Automated Threshold Determination of Auditory Evoked Brainstem Responses by Cross-correlation Analysis with Varying Sweep Number

Haoyu Wang[1-4], Bei Li[1], Xu Ding[2,3], Xueling Wang[1-3], Zhiwu Huang[1-3], Yunfeng Hua[2-4], Lei Song[1-3], Hao Wu[1-4]

[1]Department of Otolaryngology-Head and Neck Surgery, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[2]Ear Institute, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[3]Shanghai Key Laboratory of Translational Medicine on Ear and Nose Diseases, Shanghai, China

[4]Shanghai Institute of Precision Medicine, Shanghai Ninth People's Hospital, Shanghai, China

Haoyu Wang and Bei Li contributed equally to this work.

Correspondence to: Yunfeng Hua, Lei Song and Hao Wu, Shanghai Ninth People's Hospital, 115 Jinzun Road, 200125 Shanghai, China.

Email: yunfeng.hua@shsmu.edu.cn, lei.song@yale.edu, wuhao@shsmu.edu.cn

ABSTRACT

Auditory brainstem response (ABR) is widely employed to evaluate the hearing function, both in clinics and basic research. Despite many attempts for automation over decades, reliable determination of threshold stimulus level still relies on human visual identification of waveform, which oftentimes is subjective. Here, we report a robust procedure for automatic and accurate threshold determination in both mouse and human ABR. Contrary to prior approaches, in our new threshold determination algorithm, the on-going averaging is stopped once the waveform is confirmed by a cross-correlation time shift approach. The flexible ending sweep numbers for different stimuli is used to inform the threshold determination. We found a good match of the threshold readings between the algorithm and the human judges. Moreover, in the algorithm, smaller sweep number is required for strong response from supra-threshold level, and thus a considerable portion of sweeps can be saved in comparison to the case with level averaging of a fix number. These features are attractive and implementation of this method in commercial devices will make the ABR test procedure more objective and efficient.

Keywords: auditory brainstem response, threshold determination, cross-correlation, automation

INTRODUCTION

The auditory brainstem responses (ABRs) are brain electrical potential changes due to synchronous neuronal activities evoked by supra-threshold acoustic stimuli (Jewett et al., 1970). These responses are detectable using non-invasive surface electrodes placed on the scalp of the test subject, and thereby widely employed to assess the hearing function. In rodents and cats, typical ABR waveform is composed of initial five peaks in the early onset of sound evoked potentials, followed by broader, later waves that represent synchronous activities arising from projections along the auditory ascending pathway including auditory nerve, cochlear nucleus, superior olivary complex, lateral lemniscus and inferior colliculus, respectively (Henry, 1979; Melcher et al., 1996), whereas in human slightly different peak generators were demonstrated with intracranial recordings (Moller and Jannetta, 1983) and neuromagnetic responses (Parkkonen et al., 2009). Thus, features like ABR wave latencies and amplitudes provide clinical-significant information, for instance site of lesions or tumors in the auditory system (Lewis et al., 2015; Roeser et al., 2007) based on how the properties of waveforms are altered.

Although the ABR is an objective measurement, at near-threshold the waveform recognition involves human interpretation. Currently, professionals are still required to supervise recording and visually identify the obtained waveforms which is labor-intensive. Besides, such interpretations oftentimes are subjective and can introduce errors that vary from person to person. When bias due to the skill and the experience of the interpretators is involved, the variation is not trivial, especially for cases with untypical waveform or high background noise (Vidler and Parkert, 2004). As precise and objective measurement of small hearing threshold elevation became critical for diagnosis of progressive hearing loss (Barreira-Nielsen et al., 2016), hidden hearing

loss (Kujawa and Liberman, 2009; Mehraei et al., 2016; Ridley et al., 2018; Sergeyenko et al., 2013), age-related hearing loss (Gates and Mills, 2005; Sergeyenko et al., 2013) and tinnitus (Bramhall et al., 2018; Castaneda et al., 2019), automated approaches with high precision and reliability are in demand to objectify the ABR threshold determination. Over decades, many attempts were made including: (1) quantification of the waveform similarity by comparison to existing templates (Davey et al., 2007; Elberling, 1979; Valderrama et al., 2014) as well as based on features learned by artificial neural network (Alpsan and Ozdamar, 1991; McKearney and MacKinnon, 2019) from human annotated datasets; (2) quantification of the waveform stability by cross-correlation function between single-sweeps (Bershad and Rockmore, 1974; Weber and Fletcher, 1980), interleaved responses (Berninger et al., 2014; Xu et al., 1995) or responses at adjacent stimulus levels (Suthakar and Liberman, 2019); (3) the 'signal quality' through scoring procedures like F-ratios (Cebulla et al., 2000; Don and Elberling, 1994; Elberling and Don, 1984; Sininger, 1993). Due to inconsistencies in waveform and signal-to-noise-ratio (SNR) introduced by differences in test subject conditions, electrode placement and impedance, as well as acquisition settings, the accurate threshold determination is only possible under a narrow range of experimental settings, hampering direct comparisons of ABR data and results across laboratories.

In this study, we proposed a novel approach which detects time-locked ABR waveforms via a time shift cross-correlation approach during on-going sweep averaging. Sweep averaging is terminated upon reaching a criterion for a detectable waveform at different stimulus levels and the threshold estimation was also carried out by the algorithm. The collected results were validated by human experts on the same mouse or human subjects. To verify, the total numbers of sweeps (as an

indicator of test duration) were compared between cases in the algorithm and with a fixed sweep number for level averaging and prove the algorithm effective.

MATERIALS AND METHODS

Animals, Human Participants and Ethics

C57BL/6 mice were purchased from Sino-British SIPPR/BK Lab Animal Ltd. (Shanghai, China). The telomerase-knock-out mice were kindly donated by Prof. Lin Liu (Nankai University, China) and bred in house. Human participants were recruited from Shanghai Ninth People's Hospital and consent forms were signed before the experiment. This study was conducted at the Ear Institute and the Hearing and Speech Center of the hospital. All procedures were reviewed and approved by the Institutional Authority for Laboratory Animal Care (HKDL2018503) and the Hospital Ethics Committee for Medical Research (SH9H-2019-T79-1).

