An Efficient and Robust Approach to Detect Auditory Evoked Brainstem Responses using Adaptive Averaging

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

[1]Department of Otolaryngology-Head and Neck Surgery, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, China
[2]Ear Institute, Shanghai Jiao Tong University School of Medicine, China
[3]Shanghai Key Laboratory of Translational Medicine on Ear and Nose Diseases, China
[4]Shanghai Institute of Precision Medicine, Shanghai Ninth People's Hospital, China
[#] These authors contributed equally to this work
* Corresponding authors. E-mail: yunfeng.hua@shsmu.edu.cn (Y.H.) and haowu@sh-jei.org (H.W.)

Abstract

Objective: Auditory brainstem response (ABR) is widely employed to evaluate auditory function of patients and lab animals. Despite attempts in automation, hearing threshold estimation still relies on professional inspection of audiograms. The aim of this study is to develop an automated procedure to detect the hearing threshold.

Study Sample: ABR tests were performed on eight mice and four human participants.

Design: Adaptive averaging strategy was proposed. It models the number of sweeps for averaging that leads to a detectable ABR at progressively reduced loudness level.

Results: Noise-overwhelmed responses at threshold levels lead to an exponential increase in this number, allowing automated threshold estimation with high precision. Furthermore, up to 70% sweeps were saved by stopping ongoing averaging upon detected responses.

Conclusion: Implementation of this method in commercial devices will automate the ABR test in a more reliable and cost-effective fashion.

Introduction

The auditory brainstem responses (ABRs) are brain electrical potential changes due to the 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, making the test wide-employed to exam the hearing function, particularly for infants, adults with learning disabilities and patients undergoing operation, to whom a test through communication or body movements is not applicable (Jacobson et al., 1990). Typical ABR waveform is composed of five peaks in the early onset of sound evoked potentials, corresponding to synchronous activities arising from auditory nerve, cochlear nucleus, superior olivary complex, lateral lemniscus and inferior colliculus, respectively. Thereby, features of the ABR waveform including peak latencies and amplitudes provide clinical-significant information, for instance hidden hearing loss (Mehraei et al., 2016; Ridley et al., 2018), tinnitus (Bramhall et al., 2018; Castaneda et al., 2019), site of lesions or tumors in the auditory system (Roeser et al., 2007) based on the way how the waveform is altered.

Although the test itself is considered objective, the waveform recognition at near-threshold levels is not always trivial. Currently, professionals are still required to conduct recording and audiogram assessment, which raises largest cost component. Moreover, often limited consensus about the audiogram assessments occurs (Vidler and Parkert, 2004), because accurate waveform recognition highly depends on the skill and the experience of the assessors, especially when untypical waveforms or

high background activities are encountered. As diagnosis of progressive and acquired hearing loss in preschool children (Lu et al., 2014; Lu et al., 2011) as well as categorizing temporal threshold shifts caused by noise exposure (Le Prell, 2019) require comprehensive hearing test with high accuracy and robustness, there is an increasing demand for automatic approaches making the ABR test less labor-intense.

The fundamental challenge of automated ABR analysis is contamination of high-level background and intrinsic noise which requires exclusively averaging over hundreds of sweeps to improve the signal-to-noise ratio (SNR) for unambiguous waveform recognition. Hence, by default audiograms are provided with statistical measures like well-known single- and multiple-point F-ratio (F_{sp} and F_{mp}) in order to evaluate the signal and noise characteristics (Don and Elberling, 1994; Elberling and Don, 1984). Over decades, several methods have been proposed to automate the procedure. Feature-based strategies detect responses by quantifying the similarity of the measured waveforms to either existing templates (Valderrama et al., 2014) or waveforms evoked by higher loudness level (Suthakar and Liberman, 2019), as well as by recognizing features learned by artificial neural network from human-annotated datasets (Alpsan and Ozdamar, 1991; McKearney and MacKinnon, 2019). However, the detection accuracy is often limited owing to waveform heterogeneity, varying data quality, and inconsistent training data (McKearney and MacKinnon, 2019). Alternatively, statistics-based strategies quantify either inter-sweep variability using correlation function (Bershad and Rockmore, 1974; Weber and Fletcher, 1980; Xu et al., 1995) or the 'signal quality' through scoring procedures like F-ratios (Don and

Elberling, 1994; Elberling and Don, 1984), but both require calibrations to predetermine arbitrary boundary for signal thresholding. Although some approaches have been integrated in commercial devices, reliable assessment still involves human supervision due to frequent detection errors.

Here we developed a novel procedure which detects ABRs in an SNR-adaptive fashion using iterative averaging strategy on single-sweeps. By modeling the minimal-required sweep numbers for detectable evoked responses, automated ABR threshold estimation was achieved. Tests were done with the ABR recordings from mice and human participants, and algorithm outcomes were found matching the expert assessments with a maximal aberration of 5 dB. In addition, the numbers of adaptive averaged sweeps are instructive for an efficient sampling strategy. When implemented in commercial devices, saving up to 69 % redundant sweeps is expected if ongoing averaging is stopped upon detected responses.

