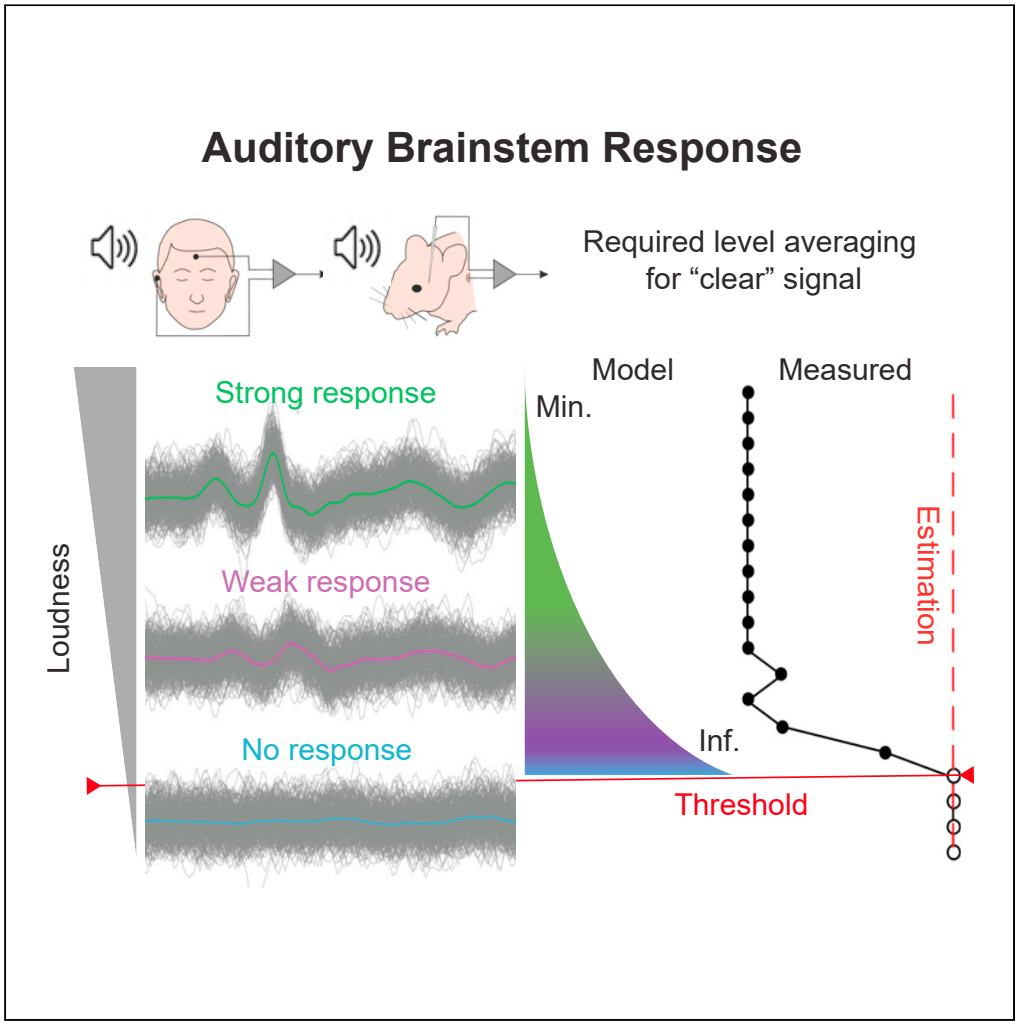
iScience

CellPress
OPEN ACCESS

Article

Real-time threshold determination of auditory brainstem responses by cross-correlation analysis



Haoyu Wang, Bei Li, Yan Lu, ..., Zhiwu Huang, Lei Song, Yunfeng Hua

lei.song@yale.edu (L.S.)
yunfeng.hua@shsmu.edu.cn (Y.H.)

Highlights

Automatic threshold determination of auditory brainstem response (ABR)

Detection of "clear" responses from iteratively averaged level representation

Wide application in both animal and human ABR tests

Stop on-going level averaging based on detection outcome

Wang et al., iScience 24, 103285
November 19, 2021
https://doi.org/10.1016/j.isci.2021.103285

iScience

CellPress
OPEN ACCESS

Article

Real-time threshold determination of auditory brainstem responses by cross-correlation analysis

Haoyu Wang,[1,2,3,4,6] Bei Li,[1,6] Yan Lu,[4] Kun Han,[1] Haibin Sheng,[1,2,3] Jialei Zhou,[5] Yumeng Qi,[4] Xueling Wang,[1,2,3] Zhiwu Huang,[1,2,3] Lei Song,[1,2,3,*] and Yunfeng Hua[1,2,3,4,7,*]

SUMMARY

Auditory brainstem response (ABR) serves as an objective indication of auditory perception at a given sound level and is nowadays widely used in hearing function assessment. Despite efforts for automation over decades, ABR threshold determination by machine algorithms remains unreliable and thereby one still relies on visual identification by trained personnel. Here, we described a procedure for automatic threshold determination that can be used in both animal and human ABR tests. The method terminates level averaging of ABR recordings upon detection of time-locked waveform through cross-correlation analysis. The threshold level was then indicated by a dramatic increase in the sweep numbers required to produce "qualified" level averaging. A good match was obtained between the algorithm outcome and the human readouts. Moreover, the method varies the level averaging based on the cross-correlation, thereby adapting to the signal-to-noise ratio of sweep recordings. These features empower a robust and fully automated ABR test.

INTRODUCTION

The auditory brainstem responses (ABRs) are brain electrical potential changes due to synchronous neuronal activities evoked by suprathreshold 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 used for hearing function assessment. In rodent and cat, a typical ABR waveform is composed of initial five peaks in the early onset of sound-evoked potentials, representing 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 human has slightly different peak generators as demonstrated with intracranial recordings (Moller and Jannetta, 1983) and neuromagnetic responses (Parkkonen et al., 2009). Nowadays, ABRs are of high clinical relevance for objective analysis of hearing function, especially for newborn hearing screening, screening for auditory neuropathy, acoustic neuroma, and central hearing loss (Lewis et al., 2015; Roeser et al., 2007), as well as possible in the future for assessing "hidden hearing loss" (Kujawa and Liberman, 2009; Mehraei et al., 2016; Ridley et al., 2018).

