Mao, D., H. Innes-Brown, M. A. Petoe, Y. T. Wong, and C. M. McKay. 2018. Cortical auditory evoked potential time-frequency growth functions for fully objective hearing threshold estimation. *Hearing Research*. **370**: 74-83.

doi: 10.1016/j.heares.2018.09.006

Accepted Manuscript

Cortical auditory evoked potential time-frequency growth functions for fully objective hearing threshold estimation

Darren Mao, Hamish Innes-Brown, Matthew A. Petoe, Yan T. Wong, Colette M. McKay

PII: S0378-5955(18)30252-1

DOI: 10.1016/j.heares.2018.09.006

Reference: HEARES 7619

To appear in: *Hearing Research*

Received Date: 8 June 2018

Revised Date: 24 August 2018

Accepted Date: 26 September 2018

Please cite this article as: Mao, D., Innes-Brown, H., Petoe, M.A., Wong, Y.T., McKay, C.M., Cortical auditory evoked potential time-frequency growth functions for fully objective hearing threshold estimation, *Hearing Research* (2018), doi: https://doi.org/10.1016/j.heares.2018.09.006.

Cortical auditory evoked potential time-frequency growth functions for fully objective hearing threshold estimation

Darren Mao[a,b], Hamish Innes-Brown[b,c], Matthew A. Petoe[b,c], Yan T. Wong[a,d], Colette M. McKay[b,c]

[a] Department of Biomedical Engineering, University of Melbourne, Parkville VIC 3010, Australia

[b] The Bionics Institute, 384-388 Albert St, East Melbourne VIC 3002, Australia

[c] Department of Medical Bionics, University of Melbourne, Parkville, VIC 3010, Australia

[d] Department of Physiology, Department of Electrical and Computer Systems Engineering, and the Biomedicine Discovery Institute, Monash University, Clayton VIC 3168, Australia

Corresponding author: Darren Mao, darren.mao92@gmail.com, Mailing address: The Bionics Institute, 384-488 Albert St, East Melbourne, VIC 3002, Australia

Abstract

Cortical auditory evoked potential (CAEPs) thresholds have been shown to correlate well with behaviourally determined hearing thresholds. Growth functions of CAEPs show promise as an alternative to single level detection for objective hearing threshold estimation; however, the accuracy and clinical relevance of this method is not well examined.

In this study, we used temporal and spectral CAEP features to generate feature growth functions. Spectral features may be more robust than traditional peak-picking methods where CAEP morphology is variable, such as in children or hearing device users. Behavioural hearing thresholds were obtained and CAEPs were recorded in response to a 1 kHz puretone from twenty adults with no hearing loss. Four features, peak-to-peak amplitude, root-mean-square, peak spectral power and peak phase-locking value (PLV) were extracted from the CAEPs. Functions relating each feature with stimulus level were used to calculate objective hearing threshold estimates. We assessed the performance of each feature by calculating the difference between the objective estimate and the behaviourally-determined threshold.

We compared the accuracy of the estimates using each feature and found that the peak PLV feature performed best, with a mean threshold error of 2.7 dB and standard deviation of 5.9 dB across subjects from behavioural threshold. We also examined the relation between recording time, data quality and threshold estimate errors, and found that on average for a single threshold, 12.7 minutes of recording was needed for a 95% confidence that the threshold estimate was within 20 dB of the behavioural threshold, using the peak-to-peak amplitude feature, while 14 minutes is needed for the peak PLV feature. These results show that the PLV of CAEPs can be used to find a clinically relevant hearing threshold estimate. Its potential stability in differing morphology may be an advantage in testing infants or cochlear implant users.

Keywords: Hearing threshold; objective audiometery; electroencephalography; spectral analysis; phase-locking value; growth functions;

1 Introduction

Accurate estimations of hearing thresholds are important for the optimal adjustment of hearing aids and cochlear implants to suit each individual patient's hearing characteristics. Behavioural threshold estimates are the current gold standard and in most cases are easy to administer, but the accuracy of the result can be impaired by an inability to respond (as in the very young), lack of co-operation, or confusion on the part of the respondee. For these reasons, objective measures of hearing threshold are desirable. In this work, we use the cortical auditory evoked potential recorded from a single active channel as an objective measure used to estimate hearing thresholds, and compare the performance of different features of the cortical auditory evoked potential in threshold estimation using a growth function fitting method.