Materials and methods

Animals, human participants and ethics

C57BL/6 mice were purchased from Sino-British SIPPR/BK Lab.Animal Ltd. (Shanghai, China). The terc-/- mice were kindly donated by Prof. Lin Liu (Nankai University, China) and bred in house. Participants were recruited from Shanghai Ninth People's Hospital and consent forms were signed before the experiment. This study is conduct 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 recordings

Mouse ABRs were recorded using TDT RZ6/BioSigRZ system (Tuck-Davis Tech. Inc., US) in a sound-proof chamber. 7-week-old mice were anesthetized through intraperitoneal injection of hydrate chloride (500 mg/kg). During the test, animal body temperature was maintained at 37°C using a regulated heating pad (Harvard Apparatus, US) and rectal thermal probe. Electrical potentials were registered via subdermal needle electrodes (Rochester Electro-Med. Inc., US) placed at animal vertex (record electrode), left infra-auricular mastoid (reference electrode) and right shoulder region (ground electrode). 3-ms tone pips at 16 kHz were delivered via a multi-field magnetic speaker (Tuck-Davis Tech. Inc., US) positioned at 10 cm from the animal vertex. Evoked potentials upon repeated stimuli (50 Hz) were sampled at

24 kHz. 500 sweeps were acquired at each test loudness level starting from 90 dB to 0 dB with 5-dB decrement. For one animal, we acquired potentials at predetermined threshold loudness levels (±10 dB with 1-dB decrement).

Human ABRs were recorded via commercial ABR device (Intelligent Hearing Systems, US) with Smart EP software from 4 adults aged between 21 and 29 with normal hearing function. An electromagnetically shielded insert earphone (ER-3) was applied to deliver click stimulation to the left ear at a rate of 37.1/sec. The recorded potentials were amplified by a factor of 100,000 and filtered with 100 Hz (high-pass) and 3000 Hz (low-pass) filters. Averaged responses over 500, 1000, and 2000 sweeps were acquired three times at each loudness level starting from 60 dB to 0 dB with 5-dB decrement.

Analysis of single-sweep correlation

Mouse ABR sweeps were analyzed using self-written routines in MATLAB (MathWorks Inc., US). The rejection amplitude of artifacts was set to 55 μV and a high-pass filter (smoothing splines with a smoothing parameter of 0.5) was applied to the raw recordings. The pairwise correlation coefficients were computed using MATLAB function in 350 sweeps within a temporal window of 1-6 ms post-stimulation-onset, resulting in 61,075 coefficients at each loudness level. The histograms of the coefficients were fitted with normal distributions.

Threshold detection using adaptive averaging method

For mouse ABR, all 350 sweeps were loaded progressively by seven iterations (50 sweeps per iteration). Within each iteration loaded sweeps were subdivided randomly into two groups for separated averaging. Cross-correlation with group averages was computed and time-locked components were detected by whether maximum overlap of group averages was found with neglectable lag (< 1% of the analyzed sweep length, 0-6 ms post-stimulation onset). In order to reject coincidently overlapped signals with desired lag, cross-correlation operation was performed three more times with regrouping. For each loudness level, iterations were counted till either time-locked response was detected or executed iterations reached maximum owing to total number of sweeps. A sigmoid function (1) was employed to model the executed iteration counts after normalization *NC* at the loudness level *S* and function growth of 0.9 corresponds to the presumable ABR threshold *T*. The best fit-parameter $k = 0.6$.

$$NC_S = \frac{1}{1+e^{-k(S-T)}} \qquad (1)$$

As for human ABR, average responses over 500, 1000 and 2000 sweeps were loaded iteratively. Three cross-correlation operations were done with the combination of two out of three average responses which were acquired consecutively. The maximum iteration count was seven which is limited by maximum average of 3500 sweeps. Note that for iterations requiring averages over 1500 or more than 2000 sweeps, the virtual average responses *avg(M+N)* of *M* plus *N* sweeps were generated using weighted averaging of *avg(M)* and *avg (N)* with following equation (2).

$$avg(M+N) = \frac{M \bullet avg(M) + N \bullet avg(N)}{M+N} \qquad (2)$$

The lag condition for time-locked component was set to 2% of the analyzed sweep

length (0-10 ms post-stimulation onset). Same as for the mouse ABR, the human ABR thresholds were interpolated with the best fit at the function growth of 0.9.

Audiogram assessment

Human ground truth for the ABR threshold was obtained from audiogram inspection by five independent clinicians who were blinded to the test subject identities. The audiograms contained average responses either over total sweeps (the conventional averaging) or over the adaptively averaged sweeps used in the algorithm (the adaptive averaging). The means were calculated from collected results.