Although the ABR test itself is an objective measurement, the determination of threshold involves human interpretation of the ABR waveform. The readout of ABR threshold requires a trained personnel to supervise waveform recognition, which is labor demanding. Besides, such interpretations oftentimes are subjective and may introduce errors that vary from person to person depending on his/her skill and experience, especially in cases with atypical waveform or with high background noise (Vidler and Parkert, 2004). In order to accurately detect mild hearing threshold elevation in the diagnosis of, e.g., progressive hearing loss (Barreira-Nielsen et al., 2016) and age-related hearing loss (Gates and Mills, 2005; Sergeyenko et al., 2013), unbiased automatic approaches with high precision and reliability are essential, particularly when screening is involved. Over decades, many attempts were made to automate the procedure based on, e.g., (1) waveform similarity by means of comparing either existing templates (Davey et al., 2007; Elberling, 1979; Valderrama et al., 2014) or matching features learned by artificial neural network from human annotation (Acır et al., 2006; Alpsan and Ozdamar, 1991; McKearney and MacKinnon, 2019; Sanchez et al., 1995; Vannier et al., 2002); (2) waveform stability quantified by cross-correlation function between single-sweeps

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

[5]Department of Otorhinolaryngology-Head & Neck Surgery, Shanghai Children's Hospital, Shanghai Jiao Tong University, Shanghai, China

[6]These authors contributed equally

[7]Lead contact

*Correspondence: lei.song@yale.edu (L.S.), yunfeng.hua@shsmu.edu.cn (Y.H.)

https://doi.org/10.1016/j.isci.2021.103285

(Bershad and Rockmore, 1974; Weber and Fletcher, 1980), interleaved responses (Berninger et al., 2014; Ozdamar et al., 1994; 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); (4) neurophysiological parameters from fitting the responses to different stimulus intensities (Nizami, 2002; Schilling et al., 2019). However, owing to heterogeneity in inter-subject waveform and signal-to-noise-ratio (SNR) introduced by variations in test subject conditions, electrode placement/impedance, as well as acquisition settings, the accurate threshold determination is only possible under a narrow range of experimental settings, which hampers direct comparisons of ABR results across laboratories.

In this study, we proposed an automated approach for real-time ABR threshold determination. Instead of using a prefixed sweep number, level averaging was instructed based on the outcomes of cross-correlation analysis during ongoing sweep acquisition. Near-threshold stimuli feature a sharp increase in the end sweep number upon the detection of time-locked response. We further explored the potential of using this feature for the threshold determination in human subjects, and the algorithm outcomes were validated by the human readouts of the same ABR level representations.

RESULTS

Termination of on-going averaging by cross-correlation analysis

ABRs registered by surface electrodes are embedded in high-level background activities and system noise. Smooth baseline and clear waveform, if present, are obtained after averaging over hundreds of sweeps. The number for averaging, however, depends on the amplitude of the evoked response, which varies between test subjects due to variations in, for instance, skull sizes, electrode impedances, and placement that determine the distance from the peak generator and the vector projection to the electrodes. Within one recording session, these experimental parameters are usually fixed and the baseline activities at different stimulus levels are comparable, allowing signal-to-noise ratio (SNR) quantification. It is expected that strong responses from high-level stimuli quickly reach confident SNR level, whereas weak response evoked by low level stimulus requires more averaging and for subthreshold recordings the SNR cannot be improved by level averaging (Figure 1A).

Based on the above notions, we designed a procedure to test whether the change in sweep number is required for the average response at different stimulus levels to reach a certain SNR level. Such change can produce unbiased confident threshold reading. In detail, at a given stimulus level recorded sweeps are loaded into three memory buffers (Figure 1B, yellow boxes) and cross-correlation coefficients (CCs) are computed between two of three group averages (green boxes, see Figure S1 for examples). If a time-locked response, irrespective of wave latencies and shapes, is present, two group averages overlap without certain time shift, and thereby the obtained CC peak will be found within a small range of signal lag from zero (L, magenta boxes). Three parallel runs (red box) can effectively reject false positives caused by overlapped random noise with similar peak latencies. Next, the correlation analysis is iterated with increasing sweep numbers (the inner loop). When the measured signal lags in all three runs are smaller than L, the response was then considered as time locked by the algorithm and indexed with the end sweep number. Upon the absence of a time-locked response, the upper limit for iteration count (N) is used to terminate nonproductive attempts. Finally, the outer loop is implemented to scan the threshold response with decreasing stimulus levels and the stop command is triggered upon two consecutive levels that reached the iteration limit.

ABR threshold determination in mouse

To test whether the proposed algorithm could determine the ABR threshold reliably, we recorded from ten mice (three wild-type adult C57BL/6 mice of normal hearing, two wild-type adult CBA mice experienced noise exposure and five telomerase knock-out mice with early-onset age-related hearing loss; data are pooled in this study) dual-sweep ABR sets (minimum level averaging allowed by the BigSig software) starting from 90 dB (sound pressure level, SPL) to 0 dB with a step size of 5 dB. The raw data were corrected for baseline fluctuations through a smoothing spline fit before being processed by the algorithm.

Figure 2 presents a walkthrough of the procedure. First, an example of level series is plotted (Figure 2A) and the visually identified threshold denoted at 30 dB SPL with an asterisk. In the algorithm, three level averages (Figure 2B) were used to compute the correlation coefficient (CC). The changes in CC peak amplitude (Figure 2C) and the corresponding signal lag (Figure 2D) upon different stimulus levels were then obtained to further