Objective hearing threshold estimates may be derived from auditory-evoked electroencephalography (EEG) measures such as the compound action potential (CAP) (Goldstein Jr et al., 1958), the auditory brainstem response (ABR) (Jewett et al., 1970), the auditory steady state response (ASSR) (Galambos et al., 1981), or the cortical auditory evoked potential (CAEP) (Davis, 1965). Electrophysiological recordings have the advantage that they do not require the participant to give feedback. When assessing the validity of these methods against behavioural thresholds, the CAEP, which is more likely than subcortical potentials to correspond to perceptionof the stimulus, appears to correlate with behavioral thresholds similarly or better than the ABR (Hoth, 1993; Pratt et al., 1978; Van Dun et al., 2015) and the ASSR (Tomlin et al., 2006; Yeung et al., 2007). The CAP (generated in the auditory nerve) is only utilised for cochlear implant users, where it is better known as the electrically-evoked CAP (ECAP), and is less correlated with behavioral hearing thresholds than either CAEP or ABR (Brown et al., 2000; McKay et al., 2017; McKay et al., 2013; Miller et al., 2008). In addition, the CAEP have shown promise in situations where ABR thresholds cannot be obtained (e.g. those with auditory neuropathy (Pearce et al., 2007; Starr et al., 1996)), or where ABR thresholds are not well correlated with behavioural measures (e.g. cochlear implant users, especially at higher stimulation rates (Abbas et al., 1991; Brown et al., 1999)).

Methods of hearing threshold estimation based on physiological responses use a common framework regardless of whether the estimation technique is by visual inspection or is automated by an algorithm. Estimation methods can generally be split into three steps: First,

multiple recordings are acquired in response to each of several sound intensity levels. Second, specific features are extracted from the EEG recording. Examples of these features include the frequently used N1-P2 potential difference (Picton et al., 1976), the phase coherence between individual response epochs of the same stimulus(Picton et al., 1987; Tallon-Baudry et al., 1996), the spectral power of the ASSR (Yeung et al., 2007), or the global field power across a full-scalp electrode montage (Visram et al., 2015). Third and finally, a set of decision rules are applied to find the threshold by either statistically testing for the presence of the feature at each single intensity level (Golding et al., 2009), or by fitting a growth function to extrapolate beyond the intensity levels that were tested to find threshold (Visram et al., 2015). Single intensity level testing requires using an adaptive procedure to find the lowest intensity level that elicits a response (Van Dun et al., 2015), however there is no literature on the optimal intensity levels to present for growth function fitting.

Previous studies have explored many combinations of the feature extraction and threshold estimation steps described above. By far the most commonly used feature in clinical use for threshold estimation is visual inspection of the N1-P2 morphology. These studies have consistently found a 10 dB standard deviation in difference between estimated threshold and behavioural threshold for normal hearing and hearing impaired adults using acoustic stimuli, although all have isolated exceptions where the estimate disagreed with behavioural estimates by over 15 dB or where no response could be recorded (Coles et al., 1984; Hoth, 1993; Rickards, 1996; Yeung et al., 2007). In objective detection, the amplitude of the time-domain response remains the most straightforward and most used feature (Hoth, 1993; Ross et al., 1999; Van Dun et al., 2015). However, other features such as single epoch phase (Ross et al., 1999), signal to noise ratio, reproducibility, and spectral power (Hoth, 1993) have been combined with amplitude measures, achieving results similar to those of visual detection by experienced examiners.

To further advance the robustness of objective threshold estimation, features based on the time-frequency content of the CAEP are of particular interest, as there has been no direct comparison of their performance against time-domain analyses. In addition, although power change from baseline (Hoth, 1993; Makeig, 1993) and phase measures (Ross et al., 1999) have been previously used in CAEP based analyses, the relationship of these features with stimulus intensity is largely unexplored. For example, Ross et al. (1999) reported an

increased statistical significance of phase locking compared to baseline phase locking when stimulus intensity was increased, but did not quantify the degree of phase locking.