Results

Correlations between single-sweep ABR

As ABRs are usually embedded in high-level background activities and system noise, smooth baseline and clear waveform, if present, are usually obtained after averaging over hundreds of sweeps (**Figure 1A**). From mice we recorded single sweeps upon 16 kHz tone-pips at different loudness levels. After exclusion of a small fraction with movement-induced excessive amplitudes, more than 350 sweeps were then filtered sequentially through a high-pass filter (Nishida et al., 1993) for removal of low frequency fluctuations in short-latency components (**Suppl. Figure 1**). Pairwise correlation coefficients (CCs) between single sweeps were computed. The CCs at the same loudness levels were found following normal distribution, of which supra-threshold stimuli caused positive shifts (**Figure 1B**), indicating elevated degrees of correlation. Together with prior study (Galbraith, 2001), this result suggests that single-sweep ABRs are highly correlated and quantifiable. Moreover, gaussian fits resulted in similar peak widths (0.1778 ± 0.0154) which was a presumable consequence of constant noise level.

Thresholding ABR using CC and F_{sp}

Next, we tried to estimate the ABR threshold by quantifying CCs. Comprehensive ABR tests were performed starting from 90 dB to 0 dB on three wildtype (wt) and five telomerase-knock-out mice (terc-/-) with high deafness risk. Example audiograms of a wt and a terc-/- animal were illustrated in **Figure 2A** with corresponding medians

of CCs in **Figure 2B**. A rising phase of CCs was found postponed by about 30 dB, consistent with the threshold elevation suggested by audiogram inspection. This result supported CC as a promising tool for ABR threshold detection, as derived in early theoretical work (Bershad and Rockmore, 1974). In order to obtain an arbitrary boundary of CC for supra-threshold responses, we aligned the CCs to the human-assessed thresholds (**Figure 2C**) and found supra-threshold responses having a minimal CC of 0.0119 ± 0.0034. Although this empirically determined boundary can be utilized to threshold responses, limited accuracy was expected owing to large inter-trial variability presumably caused by experimental condition changes (such as electrode position, depth of anesthesia, etc.) during multiple recording sessions. Varying SNR was found being a common issue for approaches if this kind. When F_{sp} was applied in the replacement of CC (**Figure 2D** & **E**) to the same dataset, large variation (3.09 ± 1.53) persisted in the calibrated detection boundary, hampering accurate threshold estimation.

Threshold detection using adaptive averaging method

As the noise level is suppressed exponentially by averaging, it is rational to assume that to reach certain SNR level weak responses require more sweeps for averaging than strong responses. Next, we designed a novel procedure of detecting ABRs with SNR-adaptive averaging strategy (**Figure 3A**). At each test loudness level, recorded sweeps were randomly divided into two groups and cross-correlation was measured between group averages (**Figure 3C,** see **Figure 3B** for all-sweep averages). For

detectable responses, one expected neglectable lags where maximum correlation was located (**Figure 3D**). To reject potential false positives owing to coincidently overlapped noise peaks at desired lag, three parallel measurements were performed with regrouped sweeps (**Figure 3A**). For the lack of a detectable response, additional sweeps were added iteratively to the random-split averaging till time-locked component was confirmed by three parallel runs or the upper iteration limit was reached (**Figure 3A**). As illustrated in **Figure 3E**, the counts of executed iterations were found increasing rapidly at near-threshold loudness levels and reaching a plateau of maximum. The sharp transition narrowed down loudness levels close to the ABR threshold. Further attempt was made to estimate threshold beyond the loudness decrement by means of modeling the iteration counts (**Figure 3F**). From one animal, we acquired sweeps around predetermined threshold loudness levels with 1-dB decrement and fitted the executed iteration counts with a sigmoid function. Comparable fit parameters were obtained irrespective of 1-dB or 5-dB loudness decrement (**Figure 3F, inset**), validating the interpolation approach. Alternatively, similar results were obtained when arbitrary F_{sp} was utilized for response detection (**Suppl. Figure 2**).

Accuracy and efficiency of the automated threshold detection

To evaluate the performance of the adaptive averaging method, we compared the ABR thresholds determined by the algorithm and audiogram inspection on both mouse and human ABR. Note that for human ABR, average responses over different

sweep numbers were acquired in individual runs instead of from random-split averaging, as the commercial devices were not optimized for single-sweep registration and thereby acquiring single-sweep dataset was not feasible within reasonable duration (see **Suppl. Figure 3A** for the modified flowchart). We found that the ABR threshold estimated by adaptive averaging method matched the mean audiogram readouts of five independent clinicians with a maximum aberration of 5 dB (2.5 ± 1.6, **Figure 4A** & **Suppl. Figure 4A**), indicating a detection accuracy of near-human-performance. Besides, inspection of audiograms with adaptively averaged ABRs gave almost identical results (**Figure 4B** & **Suppl. Figure 4B**), implying reliable threshold detection with adaptively averaged sweeps. When compared to the conventional averaging with a predefined sweep number, the adaptive averaging required 69.39 ± 17.45 % and 36.19 ± 17.53 % less sweeps of mouse and human ABR without compromising threshold detection accuracy (**Figure. 4C & Suppl. Figure 4C**).