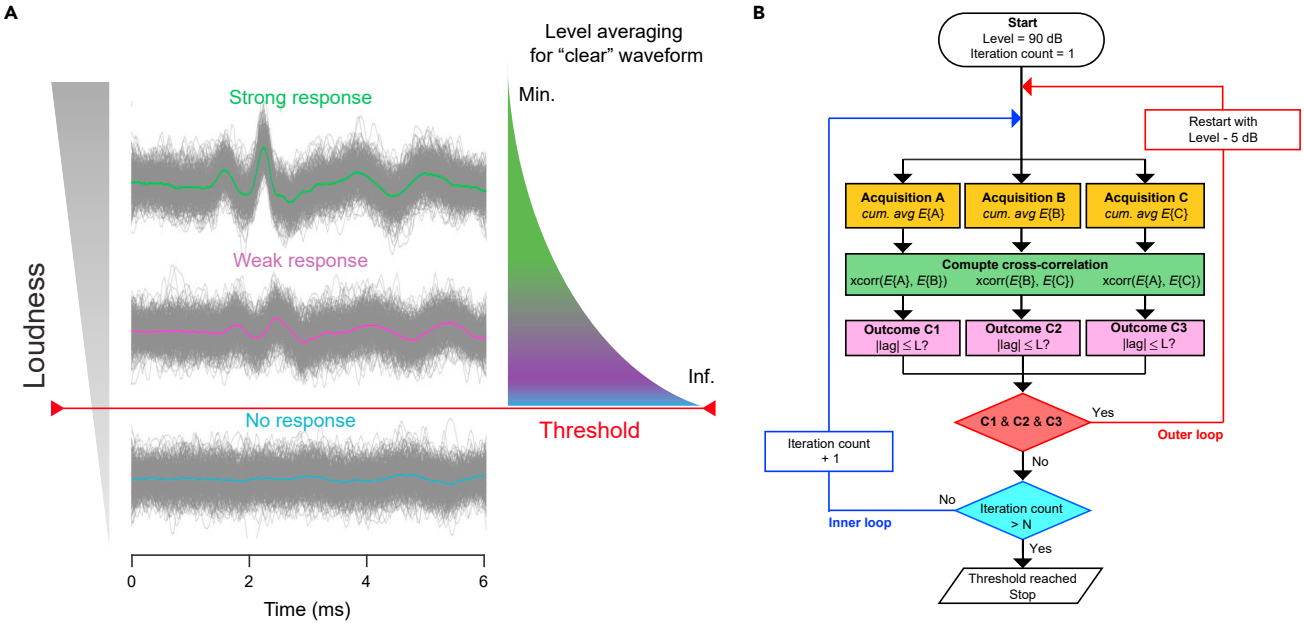


Figure 1. Principle and algorithm design for automated threshold determination

(A) Typical level representation of ABRs. To obtain stable waveform, one needs less level averaging of sweep recordings at strong stimulus (green) than that of weak response (magenta), whereas level averaging cannot improve the SNR of subthreshold recordings containing no response (blue).

(B) Algorithm flowchart. The test starts with the highest stimulus level (e.g., 90 dB SPL for mouse ABR) and a reset iteration count. Recorded sweep batches (e.g., 120 sweeps for mouse ABR) are cumulatively averaged in three data buffers (yellow boxes, *E*{A}, *E*{B} and *E*{C}). Cross-correlation operation (*xcorr*) is carried out in each two of three group averages (green boxes), which yields three CC peaks and their corresponding signal lags. If the absolute lag is less than the allowed value ($L \leq 2$ data points for mouse ABR), the outcome (magenta boxes, C1–C3) returns true, suggesting a positive response. In cases of all three positive outcomes (red box), the procedure descends to lower stimulus level (red line, the outer loop), otherwise the same stimulus level will be repeated with more sweeps (the inner loop, blue line). An iteration upper limit (blue box, e.g., N = 7 for mouse ABR) is implemented to jump out the inner loop upon subthreshold stimulus and flag no response.

explore whether the two derivatives can predict the threshold. Upon reducing the stimulus strength, a monotonic decrease was observed in the CC peak amplitudes, whereas near-threshold levels feature a sudden jump from small absolute lags ($\leq$2 data points or less than $\pm$0.082 ms for suprathreshold levels) to big absolute lags with large variation at subthreshold levels. This result suggested that the signal lag is a better response detector at near-threshold levels, thus justified its use in the algorithm to demarcate the threshold boundary.

Next, to test whether the algorithm could find the ABR threshold in real time, we grouped 60 repeats (120 sweeps) as a batch then increment the group averaging by iterations (multiply by 120 for end sweep number) until positive results, which were scored by triple cross-correlation analysis, were returned by the algorithm. As shown in Figure 2E, the normalized count was small at the suprathreshold levels, whereas it increased rapidly at near-threshold levels and reached its upper limit at subthreshold levels. Therefore, the estimated threshold was defined above the stimulus level at which the iteration upper limit was reached (the highest subthreshold level). Further attempt was made to model the change in iteration counts at near-threshold levels. We acquired an ABR dataset with 1-dB step size (Figure S2A) and fitted with both exponential and sigmoidal functions (Figure S2B). The visually identified threshold was found approximately at the stimulus level, which yields 1.0 and 0.9 on the best-fitted exponential or sigmoidal function. In addition, we found the algorithm outcome was not strongly influenced by either the allowed lag for response detection (Figure S3A) or the applied iteration upper limit (Figure S3B) in the cross-correlation analysis, suggesting that the threshold determination is robust and does not require parameter fine tuning.

ABR threshold determination in human

To test whether the automated approach applied to human ABR, we acquired ABR datasets from eight human participants. Because intermediate single sweeps are not available on the commercial device, alternatively we provided the algorithms with recorded average responses over different sweep numbers (see method details). Example average responses (Figure 3A), as well as group averages (Figure 3B), are shown