Threshold estimation can benefit from utilising the relationship between features and stimulus intensity. This relationship is found by fitting growth functions to observed features of EEG recordings in response to sound stimuli of different intensities (see Figure 1). In comparison to methods that use single level statistical testing (e.g. Van Dun et al. (2015)) the growth function technique has the advantage that extrapolation of the growth function can give finer resolution of threshold estimates, and each tested level does not have to reach statistical significance. Visram et al. (2015) have used a growth function method successfully in cochlear implant users; however, 64 EEG channels were needed to estimate global field power. Use of fewer or single recording channel pairs may allow their direct incorporation onto hearing aids or cochlear implants, and hence audiometry in aided hearing without additional recording equipment. Growth functions of CAEPs are otherwise not widely reported due to the the time consuming nature of acquiring the necessary responses (Abbas et al., 2015).

In the present study, we explored time-frequency features of the EEG acquired from single EEG channels and their associated growth with stimulus intensity in order to estimate hearing thresholds in 20 normally hearing adults. We compared two time-domain and two time-frequency domain features: in the time domain, the peak-to-peak amplitude (N1-P2) and the total power in the response (root-mean-square value); and in the frequency domain, the post-stimulus peak spectral power and peak phase-locking value. We objectively estimated hearing thresholds with each of these features. We compared the performance of these four features in threshold estimate accuracy and established the relation between threshold estimate accuracy and recording time. We hypothesised that growth function fits based on time-frequency features would provide more accurate estimates of hearing threshold than growth function fits based on time domain features.

2 Methods

<<Insert Figure 1 about here, grayscale>>

2.1 Subjects

20 subjects, with ages ranging from 21 to 37 years (mean 25 years), participated in this study. This study was conducted under the approval of the Human Research Ethics Committee of the Royal Victorian Eye and Ear Hospital, and each participant gave written informed consent. An audiogram was performed for both ears at octave frequencies between 250 and 8000 Hz with an audiometer and all subjects had thresholds less than 20 dB HL. The ear with the lower pure tone average threshold between 250 and 4000 Hz was selected for subsequent testing, resulting in 12 right ears and eight left ears.

2.2 Sound stimuli

The test stimulus was a 1000 Hz pure tone of 50ms duration and had 4-ms linear rising and falling edges, and phases were consistent for each stimulus. The stimuli were presented through Matlab (version 2015a, MathWorks, MA, USA), an RME Fireface 400 sound card (RME Audio, Germany) and Etymotic Research ER2 insert earphones (Etymotic Research, IL, USA). Stimulus calibration was performed with the GRAS 43AC Ear Simulator Kit and the Norsonic Nor140 Sound Analyser by taking a reading at a 1/3 octave band around a 1 kHz centre frequency. The dB SPL of a full-scale pure tone was recorded and used to scale stimuli to the required dB SPL levels in Matlab.

2.3 Behavioural threshold estimate

Behavioural thresholds for the experimental stimulus were found with an adaptive three-interval forced-choice task. One of three 200-ms intervals, chosen randomly, contained the sound stimulus, where a 150ms silent period was appended to the end of the 50ms tone, while the other two contained silence. Subjects were asked to select the interval with sound or to guess if unsure. Stimuli began at 50 dB SPL, and the level was increased after one incorrect response or decreased after two consecutive correct responses. There were ten turning points, the first two being in steps of 10 dB and 5 dB, followed by eight 2 dB steps. The behavioural threshold was taken as the average of the last six turning points and is referred to as 0 dB sensation level (dB SL) in this paper.

2.4 EEG data

Nine stimulus intensities relative to 0 dB SL were presented: -20, -5, 0, 2, 5, 10, 20, 40 and 60 dB SL. The -5, 0, and 2 dB SL levels were excluded from growth function analyses as they were deemed to not contribute to the growth function fits; instead they were used in