Discussion

Attempts have been made to automate ABR analysis since decades. Several techniques were proposed including the usage of correlation function to detect time-locked ABRs from random fluctuations. In this study we found pairwise CCs following single normal distribution and reflecting the power of evoked responses over spontaneous activities. The observed constant distribution broadening (**Figure 1B**) is presumably caused by weak correlations of background noise, implying high fidelity of auditory transmission and stable noise level of single recording session, otherwise multiple peaks or varying peak width would appear. That means it is crucial to keep identical experimental settings including ABR device, electrode position, anesthesia depth, etc. for robust response detection, which is however impractical for changing test subjects and multiple recording sessions. In fact, we showed large variations in the CC and F_{sp} values at the threshold level (**Figure 2C** & **2E**) which was about the same magnitude to the differences caused by 10-20 dB peri-threshold loudness changes, compromising detection accuracy and thereby hampering a plug-in application of these approaches.

We solved this long-lasting limitation by using adaptive averaging strategy which keeps the ABR detection boundary unchanged and adapts averaging to the SNR of recordings. Prior studies used the cross-correlation lag between either (1) two consecutive recordings (Xu et al., 1995) or (2) recordings at two close loudness levels (Suthakar and Liberman, 2019) to detect auditory evoked responses. Here we utilized the same detection method but in a new procedure of iterative averaging: one starts

with sweeps of unknown signal power over a constant noise level. As the noise was suppressed progressively by iterations with increment of sweeps for averaging, true response is no longer overwhelmed and becomes detectable by means of either cross-correlation (**Figure 3**) or F_{sp} measures (**Suppl. Figure 2**). Thereby, converging lag requires fewer iterations if strong responses are present than weak responses, whereas non-converging lag suggests the absence of any time-locked responses. This was supported by our observation of minimum iteration counts at supra-threshold levels (**Figure 3E**) and persisting fluctuation in lags at sub-threshold levels (**Figure 3D**). The sharp transition of the executed iteration counts (equivalent to sweep numbers in averaging) was found sensitive to the near-threshold loudness levels, enabling an accurate and calibration-free threshold detection. We showed that precise threshold estimation was achieved beyond the test loudness decrement by the model interpolation (**Figure 3F**), in our case up to 1 dB. Further extension of this method is to combine with sampling strategy of progressively reduced loudness decrement and increased sweep increment (Cebulla and Sturzebecher, 2015) at peri-threshold loudness levels in a target-precision-dependent manner.

For both mouse and human ABR, the adaptive averaging method was proven reliable in threshold detection and adaptively averaged responses does not cause notable aberration from audiogram assessment (**Figure 4** & **Suppl. Figure 4**). More importantly, it provides audiograms with sufficient SNRs for unambiguous waveform recognition by both human and machine. Such standardized data quality will benefit machine-learning-based approaches by minimizing discrepancy in human-annotations

(McKearney and MacKinnon, 2019). When to stop averaging is an important decision to make during ABR recording, it not only makes the ABR test more efficient by avoiding prolonged acquisition and redundant data, but also specifies the confidence level of waveform interpretation (Don and Elberling, 1996; Madsen et al., 2018). As the sweep number in adaptive averaging depends on the SNR of recordings, instruction of stop averaging was provided upon detected responses. Our quantification suggested up to 69 % mouse ABR sweeps and 36 % human ABR sweeps were saved using the adaptive averaging (**Figure 4C** & **Suppl. Figure 4C**), implying the potentials of bursting the regular ABR test in speed and efficiency.

Acknowledgements

We would like to thank Yun Li, Kun Han, Yan Ren, Lu Yang and Haifeng Li from Shanghai Ninth People's Hospital for help with audiogram assessment. We would like to thank Drs. Guangming Chen and Lin Liu for providing terc-/- mice. This study was supported by The Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (QD2018015 to Y.H.), Elite Program at Shanghai Ninth People's Hospital (JY201802 to Y.H., Z.H.), The National Science Foundation for Young Scientists of China (81800901 to Y.H., 81700903 to B.L) and SHIPM-mu Fund (JC201808 to B.L.).

Declarations of interest

The authors have declared that no conflicts of interest exist.

Author contribution

Y.H. H.W. designed the study; H.Y.W., B.L., X.D. acquired the data; Y.H., H.Y.W. wrote the algorithm; Y.H., H.Y.W., X.W., Z.H., analyzed and interpreted the data. Y.H., H.Y.W., H.W. wrote the paper.