ABR Recording

Mouse ABRs were recorded via a TDT RZ6/BioSigRZ system (Tuck-Davis Tech. Inc., US) in a sound-proof chamber as previously described (Lin et al., 2019). In brief, 7-week-old mice were anesthetized through intraperitoneal injection of Chloral hydrate (500 mg/kg). During the recording, animal body temperature was maintained at 37 °C using a regulated heating pad (Harvard Apparatus, US) with a rectal thermal probe placed under the animal's body. Evoked potentials were registered via subdermal needle electrodes (Rochester Electro-Med. Inc., US) placed at the animal's vertex (active electrode), left infra-auricular mastoid (reference electrode) and right shoulder region (ground electrode). 3-ms tone pips at 16 kHz were delivered via an MFI speaker (Tuck-Davis Tech. Inc., US) positioned in the front 10 cm from the animal's vertex. Acoustic stimuli were presented 20 stimuli per second and the evoked potentials were sampled at 24 kHz. Artifact rejection level was set at < 35 % (mean rejection voltage 20.5 μV). Sound level series were acquired starting from 90 to 0 dB

SPL (sound pressure level) with 5-dB step size. For one animal, the stimulus level series were repeated from +10 to –10 dB SPL around the estimated threshold with 1-dB step size (Fig. 3B).

Human ABRs were recorded by a commercial ABR device (Intelligent Hearing Systems, US) with Smart EP software from four volunteers aged 21-29 years without the knowledge of their medical conditions. Click sound stimulation (100 μs duration, rectangular envelopes) was generated and presented monaurally through ER3 insert earphones with foam tips. Stimuli were presented at a rate of 37.1/s with alternating polarity. Electrode impedance was < 5 kΩ and inter-electrode impedance was within ± 1 kΩ. The artifact rejection level was < 31% (rejection voltage 31 μV) to exclude contaminations from EEG and myogenic potentials. The evoked potentials were collected with 40 kHz sampling rate and × 100,000 amplification. The bandpass filter was set at 100 - 3000 Hz. Average responses over 500, 1000, and 2000 sweeps were acquired and repeated three times for the level series starting from 60 to 0 dB SPL with 5-dB step size.

Cross-correlation Analysis in Mouse ABR

Sweeps were randomly subdivided into two groups. Cross-correlation operations (MATLAB Central File Exchange Function *xcorr*, MathWorks, US) were applied to subgroup averages. The result of this operation yielded the correlation coefficient as a function of time shift between two signals (ABR subgroup averages). The time shift (signal lag) of the maximal coefficient was used to judge the reproducibility of ABRs. As the responses are time-locked to the acoustic stimuli, a neglectable time shift is expected. In this study, maximum allowed lag (L) for a true waveform was within one data point from time zero (equivalent to ± 0.042 ms, 1 % of the analyzed temporal

143 window) due to system sampling error. As noise peaks could also coincidently
144 overlapped within the desired time shift, three parallel cross-correlation runs with
145 regrouped sweeps were implemented and false positives were rejected upon
146 inconsistent lag values. In addition, the correlation coefficient peak amplitude was
147 included as an independent variable (Fig. 2C).

148 At each stimulus level, averaging with increasing sweep numbers was iterated
149 and the ending sweep number was noted upon a detectable response by the cross-
150 correlation approach described above. Each iteration consists of 50 sweeps and the
151 upper limit was set to include seven iterations (350 sweeps). The estimated threshold
152 was just above the stimulus level at which the upper iteration limit was reached. A
153 more precise threshold determination was done by modeling the change of the ending
154 sweep numbers upon level series. For that both sigmoidal (1) and exponential
155 functions (2) were employed to fit the relationship between the normalized iteration
156 count *C'* and the stimulus level *S* using a nonlinear least square method in MATLAB
157 (MathWorks, US). In the functions, $\alpha_1 = 0.6$ and $\alpha_2 = 0.25$ were fixed for calibrated
158 lag criterion ($L = 1$), while β_1 and β_2 were obtained by fitting. The estimated threshold
159 was the corresponding *S* with the sigmoidal function value of 0.9 or the exponential
160 function value of 1.0.

161 $$Sigmoid\ model: C'(S) = \frac{1}{1+e^{\alpha_1(S-\beta_1)}} \quad (1)$$

162 $$Exponential\ model: C'(S) = e^{-\alpha_2(S-\beta_2)} \quad (2)$$

163

164 ## Cross-correlation Analysis in Human ABR

165 For human ABR, average responses were recorded sequentially and used as inputs in
166 the algorithm with minor modification (Fig. S3). Instead of regrouping single sweeps
167 as in the mouse ABR, the three combinations of two out of three average responses

(*E*{A}, *E*{B}, *E*{C} in Fig. S3) were used for the parallel cross-correlation runs to reject false positives caused by noise peaks. When the lag condition is not fulfilled, averages over more sweeps (with a step size of 500) were used for further iterations until the upper limit of 3500 was reached. The average responses over 500, 1000 and 2000 sweeps were recorded, whereas responses over 1500, 2500, 3000 and 3500 sweeps could be obtained by weighted averaging (3) where $E\{m\}$, $E\{n\}$ and $E\{m + n\}$ denote the time averages over *m*, *n* and *m* + n sweeps, respectively.

$$E\{m+n\} = \frac{m \bullet E\{m\} + n \bullet E\{n\}}{m+n} \quad (3)$$

The maximum allowed lag (*L*) for a true response was within seven data points from time zero (equivalent to ± 0.175 ms, 2 % of the analyzed temporal window). The estimated threshold was the lowest level with a detectable waveform.

Visual identification of ABR Threshold by Human Judges

To estimate the ground-truth thresholds of the recorded mouse and human ABRs, five clinicians were asked to independently assess the average responses and report the visually identified thresholds. The identities of the test subjects were blinded to the judges. The average responses of all level series were provided, of which either constant number of sweeps (the conventional averaging) or ending number determined by the algorithm (the algorithm termination averaging) was used. The thresholds were determined by three out of five execution judges (with the highest and the lowest value excluded) and used to evaluate the accuracy of the algorithm outcomes.

RESULTS

Cross-Correlation Analysis in On-going Averaging

ABRs are embedded in high-level background activities and system noise. Smooth baseline and clear waveform, if present, are obtained usually after averaging over hundreds of sweeps. The required number of sweeps in averaging, however, not only depends on the amplitude of the evoked response but also varies between test subjects due to variations in, for instance, skull sizes, electrode impedances and placement that determined the distances from the generator, how far reach the electrode pick up the far-field signals and the angles of the vector projections. Within an ABR recording session, these experimental parameters are fixed and the SNR of recorded sweeps between stimulus levels can be quantitatively compared. It is expected that weak response evoked by low level stimulus requires more sweeps to average than those strong responses from high level stimuli to reach similar SNR level, whereas averaging fails to improve the SNR when a response is absent. Based on this fact, we designed a novel procedure to estimate the threshold stimulus level by monitoring the change in sweep number which is required for the average response to reach a stable SNR level.