single level statistical testing. Additionally, the -20 dB SL level was only used to calculate baseline levels and was not used in the growth function fitting. To minimize possible effects of training or fatigue, stimuli were presented in a random order with an inter-stimulus interval (ISI) randomly jittered between 1.35 and 1.65 seconds. EEG data were acquired with a BioSemi ActiveTwo EEG system with data acquired from Cz (in the international 10-20 configuration) and left and right mastoids, with the Common Mode Sense (CMS) and Driver Right Leg (DRL) placed on either side of the midpoint between Pz and POz. Data were sampled at 2048 Hz, with a fifth order sinc response anti-aliasing filter with -3dB cut-off at 400 Hz. Data were acquired at each stimulus intensity until the residual noise level (Elberling et al., 1984) either reached 0.4 μV^2, or 0.6 μV^2 if the Hotelling T^2 statistic (Golding et al., 2009; Hotelling, 1992) concomitantly returned $p < 10^{-6}$. This ensured that the data quality across all subjects were at a consistent noise level. Recording was performed in an electrically isolated acoustic chamber. The BioSemi A/D box was connected to computer equipment outside the room via an optic cable. Subjects were given a silent, captioned film to watch during the experiment, and were instructed to stay awake and as still as possible. The noise from the stimulus generation system was measured to be 6.0 dB SPL.

2.5 Data post-processing

Data were re-referenced to the average of the two mastoid channels, and zero-phase filtered between 1 and 45 Hz with separate low and high-pass elliptical filters designed to have a passband ripple of <1% and stopband attenuation of >40dB. Data were then down-sampled to 256 Hz and epoched in a window -800 ms to +1200 ms relative to stimulus onset. Finally, any epochs with amplitudes that exceeded ±100µV were rejected as artefact contaminated epochs. After these analyses, the dataset included data from 20 subjects and nine stimulus levels. We will henceforth refer to each set of epochs from one subject and one stimulation level as a “block” of epochs. The number of epochs ranged from 130 to 451 per block (mean: 265, standard deviation: 86).

2.6 Feature extraction

Four response features of the EEG were considered in this study; two in the time domain, and two in the time-frequency domain. The two time domain features were both calculated on the epoch block mean (average of all epochs across each time point in a block): peak-to-peak amplitude values were calculated as the difference between the minimum and maximum voltage potential in a 50 to 500 ms post-stimulus window, and the root mean square (RMS)

value of the block mean was calculated on the same 50 to 500 ms post-stimulus window (Figure 2A). The two time-frequency domain features, peak spectral power and peak phase locking value (PLV), were calculated by first taking the short-time Fourier transform using Matlab's spectrogram function on each epoch. A 400 ms Hamming window with a time-step of 20 ms was used. Power spectrograms were calculated by taking ten times the log of each epoch's time-frequency magnitude, averaging across all epochs, and then normalizing the spectrogram by subtracting the pre-stimulus power calculated at 300 ms before stimulus onset. The peak spectral power feature was defined as the peak value in the power spectrogram between 1-20 Hz and between 50 and 500 ms, where a post-stimulus time is the location of the centre of the Hamming window (Figure 2B). The phase-locking spectrograms were calculated by taking each of the N epoch's time-frequency phase (θ) and applying the following formula to calculate the phase-locking value (PLV) at each time (t) and frequency (f) point to generate the phase-locking spectrogram (Mardia, 2014):

$$\mathrm{PLV(t,f)}=\frac{1}{\mathrm{N}}\sqrt{\left[\sum_{\mathrm{i=1}}^{\mathrm{N}}\cos\left(\theta_{\mathrm{i}}(t,f)\right)\right]^{2}+\left[\sum_{\mathrm{i=1}}^{\mathrm{N}}\sin\left(\theta_{\mathrm{i}}(t,f)\right)\right]^{2}}$$

The peak PLV feature was the peak value in the PLV spectrogram between 1-20 Hz and between 50 and 500 ms (Figure 2C).

2.7 Growth function fitting and threshold estimation

Each time a feature was extracted from a block of epochs in response to one stimulus intensity, bootstrapping was performed to establish a confidence interval for each calculated feature. The feature's distribution was estimated by randomly selecting a subset of epochs from a block with replacement for 1000 iterations (Efron et al., 1986). The standard deviation of the feature distribution was calculated (Efron et al., 1986), which gives a measure of the size of the confidence interval. We will refer to this as the 'feature noise level'. Thus, a larger feature noise level corresponded to a bigger uncertainty of the calculated feature. If a block did not have enough epochs to reach a specified feature noise level, all epochs within the block were used.