Figure Legends

Figure 1. Pairwise CCs of single-sweep ABR recordings. **(A)** Example sweeps recorded at four loudness levels (grey lines). Responses (colored lines) were obtained

by averaging over 350 sweeps. **(B)** Histograms of pairwise CCs from (A). Fitting of the histograms with normal distributions (colored lines) resulted in a mean full-width-at-haft-maximum of 0.1778 ± 0.0154.

Figure 2. Threshold estimation using CC and F_{sp} **(A)** Example audiograms of average responses over 350 sweeps recorded from a wt mouse (black lines) and a deaf terc-/- mouse (red lines). Attenuated ABR amplitudes were observed: > 25 dB for animal of normal hearing (black asterisk) and > 55 dB for deaf animal (red asterisk). **(B)** Pairwise CCs of sweeps recorded at test loudness levels. The median values were fitted with sigmoid function. A delayed raising phase of about 30 dB in loudness was observed from the terc-/- mouse (red) compared to the wt (black). **(C)** Median CCs at peri-threshold loudness levels. Eight measurements were aligned to the human-assessed thresholds. Mean CC value at the threshold was 0.0119 ± 0.0034 (varying from 0.0090 to 0.0199). **(D)** Similar to (B), F_{sp} values at test loudness levels were fitted with a sigmoid function. **(E)** Similar to (C), mean value of F_{sp} at the threshold was 3.09 ± 1.53.

Figure 3. Threshold estimation using adaptive averaging method. **(A)** Flowchart of the adaptive averaging routine, in which argA, argB, argC represent the subgroup averages (increment of 50 sweeps per iteration), xcorr the cross-correlation operation, k the condition for time-locked components (here k = 1 data point), r1, r2, r3 results from three parallel runs, as well as Max. iterations the maximum executed iterations

(here Max. iterations = 7 for total 350 sweeps). **(B)** Example audiogram averaged over 350 sweeps. The inspected threshold was around 30 dB (black asterisk). **(C)** Subgroup averages (black and grey lines) were obtained by averaging over random-split sweeps. **(D)** The absolute cross-correlation lags between subgroup averages were obtained by three-fold measurements. For supra-threshold responses (dots) neglectable lags were observed in all three parallel runs, while lags with large inter-trial variability (31.11 ± 16.12) were characteristic for sub-threshold responses (cycles). **(E)** 50 new sweeps were added iteratively for random-split averaging, until desired lag requirement was fulfilled. Executed iteration counts for detected responses (dots) and maximum counts at sub-threshold levels (cycles) were plotted. When the sub-threshold response was confirmed by two consecutive loudness levels (black cycles), abort operation was implemented to terminate recording at lower loudness levels (red cycles). **(F)** The normalized executed iteration counts were fitted by sigmoid function to interpolate the precise threshold beyond the decrement of test loudness levels. Inset: validation of the interpolated threshold with tests of 1-dB decrement. The ABR threshold corresponded to the growth of 0.9 in the best sigmoid fit.

Figure 4. Performance of the adaptive averaging method. **(A** & **B)** For mouse ABR, audiograms (n = 8) were either generated by averaging all 350 sweeps (A, conventional averaging) or adaptively averaged sweeps in the algorithm (B, adaptive averaging). The ABR thresholds assessed independently by five clinicians were

consistent with the algorithm outcomes (aberration $\delta = 2.50 \pm 1.60$ dB in A and 2.13 ± 2.47 in B). Linear fit: adjust $R^2 = 0.9587$ in A and 0.9713 in B. **(C)** Comparison between the number of sweeps used in conventional averaging (left bar) and in adaptive averaging (right bar). Note that the sweep numbers were counted for all supra-threshold and first two sub-threshold levels. The adaptive averaging required 69.39 ± 17.45 % less sweeps than the conventional averaging.

References

Alpsan, D., and Ozdamar, O. (1991). Brain-Stem Auditory Evoked-Potential Classification by Backpropagation Networks. Ieee Ijcnn, 1266-1271.

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.

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. J Acoust Soc Am *99*, 491-499.

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

Galbraith, G.C. (2001). Enhanced brainstem and cortical evoked response amplitudes: single-trial covariance analysis. Percept Mot Skills *92*, 659-672.

Jacobson, J.T., Jacobson, C.A., and Spahr, R.C. (1990). Automated and conventional ABR screening techniques in high-risk infants. J Am Acad Audiol *1*, 187-195.

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.

Le Prell, C.G. (2019). Effects of noise exposure on auditory brainstem response and speech-in-noise tasks: a review of the literature. Int J Audiol *58*, S3-S32.

Lu, J., Huang, Z., Ma, Y., Li, Y., Mei, L., Yao, G., Wang, Y., Shen, X., and Wu, H. (2014). Comparison

between hearing screening-detected cases and sporadic cases of delayed-onset hearing loss in preschool-age children. Int J Audiol *53*, 229-234.

