in the EU Directive which only uses Peak - if it is an applicable method in Germany then the text should refer to this as a German approach and should include the relevant reference to the German legislation.

VERSION 3 – AUTHOR RESPONSE

Comments of Reviewer2: Mario Burgess	**Responses (cyan)**
Page 12 line 48 ACT should be ATC	Done at page 11 line 48

Page 12 I am still not confident that the reference to the "impulse correction factor" in relation to workplace noise exposure is appropriate in this paper - it is not the way impulse noise is dealt with in the EU Directive which only uses Peak - if it is an applicable method in Germany then the text should refer to this as a German approach and should include the relevant reference to the German legislation.	At page 8 lines 47-48 in chapter **results** the term "impulse-correction-factor“ is deleted and the text changed from: "…and can be used as a correction factor for impulsive noise and its special effects on hearing (not listed in Tab. 2)" to: "… indicating an impulsive character of the communication sound" At page 11 (not 12) lines 46-51 the changed text is: … for noise exposure levels by communication the “impulse” weighted exposure levels could be used. In all cases the upper exposure action values then would be reached during communication. As the ATC time is mostly shorter than half of the total flight time and never 8 hours, the higher exposure levels will be compensated approximately by the shorter exposure time. We want to find a reason why the left ear shows hearing loss although the upper exposure action values are mostly not reached! The question is: do we need further restictions for the action values or is it appropriate to consider the impulse character of the noise at the left ear?