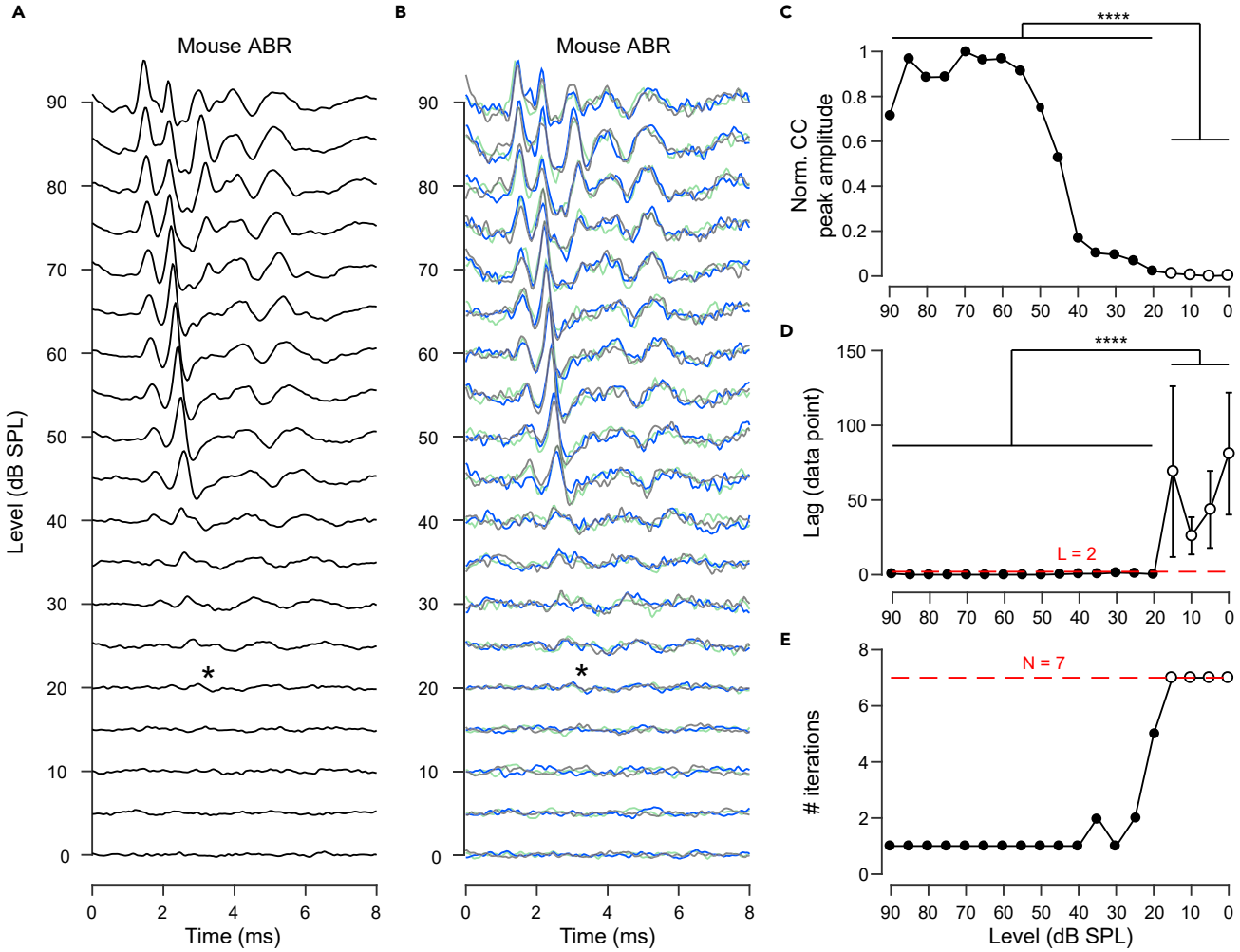


Figure 2. Cross-correlation analysis of mouse ABR

(A) Example level representation (averaging over 840 sweeps) of a mouse. The visually identified threshold level was 20 dB SPL (asterisk).

(B) Group averages generated for cross-correlation analysis. The threshold determined by the algorithm (asterisk) was 20 dB.

(C) Plot of the obtained CC peak amplitude versus the level series.

(D) Plot of signal lag at the CC peak versus the level series. At suprathreshold levels (dots, $\geq$ 20 dB), small signal lags were consistent across three parallel runs (0.33 $\pm$ 0.56 data points, n = 15 levels), whereas at subthreshold levels (cycles, < 30 dB) large absolute values and variability were observed (54.92 $\pm$ 39.36 data points, n = 4 levels, two-sample t test: ****$p < 10^{-5}$).

(E) Result from a test run of the algorithm. The iteration counts were plotted as a function of the level series. Detection of time-locked responses (L $\leq$ 2 data points) required more iterations of averaging at suprathreshold (black dots, 1.40 $\pm$ 1.06, n = 15 levels) than subthreshold stimulus levels (cycles, N = 7). After two consecutive hits at the iteration upper limit (N = 7, dash line), the algorithm flags no response for the applied stimulus level (cycles) and triggered a stop command to avoid nonproductive attempts with weaker stimuli. The algorithm-determined threshold matches visual identification. Data are represented as mean $\pm$ SD.

at tested stimulus levels (60 dB SPL to 0 dB with a step size of 5 dB). The obtained CC peak amplitudes and the signal lags are plotted as a function of the stimulus levels (Figures 3C and 3D). Note that a block of 500 sweeps was added per iteration; the upper limit of iteration count was set to seven (3,500 sweeps) and a larger decision boundary of lag ($\leq$6 data points or less than $\pm$0.3 ms) for positive responses was implemented due to broader waveforms in human ABR evoked by click sound. As shown in Figure 3E, the iteration count increases quickly to reach the upper limit near the visually identified threshold level.

Comparison between expert and algorithm-determined thresholds

In order to evaluate the performance of our method, we recruited five human experts to assess the same ABR level series independently and compared their readouts of the thresholds with the algorithm outcomes