Lu, J., Huang, Z., Yang, T., Li, Y., Mei, L., Xiang, M., Chai, Y., Li, X., Li, L., Yao, G., *et al.* (2011). Screening for delayed-onset hearing loss in preschool children who previously passed the newborn hearing screening. Int J Pediatr Otorhinolaryngol *75*, 1045-1049.

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.

Nishida, S., Nakamura, M., and Shibasaki, H. (1993). Method for single-trial recording of somatosensory evoked potentials. J Biomed Eng *15*, 257-262.

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).

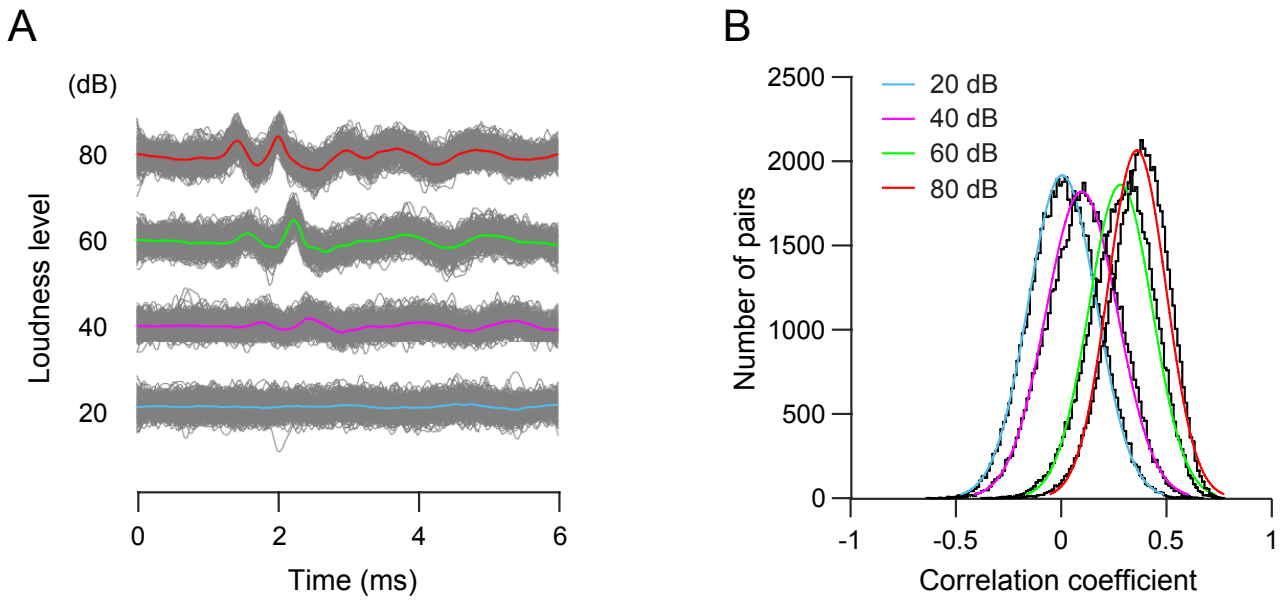
Suthakar, K., and Liberman, M.C. (2019). A simple algorithm for objective threshold determination of auditory brainstem responses. Hear Res *381*, 107782.

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.