In detail, recorded sweeps at a test stimulus level were randomly divided into two groups (Fig. 1, yellow boxes) and cross-correlation coefficients (CCs) were computed between two subgroup averages (Fig. 1, green boxes). Time-locked ABR waveform, irrespective of wave latencies and amplitudes, are detected by specifying a maximum allowed time shift (L; Fig. 1, magenta boxes) within which these subgroup averages are maximum overlapped (peak of the obtained CC). In addition, three parallel runs with regrouped sweeps are used (Fig. 1, red box) so that the false waves from randomly overlapped noise peaks of similar latencies can be rejected. Next, these

iteration steps repeat along with the increase of sweep number (Fig. 1, the inner loop), until either a true waveform is confirmed (consistent smaller lag than L) or the upper limit of iteration count (N) is reached. Here the iteration count limit is needed to avoid nonproductive attempts in the cases where waveforms are absent. Finally, the outer loop (Fig. 1) is implemented to scan the responses with decreasing stimulus levels. In this study, we start with 90 dB SPL in mice and 60 dB SPL in humans with a step size of 5 dB. The stop command was triggered upon a second attempt with exceeded iteration count, but the function was idled during initial optimization.

Threshold Determination in Mouse ABR

To test whether the algorithm could determine the threshold automatically, we recorded single-sweep ABR sets from eight mice (three wild-type adult C57BL/6 mice of normal hearing and five telomerase knock-out mice with age-related progressive hearing threshold elevation). The raw sweeps were corrected through a smoothing spline fit to remove baseline fluctuations (Fig. S1) before being processed by the algorithm.

As illustrated in Fig. 2A, an averaged mouse ABR waveform level series is plotted with the threshold of 30 dB SPL determined by human judges. Subgroup averages (Fig. 2B) were produced in the algorithm to compute the CCs. The obtained CC peak amplitudes (Fig. 2C) and corresponding signal lags (Fig. 2D) were plotted versus stimulus levels. With reducing stimulus level, the CC peak amplitude decreases monotonically, whereas the lags at supra-threshold levels are constantly within one data point (equivalent to a time shift of ± 0.042 ms from time zero). This result suggests that the cross-correlation time shift is more sensitive to the responses at near-threshold levels than the CC, thus justified its use in our algorithm.

We then determine whether the algorithm could use the ending sweep number to inform the threshold level. To enable rapid threshold determination by reducing the total number of test iterations, batches of 50 sweeps were added into the subgroup averaging by iterations until stable waveforms are reached. As shown in Fig. 2E, the iteration count (proportional to the ending sweep number) remained low at high stimulus levels but increases dramatically at near-threshold levels and reached its maximum at sub-threshold levels. Besides, we found consistent results were obtained within a large range of the allowed lag (Fig. S2A) and the maximum ending sweep number (Fig. S2B), suggesting that the new method works without fine-tuning the detection parameter.

Further attempts were made to model the iteration count for precise threshold determination between the lowest supra-threshold and the highest sub-threshold level. A sigmoidal function was employed to fit the normalized iteration counts (Fig. 3A). In order to determine the corresponding function value at the true threshold level, from one animal we acquired an ABR set with peri-threshold level series of 1-dB step size (Fig. 3B). Both exponential and sigmoidal functions were used to model the change. Note that for the exponential fit only data points at supra-threshold levels were used due to early cut-off by the maximum iteration count. The lowest supra-threshold level was found when the best-fitted exponential growth reached 1.0 or ~ 0.9 in the sigmoid growth (Fig. 3B). As fitting with sigmoid function does not require additional data exclusion, it was used to obtain the mouse threshold results for further validation of the method.

Threshold Determination in Human ABR

To test whether the new method was compatible with human ABR, we acquired ABR

sets from four human participants. Because export of single-sweep data was not an option on the commercial device we used, alternatively we use average responses over different pre-set sweep numbers (see Fig. S3 for the variant of the algorithm block diagram; see MATERIALS AND METHODS for more details). Averages (Fig. 4A) as well as subgroup averages (Fig. 4B) are shown at decreasing stimulus levels (60 dB to 0 dB SPL).

The CC peak amplitudes and the corresponding signal lags were plotted versus the stimulus levels (Fig. 4C and 4D). The sweep number increment was 500 per iteration and the upper limit was 3500 sweeps (at iteration count of seven). Note that we allowed slightly larger lag value (seven data points, equivalent to a time shift of ± 0.175 ms from time zero) for a true waveform because a broader waveform is expected for human ABR evoked by click-sound than that of mouse ABR. As shown in Fig. 4E, the iteration count increases fast to reach its maximum near the visually identified threshold level.

Comparison between Expert and Algorithm Determined Threshold

To validate the new method, we asked five human experts to independently assess the same ABR sets and compared the visually identified thresholds to those determined by the algorithm (Table 1). Scatterplots of the algorithm determined thresholds versus the visually identified thresholds showed no significant difference for both mouse and human ABR (Fig. 5A and Fig. 5C), suggesting a reliable threshold determination by our algorithm. Besides, matched thresholds were also obtained when average responses were generated with the ending sweep numbers from the algorithm (Fig. S4A and S4B). This result suggests that extensive averaging at supra-threshold levels does not improve the detection accuracy, and thus including a large number of the

sweep recording is unnecessary. We then compared the total sweep number used in two averaging methods, one with a constant sweep number (350 for mouse and 3500 for human ABR) at all stimulus levels and the other with varying sweep number based on the response detection in the algorithm. We found in algorithm a reduction of 66.76 ± 4.09 % and 53.08 ± 12.91 % (mean ± s.d.) of number of sweeps needed in mouse and human ABR, respectively (Fig. 5B and Fig. 5D).

DISCUSSION

Over decades several statistical approaches have been proposed to automate and objectify the ABR analysis. The cross-correlation approach has two advantages. First, high intra-subject waveform stability in both ABR wave latencies and amplitudes leads to the robust waveform which can be detected by cross-correlation analysis with high sensitivity. Second, it is template-free and not subject to the influence of large inter-subject waveform variability which is often presented in template-based approaches. However, prior attempts of the cross-correlation approach detect the ABR waveform with a decision boundary for a true response, for instance, minimum CC (Berninger et al., 2014; Bershad and Rockmore, 1974; Suthakar and Liberman, 2019; Weber and Fletcher, 1980) or maximum latency shift (Galbraith and Brown, 1990; Xu et al., 1995). Even with the same experimental settings, uniform SNR across recordings is not always guaranteed due to variabilities in skull size, electrode impedance and placement, as well as different sweep number used for averaging by individual experimenter. Thus, it is unlikely that calibrated decision boundary can be simply applied to another dataset without introducing detection error (in our hands up to 20 dB SPL, data not shown), limiting the application in cases like cross-institution collaboration efforts where data poolings are needed.