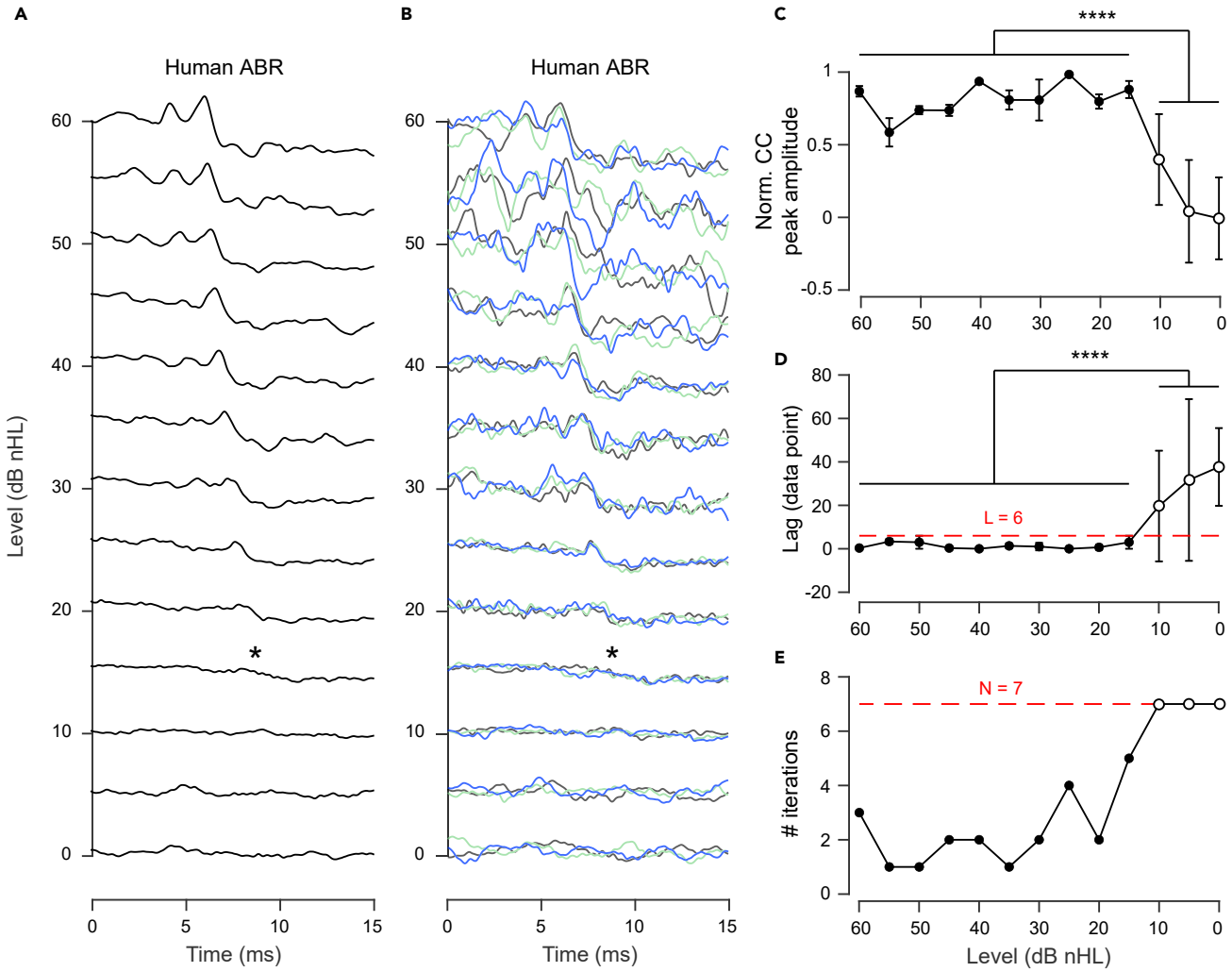


Figure 3. Threshold determination of human ABR by cross-correlation analysis

(A) Example level representations (averaging over 3,500 sweeps) of a human participant with the visually identified threshold at 15 dB (asterisk).

(B) Group averages that used in the algorithm for computing the cross-correlation. Same threshold (15 dB, asterisk) was determined by the algorithm.

(C) Plot of the obtained CC peak amplitudes versus the level series.

(D) Plot of signal lag at the CC peak versus the level series. At suprathreshold levels (dots, $\geq$ 15 dB), small mean value of the signal lags was obtained (1.30 $\pm$ 1.84 data points, n = 10 levels), whereas the subthreshold level (<15 dB, cycles) features a significantly large and variable lag value (23.00 $\pm$ 24.93 data points, n = 3 levels; two-sample t test, ****$p < 10^{-5}$).

(E) Result from an algorithm test run. The iteration counts were plotted as a function of the level series. Detection of the true responses by the cross-correlation analysis (L $\leq$ 6 data points) within the iteration upper limit (N = 7, dash line) flags the suprathreshold levels (dots), which is consistent with those identified visually. Data are represented as mean $\pm$ SD.

(Table 1). Both approaches reached similar conclusions for both mouse and human ABR (Figures 4A and 4B), validating reliability of the algorithm. Moreover, averaging over varied sweep numbers as used in the algorithm does not affect the threshold determination by either machine or human (Figures S4A and S4B). By further quantifying the minimum sweep number required by the algorithm (comparing results from the fixed number for all stimulus levels, 840 for mouse and 3,500 for human ABR), we concluded that the algorithm avoided 66.72 $\pm$ 4.98% and 43.19 $\pm$ 12.48% of the total sweeps, respectively (Figures 4C and 4D).

DISCUSSION

Over decades several statistical approaches have been proposed to automatically detect ABR waveforms. Cross-correlation is one of the most favorite methods, as it can detect temporally stable waveforms with high sensitivity and robustness in a template-free fashion. This is crucial for recognizing ABRs, which

Table 1. Comparison between algorithm outcome and human readout

	Algorithm outcomes		Human readouts					Threshold (by execution judges)	# sweeps
	Threshold	# sweeps	Judge1	Judge2	Judge3	Judge4	Judge5		
Mouse ID									
m01 (C57, wt)	20	4,080	20	20	15	20	25	20.0	14,280
m02 (C57, wt)	25	4,320	20	20	20	25	20	20.0	13,440
m03 (C57, wt)	20	4,560	15	20	15	15	20	16.7	14,280
m04 (terc$^{-/-}$)	30	4,320	25	25	25	25	25	25.0	12,600
m05 (terc$^{-/-}$)	25	3,600	25	20	15	25	25	23.3	13,440
m06 (terc$^{-/-}$)	55	2,880	55	55	55	40	60	55.0	8,400
m07 (terc$^{-/-}$)	25	4,440	55	25	25	25	25	25.0	13,440
m08 (terc$^{-/-}$)	60	2,760	60	60	55	60	55	58.3	7,560
m09 (CBA, noise exposed)	50	2,040	50	50	50	50	55	50.0	6,720
m10 (CBA, noise exposed)	70	2,640	75	70	70	70	75	71.7	5,880
Human participant ID									
h01(73 y.o.)	35	30,000	20	35	25	35	25	28.3	52,500
h02(56 y.o.)	40	54,000	30	35	35	35	25	33.3	73,500
h03(50 y.o.)	15	55,500	10	10	5	15	10	10.0	126,000
h04(58 y.o.)	20	69,000	20	15	20	25	10	18.3	115,500
h05(54 y.o.)	25	69,000	20	25	20	25	30	23.3	105,000
h06(23 y.o.)	5	40,500	5	5	5	5	0	5.0	115,500
h07(23 y.o.)	25	69,000	15	10	15	15	15	15.0	105,000
h08(30 y.o.)	35	45,000	30	30	35	30	20	30.0	84,000