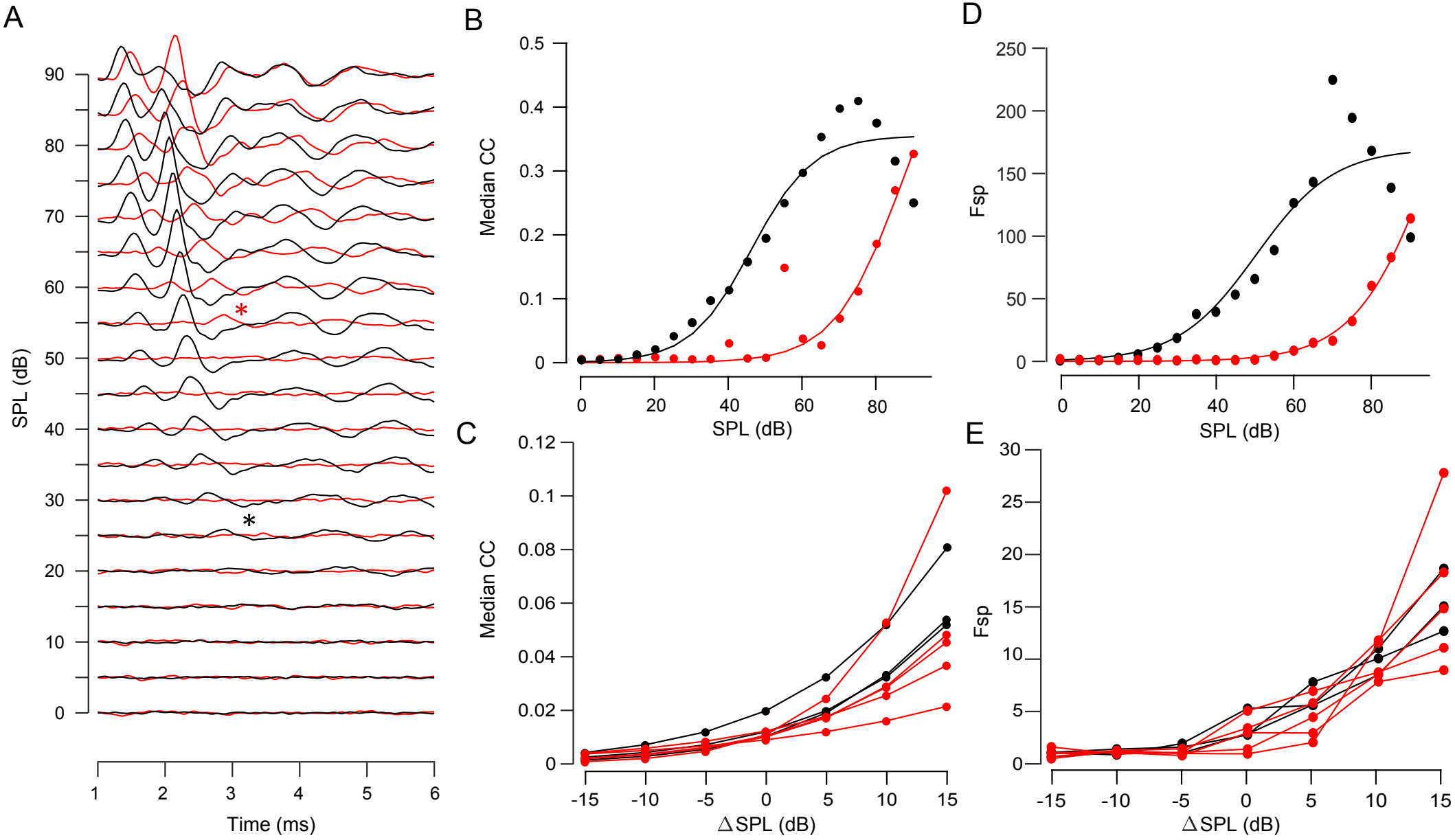
Vidler, M., and Parkert, D. (2004). Auditory brainstem response threshold estimation: subjective threshold estimation by experienced clinicians in a computer simulation of the clinical test. Int J Audiol *43*, 417-429.

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

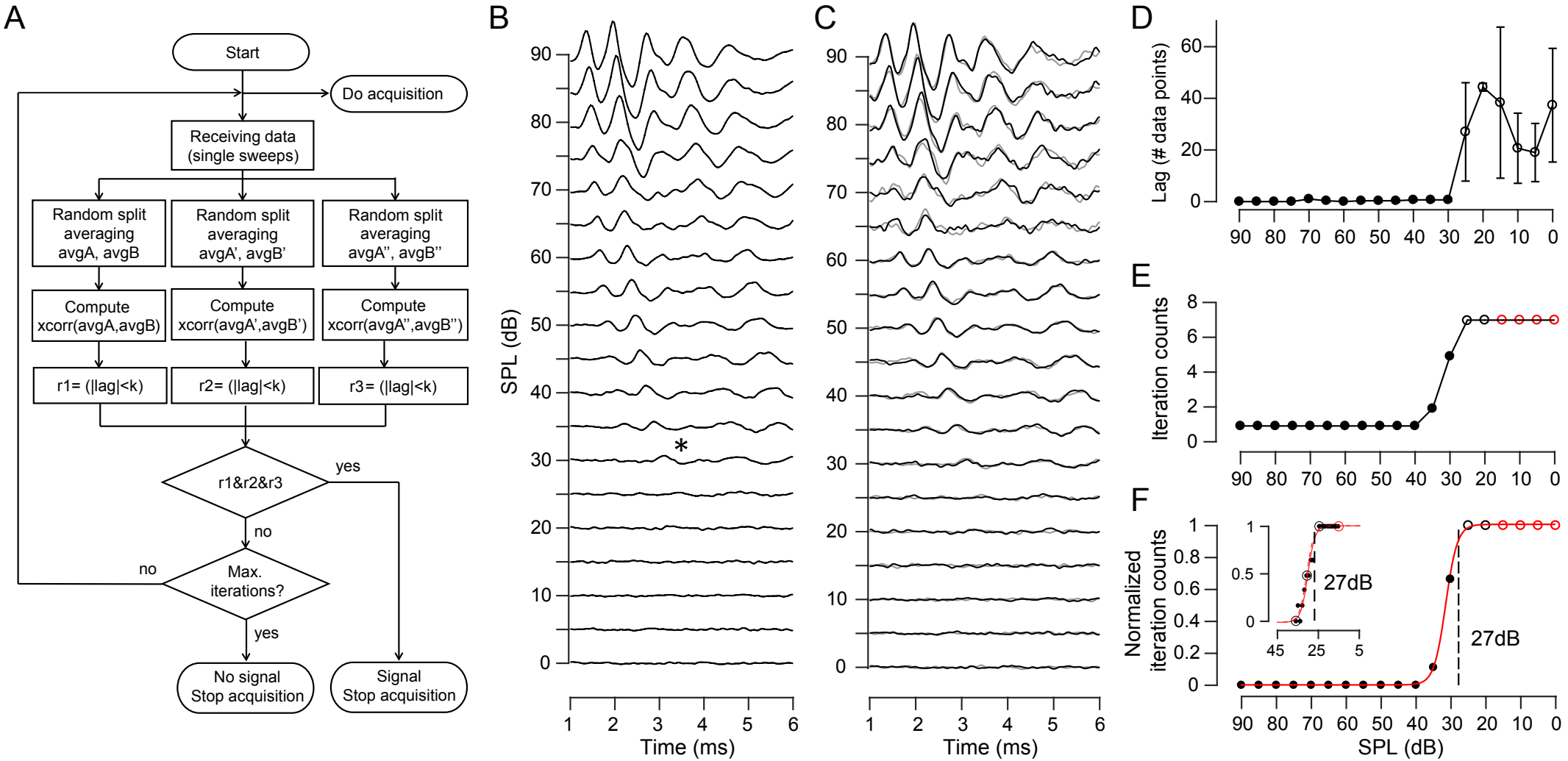
Xu, Z.M., De Vel, E., Vinck, B., and Van Cauwenberge, P. (1995). Application of cross-correlation function in the evaluation of objective MLR thresholds in the low and middle frequencies. Scand Audiol *24*, 231-236.