In contrast, our approach determines the threshold based on the relative change in the contribution of sweeps to the SNR of average response at different stimulus levels. The cross-correlation time shift approach has proven to be a reliable tool to detect the time-locked ABR waveform with high sensitivity, but in principle, it can be replaced in the algorithm by other quantifications like CC peak amplitude or F_{sp} (data not shown). The origin of the observed fast increase in the iteration count at near-threshold levels (Fig. 2E and Fig. 3E) is the compensation of a gradually reduced

response amplitude by noise suppression through averaging. As this increase reflects relative SNR change in the responses evoked by different stimulus strengths within subject, it is therefore rather insensitive to the inter-subject system variability. Besides, the new approach was proven not heavily rely on fine-tuning of the detection parameters. First, large difference in the obtained signal lags between sub- and supra-threshold levels (Fig. 2D) allows the lag selection from a large range without affecting the waveform detection (Fig. S2A). Second, the elevation of the maximum iteration count (proportional to the sweep number increase) causes little shift to the estimated threshold level when above certain value (Fig. S2B) as a result of its fast increase with reducing stimulus levels (Fig. 3B),

Next, we showed that precise threshold determination beyond the step size of level sampling was possible by modeling the sweep number change (Fig. 3B), in our case up to 1 dB in mouse ABR. Similar attempt was also made with human ABR sets, but a reliable model could not be established due to poor model fitting. Further development of this approach is to combine with level sampling strategy like progressively reduced step size (Cebulla and Sturzebecher, 2015) and increased sweep number per iteration at near-threshold levels, so that more effective data points can be used for model fitting.

In both mouse and human ABR, the new method was proven reliable in threshold determination with a maximum discrepancy of ± 5 dB to those provided by human experts. The average responses over varying sweep numbers seemed not to introduce additional difficulty in the threshold determination (Fig. 5B and Fig. 5E). That is because at the near-threshold levels, where the SNR of responses are critical for the threshold determination, an increasing number of sweeps were averaged, whereas at the supra-threshold levels the requirement of level averaging is low. Such feature is

attractive in two respects. First, it provides minimal quality control for unambiguous waveform recognition for both humans and algorithms. Such standardized data will benefit machine-learning-based approaches by minimizing annotation discrepancy in the training data (McKearney and MacKinnon, 2019). Second, when to stop averaging is an important decision during ABR recording (Don and Elberling, 1996; Madsen et al., 2018), the new method makes the ABR test more efficient by avoiding prolonged acquisition and redundant recordings.

ACKNOWLEDGMENT

We would like to thank Y. Li, K. Han, Y. Ren, L. Yang and H. Li from Shanghai Ninth People's Hospital for help with ABR assessment; Drs. G. Chen and L. Liu for providing terc$^{-/-}$ mice. This study was supported by the National Science Foundation for Young Scientists of China (81800901 to Y.H. and 81700903 to B.L.).

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of Interest The authors declare that they have no competing interests.

FIGURE LEGENDS

FIG 1. Flowchart of our algorithm for automatic threshold determination. The stimulus level starts with 90 dB SPL and the initial iteration count is one. The input is 50 sweeps recorded from a test subject. They are subdivided into two groups with corresponding subgroup averages $E\{A\}$ and $E\{B\}$ in each of three parallel runs (yellow boxes). Cross-correlation operation (*xcorr*) is applied to the subgroup averages (green boxes) and the obtained signal lags of the CC peaks are compared with the allowed time shift (L = 1 data point for mouse ABR; C1 to C3, magenta boxes). In cases when C1 to C3 are all true, the procedure starts over with lower stimulus level (red line, the outer loop), otherwise it iterates at the same stimulus level with more sweeps added (the inner loop, cyan line; 50 sweeps per iteration) until the maximum iteration count (N = 7 for mouse ABR) is reached, indicating a sub-threshold level.

FIG 2. Threshold determination by cross-correlation analysis in mouse ABR. **A** Example average responses over 350 sweeps were recorded from a mouse. The visually identified threshold level (bolded) was about 30 dB SPL. **B** Two subgroup averages were used in the algorithm for cross-correlation analysis. **C** Peak amplitude of the computed CCs was plotted as a function of the level series. **D** Signal lag of the CC peak vs. level function was plotted. At supra-threshold levels (dots) the mean signal lags from three parallel runs were close to time zero (0.28 ± 0.46 data points, mean ± s.d.), whereas significant large variability (30.50 ± 22.51 data points, mean ± s.d.) was constantly observed at sub-threshold levels (cycles). **E** Plot of the count of executed iterations vs. level function. Responses that were evoked by supra-threshold stimuli (black dots) require different number of iterations to converge lags within a

388 desired time shift (L ≤ 1 data point). After two consecutive aborts at maximum
389 iteration count (N = 7, dash line), a detectable waveform was considered absent at the
390 applied stimulus level (cycles) and stop command will be triggered to avoid
391 nonproductive attempts at lower levels (triangles).
392
393 **FIG 3.** Precise threshold determination by modeling the change in iteration count
394 with stimulus level. **A** The normalized iteration count vs. level function is fit by
395 sigmoid function (red line). The best-fit function is used to estimate the threshold by
396 level interpolation at 0.9 of the sigmoid growth. **B** Validation of the level interpolation
397 with a dataset of 1-dB spaced stimulus level. The normalized iteration count vs. level
398 function is fit by both sigmoid and exponential functions. The experimentally
399 determined threshold is approximately at the level which corresponds to 1.0 of the
400 best-fit exponential growth and 0.9 of the sigmoid growth.
401
402 **FIG 4.** Threshold determination by cross-correlation analysis in human ABR. **A**
403 Example average responses over 3500 sweeps were recorded from a human
404 participant with the visually identified threshold of 5 dB (bolded). **B** Subgroup
405 averages were used in the algorithm for the cross-correlation analysis. **C** CC peak
406 amplitude was plotted as a function of the level series. **D** Signal lag of the CC peak
407 vs. level function was plotted. At supra-threshold levels (dots) the mean signal lags
408 from three parallel runs were close to time zero (1.85 ± 2.45 data points, mean ± s.d.),
409 whereas large variability (22.67 ± 23.01 data points, mean ± s.d.) were constantly
410 observed at sub-threshold levels (cycles). **E** Plot of the executed iteration count vs.
411 level function. The supra-threshold responses converge lags within seven data points
412 from time zero after finite number of iterations (black dots).

FIG 5. Comparisons between the thresholds determined by the algorithm and human judges. **A** For mouse ABR close matches between the algorithm determined thresholds and those agreed by three out of five human experts (maximum and mean discrepancies, 4 dB and 2.00 ± 1.13 dB, mean ± s.d.). Linear fit: adjust R^2 = 0.97. **B** Comparison between the total sweeps used in the conventional level averaging with a fixed number (left bar) and in our method with varying sweep numbers at different levels (right bar). The latter requires 66.76 ± 4.09 % (mean ± s.d.) fewer sweeps than the former. Note that the sweeps were counted at all supra-threshold and two highest sub-threshold levels. **C** Similar to **A**, matched thresholds in human ABR were reported by both the algorithm and human judges (maximum and mean discrepancies, 2 dB and 0.42 ± 0.83 dB, mean ± s.d.). To illustrate overlapping data points, dots of different sizes were used. Linear fit: adjust R^2 = 1.00. **D** The total number of sweeps used in the conventional level averaging (left bar) vs. in our method (right bar). The latter requires 53.08 ± 12.91 % (mean ± s.d.) fewer sweeps than the former.