oftentimes present large intersubject variability in terms of waveform shape and latency, from different levels of background noise. However, prior attempts using this approach relied on arbitrary decision boundary for response detection, for instance, minimum required CC (Berninger et al., 2014; Bershad and Rockmore, 1974; Suthakar and Liberman, 2019; Weber and Fletcher, 1980) or maximum allowed latency shift (Galbraith and Brown, 1990; Xu et al., 1995). Preselection of such criteria can be problematic in practice because even for the same test subject, comparable SNR across recording sessions is not guaranteed owing to variabilities in electrode impedance and placement, as well as level averaging settings preferred by individual experimenter. Thus, it is unlikely that a universal response decision boundary can be applied on all ABR sets without introducing detection error, barring its usage in cases like cross-institution collaboration efforts where data pooling is needed. Our approach achieves the ABR threshold determination by monitoring the relative change in the sweep numbers that are required for detectable ABRs in cross-correlation between grouped level averages. This is, in our opinion, fundamentally different from the existing approaches mentioned above, because it does not rely on a static SNR requirement for threshold detection but rather for a dynamic instruction of the level averaging, which in turn can be used to inform the ABR threshold in a baseline-SNR-adaptive manner.

In line with prior study (Xu et al., 1995), the resulted signal lag of CC peak has proven to be a reliable criterion for detecting the time-locked responses with high sensitivity, but in principle, other quantifications like CC peak amplitude (Figure 2C) or single-point F-distribution (data not shown) can also be used in the algorithm. At near-threshold levels, a rapid increase in the end sweep number was observed (Figures 2E and 3E). It is not surprising because in order to reach confidant SNR, small responses require increased averaging for baseline noise reduction, which means that the end sweep numbers quantitatively noted by the algorithm reflect the relative change in the SNR at different recording levels. Besides, this approach does not heavily rely on the selection of detection parameters. First, sub- and supra-threshold level representations feature large differences in the obtained signal lags (Figure 2D), which offers a wide range of decision boundary selection of lag without affecting detection accuracy (Figure S3A). Second, raising the upper limit of the end sweep number only leads to a small shift from the estimated threshold (Figure S3B) due to the exponential increase of averaging at near-threshold levels (Figure S2B), So far, the

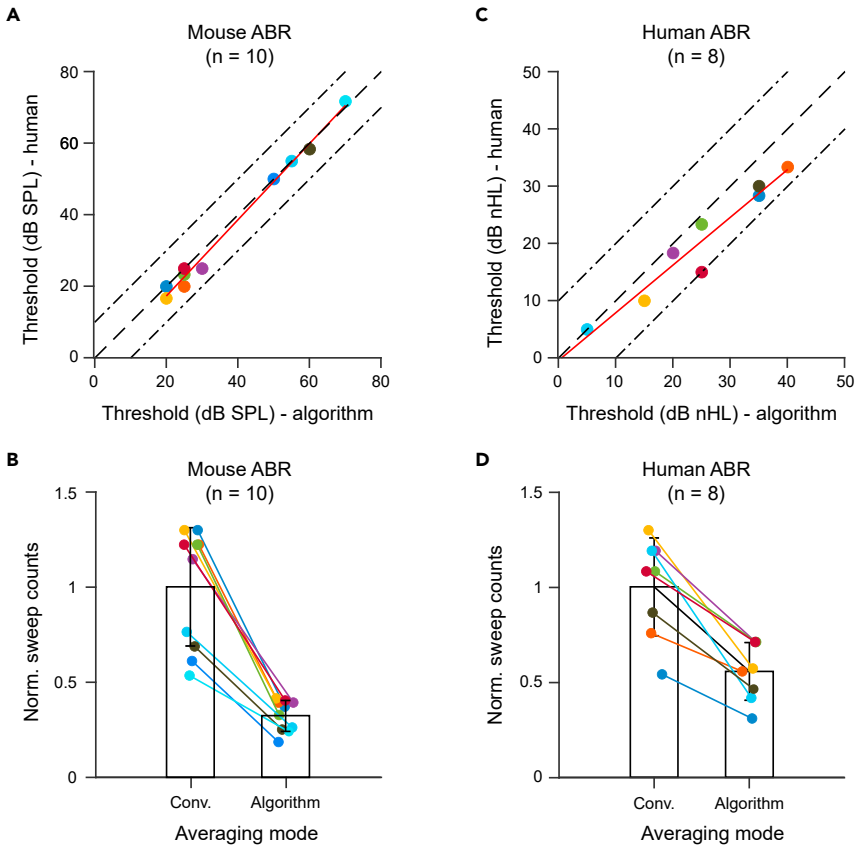


Figure 4. Comparisons between the thresholds determined by machine and human

(A) For mouse datasets, a close match was found between the algorithm-determined thresholds and those averaged from three of five independent human readouts (maximum and mean discrepancies, 5 dB and 1.83 ± 2.45 dB, n = 10 animals). Linear fit: adjust R^2 = 0.99.

(B) Comparison between the total sweeps used in the conventional level averaging (840 sweeps at each level, left bar) and in the algorithm (with varying numbers, right bar). The latter requires 66.72 ± 4.98% fewer sweeps (n = 10 animals). Note that the sweeps were counted from all suprathreshold and two successive subthreshold levels.

(C) As for human datasets, similar thresholds were reported by both the algorithm and human readout (maximum and mean discrepancies, 10 dB and 4.58 ± 3.88 dB, n = 8 human subjects). Linear fit: adjust R^2 = 0.93.

(D) The total number of sweeps used in the conventional level averaging (left bar) versus in the algorithm (right bar). The latter requires 43.19 ± 12.48% fewer sweeps (n = 8 human subjects). Data are represented as mean ± SD.

precision of threshold determination is constrained by the step size of level sampling. Modeling the obtained end sweep number change (Figure S2) and interpolating the level representation may produce higher precision. Further development of this approach is to combine with level sampling strategy including progressively reduced step size (Cebulla and Sturzebecher, 2015) and increased batch size of sweeps per iteration at near-threshold levels so that the model fitting can be improved by more effective data points in the transition. Moreover, in this study the batch size for level averaging was determined empirically (120 sweeps for mouse ABR and 500 sweeps for human ABR). The selection of this size could become more interactive, e.g., based on the algorithm outcome at higher loudness level, as we know weak responses lead to more level averaging. By doing so, cumulative system idling upon trigger of each block recording would be minimized.