Single level testing was performed with the bootstrap feature distributions as above, and for each subject. The bootstrap distributions for each level were compared to the baseline level. The baseline level was the calculated feature using the -20 dB SL response block. Responses

at each level were deemed detected when the distributions of the tested level and baseline level overlapped for less than 50 bootstrap iterations (thus calculated $p<0.05$). The threshold estimate was defined as the lowest stimulus level at which two consecutive stimulus levels were detected as responses.

Growth function estimation was performed by finding the best linear fit that described the growth of the function with stimulus level, and finding the stimulus intensity level at which this function intercected with the feature noise floor. The median of the bootstrap feature distribution was used as the estimate of the feature at each stimulus intensity level. Growth functions were generated by performing a linear least-squares curve fit to each of the four response features using the 5, 10, 20, 40 and 60 dB SL intensity levels. Then, the feature value at -20 dB SL (an intensity level which would be very unlikly to be perceived) was used as the baseline value of this particular feature, and the dB SL level that correponded to this value on the linear fit growth function was taken as the threshold estimate (see Figure 4A for diagramatic representation). Thus, the threshold estimates described in this paper are the difference between the estimated threshold and the behavioural threshold in decibels. Threshold estimates were deemed invalid, and thus removed from further analyses, if one of three criteria were met: the estimated threshold was outside the range of ±100 dB SL, the software was unable to find a valid fit, or the slope of the fit was less than 0 (negative amplitude growth). We also compared the linear curve fit to an exponential curve fit as described by Visram et al. (2015) and (Ross et al., 1999), fitting to the equation:

$$f(x) = a(1 - e^{-\frac{x-b}{c}})$$

where *f* represents the feature, and *x* represents the stimulus intensity in dB SL. *a* represents the asymptote and thus the upper limit for the features, *b* represents the x-axis shift and *c* represents the speed of feature growth. When fitting a curve for the peak PLV feature, *a* is restricted to a value between 0 and 1.

Here, the threshold estimates were deemed invalid if the estimated threshold was outside the range of ±100 dB SL, the adjusted r^2 value of the curve fit was less than 0, or the term *a* or *c* of the curve fit was less than 0. The arbitrary range of ±100 dB SL was chosen to eliminate any outliers due to unusual fits (such as one with very shallow slope) affecting the overall analysis.

The performance of the threshold estimate method was quantified by the estimate's standard deviation across subjects. This method of defining performance was chosen, rather than the mean difference between behavioral and objective thresholds, because any mean difference can be subtracted from the objective threshold as a corrective factor (Hoth, 1993; Ross et al., 1999). For example, if a method estimated thresholds to be 5 ± 10 dB SL across subjects, then we could subtract 5 dB from all estimates and the 10 dB standard deviation quantifies the error across subjects in the performance of this method.

3 Results

3.1 Behavioural data

Hearing thresholds at 1 kHz determined from the audiogram across 20 subjects were 13.0 ± 5.1 dB SPL (mean ± 1 std). Behavioural thresholds obtained by the three-interval forced-choice task with 1 kHz tone pips were found to be 14 ± 4.4 dB SPL (mean ± 1 std). A paired t-test revealed no significant difference between thresholds found with the two methods (t = 1.57, p = .133), and there was a moderate and significant correlation between the methods (r = .76, p < .001). The three-interval forced-choice behavioural thresholds were taken as 0 dB SL and thus were used to compare with the objective estimates in further analyses.

3.2 Significant power change and phase locking to sound stimuli were present in all subjects

An example of epoch-averaged time-domain data from one subject at 60 dB SL is shown in Figure 2A, which illustrates how the two time-domain features, peak-to-peak amplitude and RMS were calculated. The same response represented in the time-frequency domain (spectral power and PLV) is shown in panels B and C, respectively. The peak values in each (indicated by a black dot) are the time-frequency domain features that were extracted for growth function fitting.

The regions enclosed by the black outline in Figures 2B and 2C are the regions within which the power and phase-locking features were statistically greater than baseline for that subject, with the baseline set at the -300 ms timepoint (permutation test, 1000 iterations, normalized z score>3 for outline). The bottom panels (D-F) show the grand average time- and frequency-domain features, as well as the peak power and phase-locking locations for each of the twenty

subjects, denoted by black dots. It can be seen that the peak PLV time-frequency location was less variable across subjects than for peak power.