REFERENCES

Alpsan, D., and Ozdamar, O. (1991). Brain-Stem Auditory Evoked-Potential Classification by Backpropagation Networks. 1991 Ieee International Joint Conference on Neural Networks, Vols 1-3, 1266-1271.

Barreira-Nielsen, C., Fitzpatrick, E., Hashem, S., Whittingham, J., Barrowman, N., and Aglipay, M. (2016). Progressive Hearing Loss in Early Childhood. Ear Hear *37*, e311-321.

Berninger, E., Olofsson, A., and Leijon, A. (2014). Analysis of click-evoked auditory brainstem responses using time domain cross-correlations between interleaved responses. Ear Hear *35*, 318-329.

Bershad, N.J., and Rockmore, A.J. (1974). On estimating signal-to-noise ratio using the sample correlation coefficient. IEEE Trans Inf Theory *IT-20*, 112-113.

Bramhall, N.F., Konrad-Martin, D., and McMillan, G.P. (2018). Tinnitus and Auditory Perception After a History of Noise Exposure: Relationship to Auditory Brainstem Response Measures. Ear Hear *39*, 881-894.

Castaneda, R., Natarajan, S., Yule Jeong, S., Na Hong, B., and Ho Kang, T. (2019). Electrophysiological changes in auditory evoked potentials in rats with salicylate-induced tinnitus. Brain Res.

Cebulla, M., and Sturzebecher, E. (2015). Automated auditory response detection: Further improvement of the statistical test strategy by using progressive test steps of iteration. Int J Audiol *54*, 568-572.

Cebulla, M., Sturzebecher, E., and Wernecke, K.D. (2000). Objective detection of auditory brainstem potentials: comparison of statistical tests in the time and frequency domains. Scand Audiol *29*, 44-51.

Davey, R., McCullagh, P., Lightbody, G., and McAllister, G. (2007). Auditory brainstem response classification: a hybrid model using time and frequency features. Artif Intell Med *40*, 1-14.

Don, M., and Elberling, C. (1994). Evaluating residual background noise in human auditory brain-stem responses. J Acoust Soc Am *96*, 2746-2757.

Don, M., and Elberling, C. (1996). Use of quantitative measures of auditory brain-stem response peak amplitude and residual background noise in the decision to stop averaging. The Journal of the Acoustical Society of America *99*, 491-499.

Elberling, C. (1979). Auditory electrophysiology. The use of templates and cross correlation functions in the analysis of brain stem potentials. Scand Audiol *8*, 187-190.

Elberling, C., and Don, M. (1984). Quality estimation of averaged auditory brainstem responses. Scand Audiol *13*, 187-197.

Galbraith, G.C., and Brown, W.S. (1990). Cross-correlation and latency compensation analysis of click-evoked and frequency-following brain-stem responses in man. Electroencephalogr Clin Neurophysiol *77*, 295-308.

Gates, G.A., and Mills, J.H. (2005). Presbycusis. Lancet *366*, 1111-1120.

Henry, K.R. (1979). Auditory brainstem volume-conducted responses: origins in the laboratory mouse. J Am Aud Soc *4*, 173-178.

Jewett, D.L., Romano, M.N., and Williston, J.S. (1970). Human auditory evoked potentials: possible brain stem components detected on the scalp. Science *167*, 1517-1518.

Kujawa, S.G., and Liberman, M.C. (2009). Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. J Neurosci *29*, 14077-14085.

Lewis, J.D., Kopun, J., Neely, S.T., Schmid, K.K., and Gorga, M.P. (2015). Tone-burst auditory brainstem response wave V latencies in normal-hearing and hearing-impaired ears. J Acoust Soc Am *138*, 3210-3219.

Lin, X., Li, G., Zhang, Y., Zhao, J., Lu, J., Gao, Y., Liu, H., Li, G.L., Yang, T., Song, L., *et al.* (2019). Hearing consequences in Gjb2 knock-in mice: implications for human p.V37I mutation. Aging (Albany NY) *11*, 7416-7441.

Madsen, S.M.K., Harte, J.M., Elberling, C., and Dau, T. (2018). Accuracy of averaged auditory brainstem response amplitude and latency estimates. Int J Audiol *57*, 345-353.

McKearney, R.M., and MacKinnon, R.C. (2019). Objective auditory brainstem response classification using machine learning. Int J Audiol, 1-7.

Mehraei, G., Hickox, A.E., Bharadwaj, H.M., Goldberg, H., Verhulst, S., Liberman, M.C., and Shinn-Cunningham, B.G. (2016). Auditory Brainstem Response Latency in Noise as a Marker of Cochlear Synaptopathy. J Neurosci *36*, 3755-3764.

Melcher, J.R., Guinan, J.J., Jr., Knudson, I.M., and Kiang, N.Y. (1996). Generators of the brainstem auditory evoked potential in cat. II. Correlating lesion sites with waveform changes. Hear Res *93*, 28-51.

Moller, A.R., and Jannetta, P.J. (1983). Interpretation of brainstem auditory evoked potentials: results from intracranial recordings in humans. Scand Audiol *12*, 125-133.

Parkkonen, L., Fujiki, N., and Makela, J.P. (2009). Sources of auditory brainstem responses revisited: contribution by magnetoencephalography. Hum Brain Mapp *30*, 1772-1782.

Ridley, C.L., Kopun, J.G., Neely, S.T., Gorga, M.P., and Rasetshwane, D.M. (2018). Using Thresholds in Noise to Identify Hidden Hearing Loss in Humans. Ear Hear *39*, 829-844.

Roeser, R.J., Valente, M., and Hosford-Dunn, H. (2007). Audiology. Diagnosis, 2nd edn (New York: Thieme).

Sergeyenko, Y., Lall, K., Liberman, M.C., and Kujawa, S.G. (2013). Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline. J Neurosci *33*, 13686-13694.

Sininger, Y.S. (1993). Auditory brain stem response for objective measures of hearing. Ear Hear *14*, 23-30.

Suthakar, K., and Liberman, M.C. (2019). A simple algorithm for objective threshold determination of auditory brainstem responses. Hear Res *381*, 107782.

Valderrama, J.T., de la Torre, A., Alvarez, I., Segura, J.C., Thornton, A.R., Sainz, M., and Vargas, J.L. (2014). Automatic quality assessment and peak identification of auditory brainstem responses with fitted parametric peaks. Comput Methods Programs Biomed *114*, 262-275.