In mouse ABR, the proposed method was proven reliable in threshold determination with a mean difference of 1.84 dB (n = 10) and maximum discrepancy of ±5 dB between the algorithm outcomes and human readouts (Figure 4A). Although the detection accuracy is comparable with that of the most up-to-date approach (Suthakar and Liberman, 2019), our algorithm has the advantage of not requiring a calibrated preset criterion and has more robust outcome. Validation with human ABR resulted in only a slight decrease in detection accuracy (mean and maximum difference, 4.58 and ±10 dB; Figure 4C) than that

of mouse ABR. Moreover, the ABR sets with average responses over varying sweep numbers as in the algorithm do not seem to introduce additional difficulty for visual threshold determination (Figures S4A and S4B). This is not surprising because excessive level averaging at suprathreshold levels is not expected to improve the threshold determination. This feature is extremely attractive for two reasons. First, it provides minimal quality control for unambiguous waveform recognition for both human and machine. Such standardized data collection will benefit artificial-intelligence-based approaches by improving the quality of training data (McKearney and MacKinnon, 2019). Second, when to terminate averaging without compromising the quality of the recording is an important decision making during ABR recording (Don and Elberling, 1996; Madsen et al., 2018); this method has the potential to instruct the efficient test by avoiding nonproductive recordings at supra-threshold levels.

Limitations of the study

This is a proof-of-principle study for detecting time-locked responses in cumulatively averaged level representations using cross-correlation analysis. Nevertheless, solo use of this method is not recommended in the clinical routines, as extremely poor recording quality may introduce unexpected errors. Under such circumstances raw recordings should be accessible to the audiologists for quality control and manual debug, which in turn is crucial for further improvement of the approach. Besides, the algorithm is expected to shorten test duration by earlier stop of averaging at supra-threshold levels. The exact saved time in a real setting remains to be quantified on an ABR setup with integration of this algorithm in the control software, which is currently under engineering. Also note that comparison of the test duration with other automated routines that have been already implemented in some laboratories was not performed. As some of them also automatically reduce the sweep number when a high-amplitude response is detected, much less difference in the test duration may be expected.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals
  - Human subjects
- METHOD DETAILS
  - ABR recording
  - Cross-correlation analysis of mouse ABR
  - Cross-correlation analysis of human ABR
  - Threshold determination by human readout
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.103285.

ACKNOWLEDGMENTS

We thank Dr. Hao Wu for the support of this study. We thank Drs. Guangming Chen and Lin Liu for contributing terc$^{-/-}$ mice. We thank Eric Song for comments on the manuscript. This study was supported by Shanghai Huangpu District Industry Support Fund (XK2019011 to Y.H.), the National Science Foundation of China (81770995 to L.S., 81700903 to B.L., and 81800901 to Y.H.) and the Shanghai Key Laboratory of Translational Medicine on Ear and Nose diseases (14DZ2260300).

AUTHOR CONTRIBUTIONS

Y.H. designed the study; Y.H. and L.S. supervised the study; B.L., K.H., X.W., and Z.H. contributed the human ABR datasets; Y.L. and Y.Q. performed the mouse ABR recording; H.W. and Y.H. wrote the algorithm

and analyzed the data with the help of Y.L., Y.Q., H.S., J.Z., and L.S.; Y.H. wrote the manuscript with the help of H.W. and L.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 7, 2021
Revised: August 25, 2021
Accepted: October 13, 2021
Published: November 19, 2021

REFERENCES

Acır, N., Özdamar, Ö., and Güzeliş, C. (2006). Automatic classification of auditory brainstem responses using SVM-based feature selection algorithm for threshold detection. Eng. App Artif. Intell. *19*, 209–218.

Alpsan, D., and Ozdamar, O. (1991). Brain-Stem Auditory Evoked-Potential Classification by Backpropagation Networks, *1–3* (1991 IEEE International Joint Conference on Neural Networks), pp. 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. Theor. *20*, 112–113.

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. J. Acoust. Soc. Am. *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. *58*, 224–230.

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.

Nizami, L. (2002). Estimating auditory neuronal dynamic range using a fitted function. Hear. Res. *167*, 13–27.

Ozdamar, O., Delgado, R.E., Eilers, R.E., and Urbano, R.C. (1994). Automated electrophysiologic hearing testing using a threshold-seeking algorithm. J. Am. Acad. Audiol. *5*, 77–88.

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, Second Edition (Thieme).

Sanchez, R., Riquenes, A., and Perez-Abalo, M. (1995). Automatic detection of auditory brainstem responses using feature vectors. Int. J. Biomed. Comput. *39*, 287–297.

Schilling, A., Gerum, R., Krauss, P., Metzner, C., Tziridis, K., and Schulze, H. (2019). Objective estimation of sensory thresholds based on neurophysiological parameters. Front. Neurosci. *13*, 481.

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.

Sung, L.-Y., Chang, W.-F., Zhang, Q., Liu, C.-C., Liou, J.-Y., Chang, C.-C., Ou-Yang, H., Guo, R., Fu, H., et al. (2014). Telomere elongation and naive pluripotent stem cells achieved from telomerase haplo-insufficient

cells by somatic cell nuclear transfer. Cell Rep *9*, 1603–1609.

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.

Vannier, E., Adam, O., and Motsch, J.F. (2002). Objective detection of brainstem auditory evoked potentials with a priori information from higher presentation levels. Artif. Intell. Med. *25*, 283–301.

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.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw ABR recordings	This paper	Mendeley Data (https://doi.org/10.17632/4yb9772dff.1)
Experimental models: Organisms/strains		
Mouse: CBA/Ca	Sino-British SIPPR/BK Lab.Animal Ltd (Shanghai, China)	CBA/Ca/Bkl
Mouse: C57BL/6	Sino-British SIPPR/BK Lab.Animal Ltd (Shanghai, China)	C57BL/6/Bkl
Mouse: terc-deficient (G2 and G3)	Sung et al., 2014, Gifts from Prof. Liu Lin (Nankai University, China)	$Terc^{-/-}$
Software and algorithms		
Automatic ABR Threshold Determination	This paper	https://github.com/SHIPM-HuaLab/automatic-ABR-threshold-detection
MATLAB including Curve fitting and Statistics Toolbox	MathWorks, Inc	release 2017b

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Yunfeng Hua (yunfeng.hua@shsmu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All recorded ABR datasets supporting the current study have been deposited at Mendeley Data and are publicly available as of the date of publication. The DOI is listed in the key resources table.
- The codes were written in MATLAB scripts and shared on GitHub (https://github.com/SHIPM-HuaLab/automatic-ABR-threshold-detection).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

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).