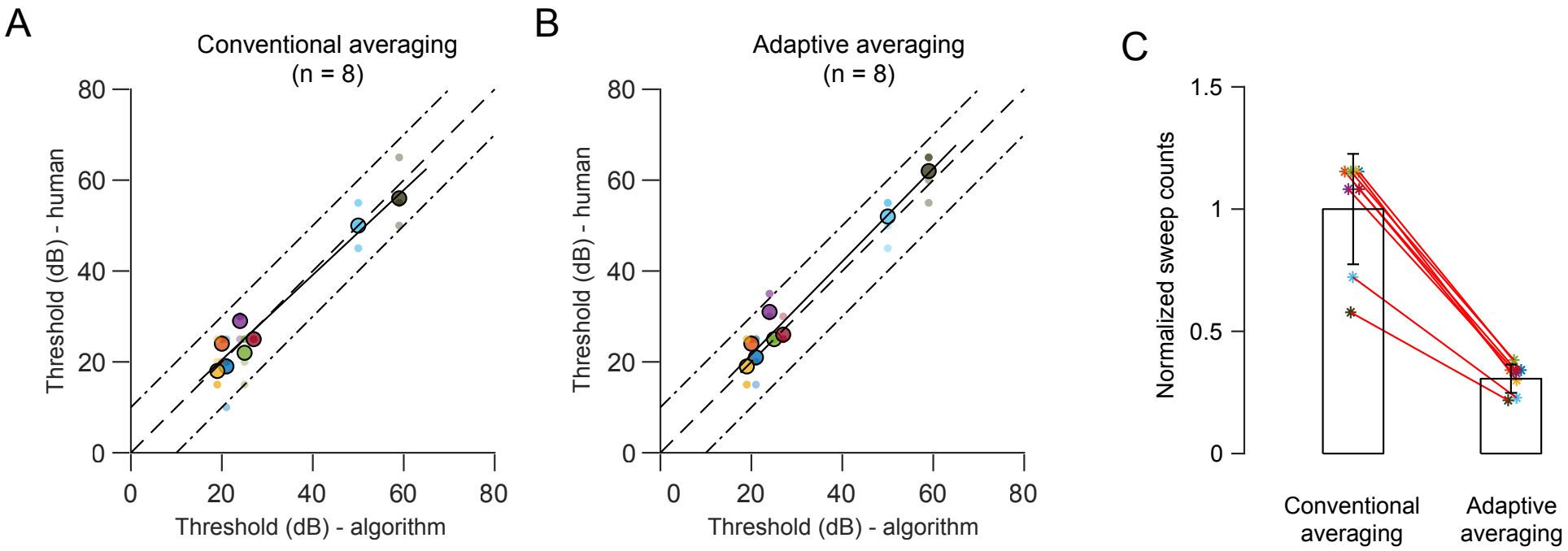
Wang et al., Fig 1



Wang et al., Fig 2



Wang et al., Fig 3



Wang et al., Fig 4