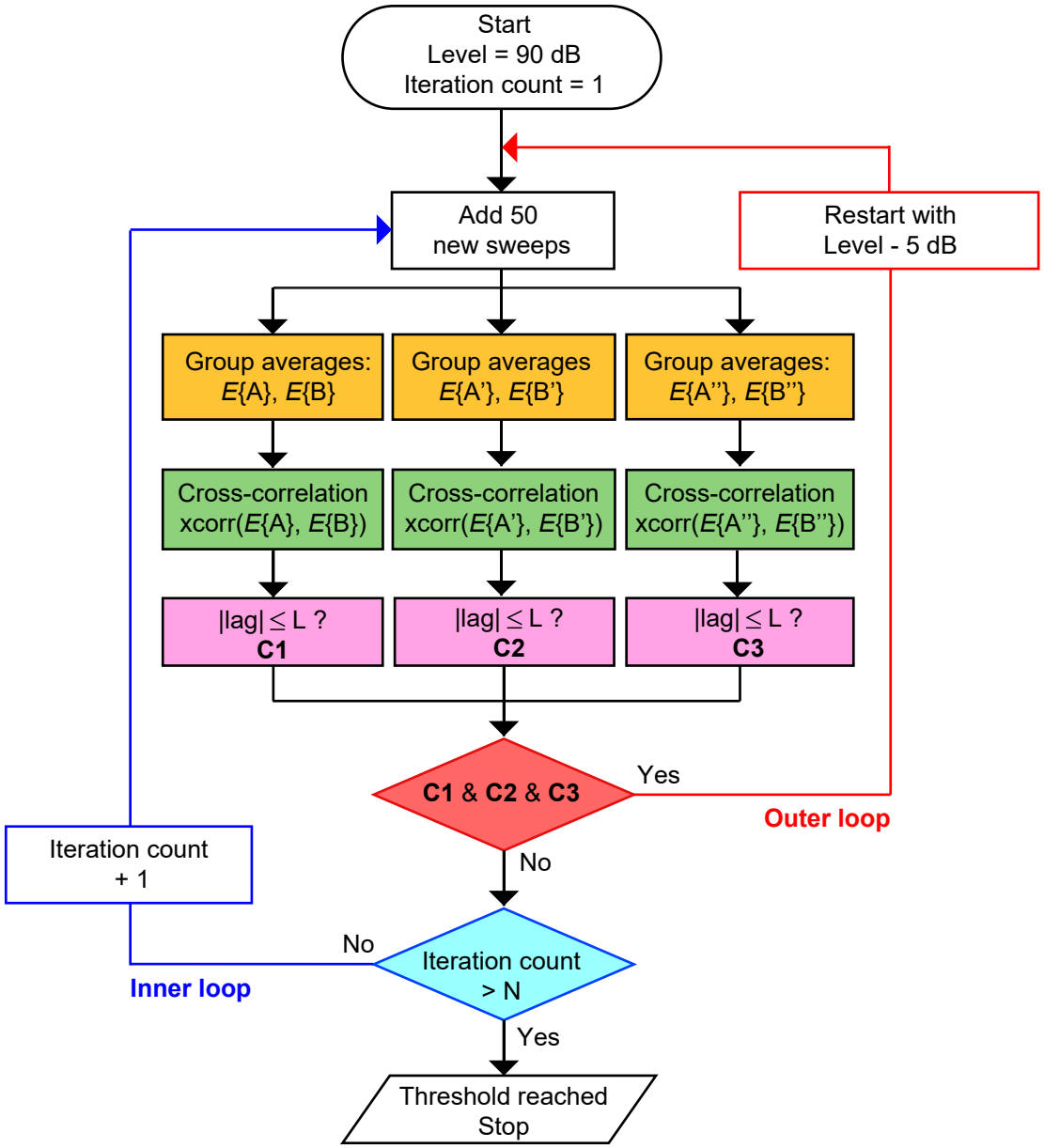
Vidler, M., and Parkert, D. (2004). Auditory brainstem response threshold estimation: subjective threshold estimation by experienced clinicians in a computer simulation of the clinical test. Int J Audiol *43*, 417-429.

Weber, B.A., and Fletcher, G.L. (1980). A computerized scoring procedure for auditory brainstem response audiometry. Ear Hear *1*, 233-236.

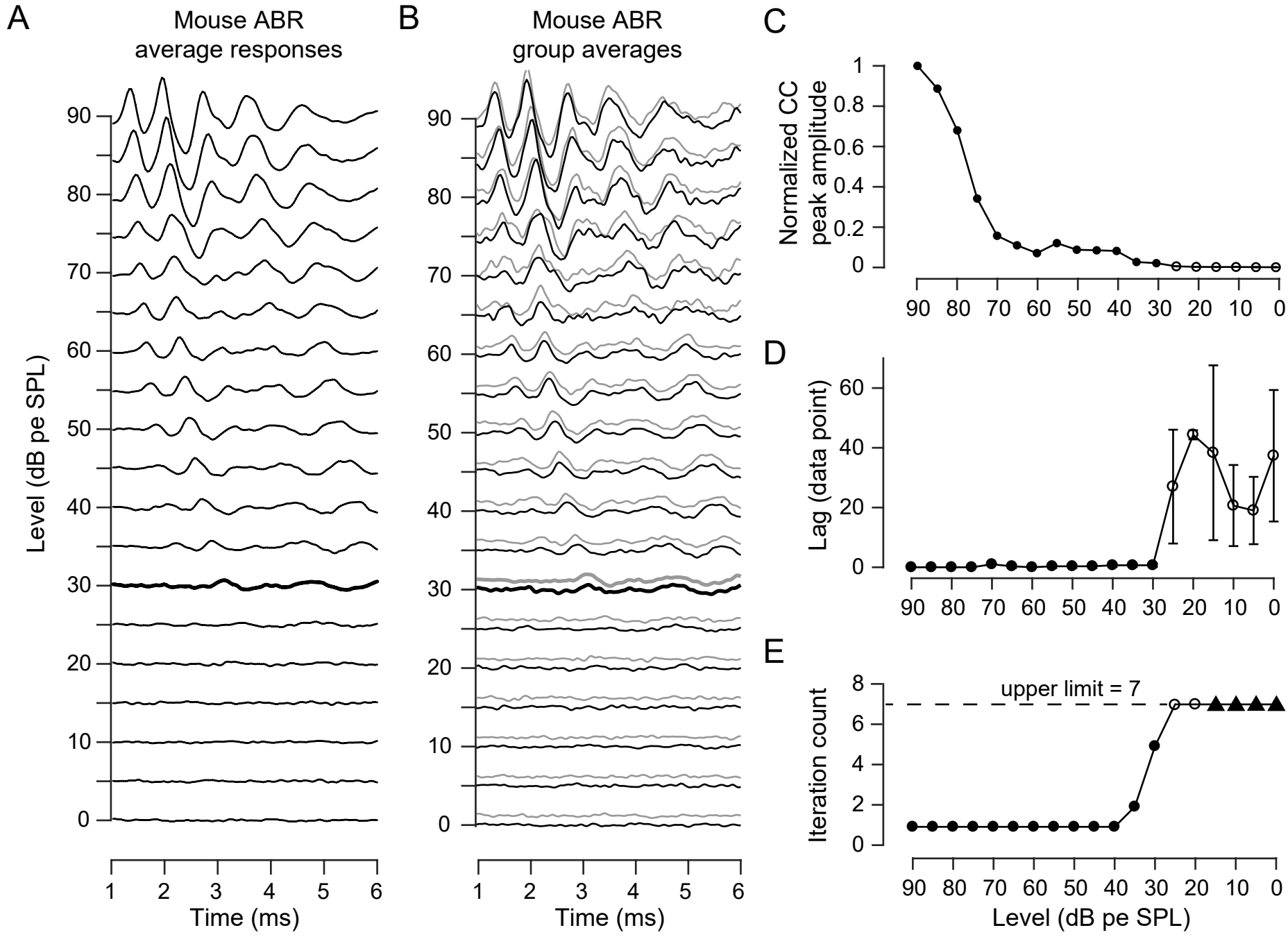
Xu, Z.M., De Vel, E., Vinck, B., and Van Cauwenberge, P. (1995). Application of cross-correlation function in the evaluation of objective MLR thresholds in the low and middle frequencies. Scand Audiol *24*, 231-236.