Animals

C57BL/6 and CBA mice were purchased from Sino-British SIPPR/BK Lab Animal Ltd. (Shanghai, China). The telomerase-knock-out mice ($terc^{-/-}$) were gifts from Prof. Lin Liu (Nankai University, China) and bred in house.

Human subjects

Eight human participants of both genders (range from 23 to 73 years old, Table 1) were recruited by Hearing and Speech Center, Shanghai Ninth People's Hospital and consent forms were signed before the experiment. This study was conducted at the Ear Institute and the Hospital Hearing and Speech Center.

METHOD DETAILS

ABR recording

Mouse ABRs were recorded via a TDT RZ6/BioSigRZ system (Tucker-Davis Tech. Inc., US) in a sound-proof chamber as previously described (Lin et al., 2019). In brief, 7-week-old animals 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 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). Recorded signals were collected by Medusa pre-amplifier then send to RZ6 station via fiber optics and amplified 5000×. Sound stimuli were generated by SigGen RP (Tucker-Davis Tech. Inc., US). 3-ms tone pips (with 1 ms rise and fall) at 16 kHz were delivered free-field via an MF1 speaker (Tucker-Davis Tech. Inc., US) positioned in the front of the animal 10 cm away from the vertex. Offline calibrations were performed by using software BioSigRZ with the microphone placed 10 cm (~4 inches) away from the speaker. The system generated correction file were then used to produce the calibration file. Acoustic stimuli with alternating polarity were presented 20 stimuli per second and the evoked potentials were collected at 24 kHz sampling rate. Each recording was averaged from two sweeps (minimum allowed by the BioSig software, No. Averages = 2 in the BioSig Acquisition Channel Setup) and repeated manually at each loudness level. Artifact rejection level was set at 35% (mean rejection voltage 20.5 μV). Sound level series started from 90 to 0 dB sound pressure level (SPL) 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 (Figure S2).

Human ABRs were recorded by a commercial ABR device (Intelligent Hearing Systems, US) with Smart EP software from eight volunteers aged 23-73 years without the knowledge of their medical conditions. 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/sec 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 20 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 of mouse ABR

Sweeps were loaded in three memory buffers. Cross-correlation (MATLAB Central File Exchange Function *xcorr*, MathWorks, US) was computed from two out of three group averages. This resulted in correlation coefficients as a function of signal lag between two group averages. As the ABRs are time-locked to the stimulus onset, they can be characterized by cross-correlation parameters including small latency delay on the lag-time axis at the point of the maximal coefficient (Xu et al., 1995). In this study, maximum allowed lag (*L*) for a true ABR signal was set to ± 2 data points from time zero (equivalent to ± 0.082 ms or 1% of the analyzed temporal window). Three such tests were implemented in parallel in order to minimize false positives caused by coincidently overlapped background activities through rejecting inconsistent lag values. Moreover, we monitored the change of correlation coefficient peak amplitude as an independent measure of SNR (Figure 2C).

The test started with the loudest stimulus level (90 dB SPL) and was reduced by a step size of 5 dB. At each given stimulus level, the sweep number used to produce the group average was increased by iterations and the end sweep number was noted upon detected response determined by the cross-correlation analysis. For each iteration, sweeps were added in batches (120 sweeps) and an upper limit was set (840 sweeps) to avoid nonproductive attempts at subthreshold levels. Upon detected responses, the iteration was then switched to lowered stimulus level, until the upper limit was reached for two consecutive nonproductive trails. The estimated threshold was the lowest applied stimulus level at which the upper iteration limit was not yet reached.

The transition of the end sweep numbers between supra- and subthreshold levels was modeled (Figure S2). Both sigmoidal (1) and exponential functions (2) were employed to fit the relationship between the normalized iteration count C' (equivalent to the sweep number) and the stimulus level S using a nonlinear least

square method in MATLAB (Curve Fitting Toolbox, MathWorks, US). In the functions, $\alpha_1 = 0.6$ and $\alpha_2 = 0.2$ were fixed for calibrated lag criterion ($L = 2$), whereas β_1 and β_2 were obtained by fitting.

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

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

Cross-correlation analysis of human ABR

For human ABR, group average responses were recorded sequentially and used directly as inputs of the algorithm. Group averages over 500, 1000 and 2000 sweeps were recorded, whereas averages over 1500, 2500, 3000 and 3500 sweeps could be obtained by linear combination (3) where $E\{m\}$, $E\{n\}$ and $E\{m + n\}$ denote the time averages over m, n and $m + n$ sweeps, respectively.

$$E\{m + n\} = \frac{m \cdot E\{m\} + n \cdot E\{n\}}{m + n} \quad \text{(Equation 3)}$$

The maximum allowed lag for a true response was set to ≤ 6 data points from time zero (equivalent to $\pm$ 0.3 ms or 2% of the analyzed temporal window). The iteration upper limit was seven, corresponding to 3500 sweeps. The estimated threshold was the lowest stimulus level with a detectable response.

Threshold determination by human readout

To estimate the ground-truth thresholds of the recorded mouse and human ABRs, average responses of all level series were provided to five clinicians to report the visually identified thresholds independently. The test subject identities were blinded to the judges. Either the fixed sweep number (the conventional averaging) or the varying numbers used in the algorithm (the algorithm averaging) was applied to compute the level averaging. The thresholds were determined by three out of five execution judges (with the highest and the lowest value excluded). The readouts were used to evaluate the accuracy of the algorithm outcomes (Table 1).

QUANTIFICATION AND STATISTICAL ANALYSIS

All data analyses including statistical tests were performed using self-written scripts and built-in functions in MATLAB (release 2017b) including the Curve Fitting and the Statistics Toolbox (MathWorks, Inc.). The group comparisons were done using two-unpaired t-tests (ttest2): Figures 2C, 2D, 3C, and 3D as well as paired t-tests (ttest): Figures 4B and 4D. The significance level of statistical tests was denoted as n.s. for p-value > 0.05, * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$ and **** for $p < 0.0001$. More statistical details of experiments can be found in the figure legends.