Table 1 Comparisons between Threshold Determination by Human Experts and the Algorithm

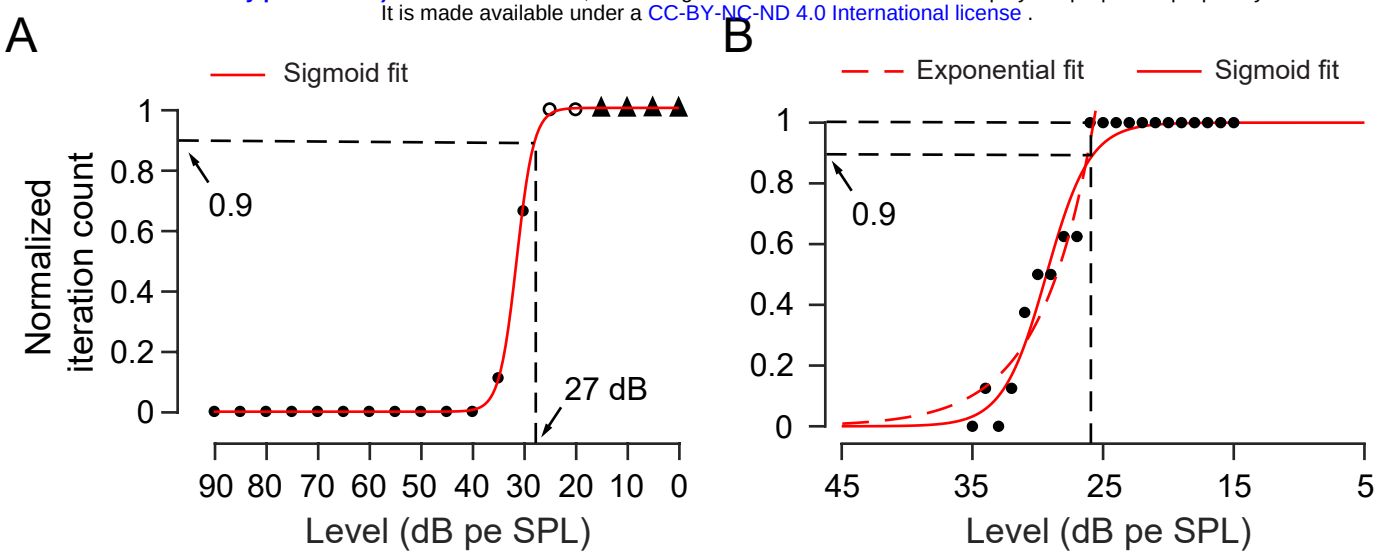
	Algorithm				Human						
	Threshold	Threshold (level interpolated)	R2 of fitting	# sweeps	judge1	judge2	judge3	judge4	judge5	Threshold (by execution judges)	# sweeps
mouse 01 (wt)	20	18	0.9122	1900	25	10	20	20	20	20	5950
mouse 02 (wt)	25	27	0.9314	2000	25	20	25	25	25	25	5600
mouse 03 (wt)	20	17	0.995	1650	25	15	20	15	15	16.67	5950
mouse 04 (terc-/-)	25	26	0.8619	2050	30	30	30	30	25	30	5600
mouse 05 (terc-/-)	30	25	0.9778	1500	25	15	25	25	20	23.33	5250
mouse 06 (terc-/-)	55	49	0.9856	1100	55	45	55	50	45	50	3500
mouse 07 (terc-/-)	30	28	0.9655	1800	25	25	25	25	25	25	5250
mouse 08 (terc-/-)	60	57	0.998	1250	55	50	65	55	55	55	3150
human participant 01	5	NA	NA	20500	5	5	5	5	5	5	45500
human participant 02	5	NA	NA	17500	5	5	5	5	5	5	45500
human participant 03	45	NA	NA	11500	40	45	45	45	40	43.33	17500
human participant 04	5	NA	NA	17500	5	5	5	5	5	5	45500



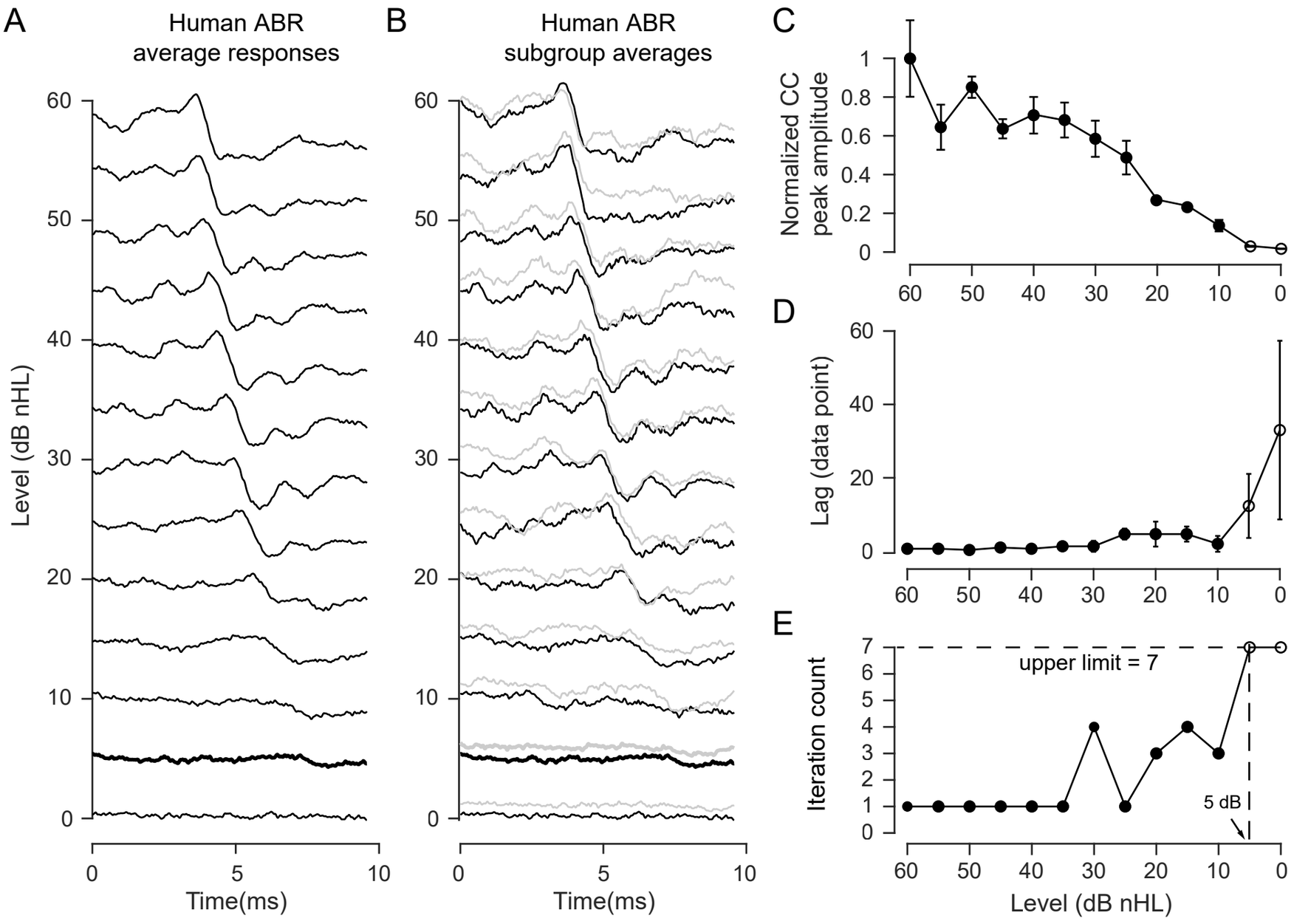
Wang et al., Fig 1



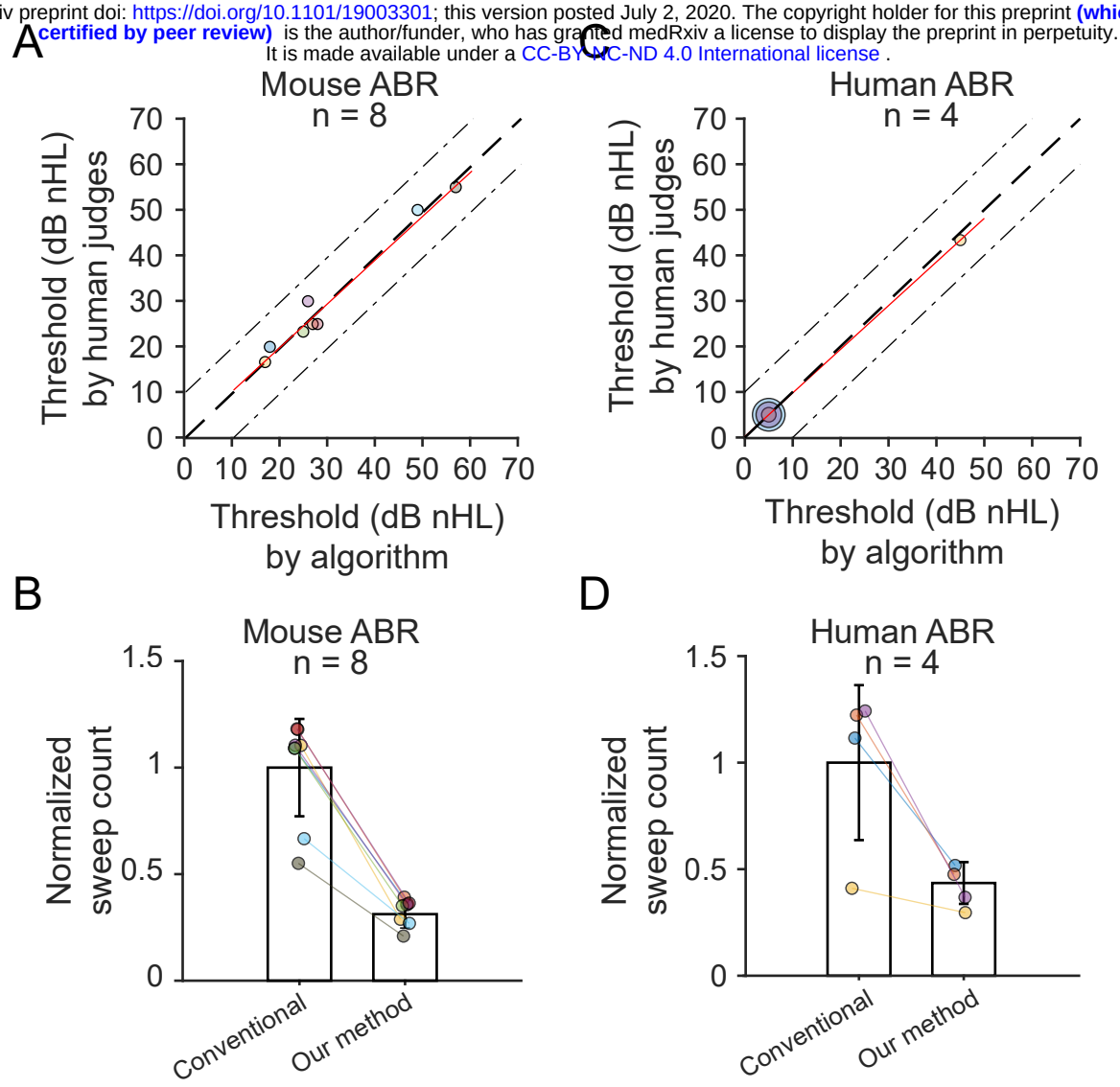
Wang et al., Fig 2



Wang et al., Fig 3



Wang et al., Fig 4



Wang et al., Fig 5