<<Insert figure 2 about here, in color>>

Each feature was extracted at every stimulus intensity for all subjects. The mean growth of the features with respect to stimulus intensity is shown in Figure 3. For all features, the values were observed to increase with increasing stimulus intensity. Furthermore, each feature in response to several near-threshold intensities were similar in value to the baseline feature level, where the baseline was indicated by the feature value at a stimulus intensity of -20 dB SL. Linear growth of each feature was observed at higher supra-threshold levels that were not obscured by the noise floor. This result prompted the selection of a linear curve fit for threshold estimation.

<<Insert figure 3 about here, in grayscale>>

3.3 Linear curve fits estimate thresholds accurately

Linear growth functions were fitted to each subject's data individually using the four features described in the Methods section. In this section, all epochs that were acquired were used in the analysis. In Figure 4, panel A shows an example of the growth function for one subject using the peak PLV feature. A linear relationship between the PLV and the stimulus intensity level can be observed. Panel B shows fitted linear growth functions for all subjects and features, and that peak PLV and peak-to-peak amplitude had the smallest variance across subjects in their threshold estimates, while peak spectral power was the worst performing measure with three invalid fits and one that was outside the displayed range.

The quality of the threshold estimation is shown in rows 1 and 2 of Table 1. Levene's test for unequal variance was performed in Minitab and showed a significant difference in the variance of the threshold estimates using the four features ($W = 5.08$, $p = 0.003$). Post-hoc analysis by performing a two-sample F-test on each pairing found that the variance of the peak power threshold estimate was significantly higher than the variance of all the other features (Bonferroni corrected, $\alpha = .05$: versus peak-to-peak amplitude: $F_{(16,19)} = 11.4$, $p < 0.001$; versus RMS: $F_{(16,19)} = 7.22$, $p < 0.001$ versus peak phase-locking value: $F_{(16,19)} = 12.1$, $p < 0.001$). Thus, the peak power feature has poorer performance compared to the threshold estimates generated using the other three features. In addition, a one-way ANOVA was

performed on the mean threshold estimates and a significant difference was found (Welch's test, $F_{(3,37)} = 3.43$, $p = 0.027$). Post-hoc analysis revealed a significantly lower mean threshold estimate when using the peak PLV compared to the peak power feature (Games-Howell test, adjusted $p = 0.029$). No other comparisons were significant.

<<Insert Figure 4 about here, in color>>

The mean goodness of fit across subjects for each feature, in adjusted r^2 value, is indicated in row 2 of Table 1. The goodness of fit values were high (above 0.91 for all features except peak power), providing evidence that a linear function is suitable to model the relationship between the intensities tested and each of the features except for peak power. To further validate our method, we compared results in three other ways. Firstly, we compared the linear curve fit with an exponential curve fit (as performed by Visram et al. (2015)), and the goodness of fits are shown in Table 1. The threshold estimates and the goodness of fits are not significantly different for all four features (Ranksum test, Bonferroni corrected at $\alpha = 0.05$). However, by visual inspection, we observed that the exponential fits were mostly linear in the region near the hearing threshold and continued linearly for the higher stimulus intensities tested. The values for the upper asymptote of the fitted exponential for peak-to-peak amplitude, for example, were all above 90 µV except for in one subject, an unlikely value for an auditory response. It was likely that the data did not sufficiently define the exponential fit, as the stimulus intensities did not reach a high enough level to for a plateau in the features to become apparent. In the special case of the peak PLV growth function where we limited the asymptote to a maximum of 1, the exponential growth functions could fit the sampled stimulus intensity points used in this study; and the growth rate was also slow enough to fit the mostly linear growth as seen in Figure 3 panel D. The implications of using different growth function shapes are included in the Discussion.

Secondly, instead of using the noise floor to estimate the threshold from the linear growth function, we used a value of zero. The resultant threshold estimates indicate that finding the intersection of the noise floor with the linear fit gave threshold estimates with lower standard deviations across subjects for peak to peak amplitude, RMS and peak PLV (*F*-test, Bonferroni corrected, $\alpha = 0.05$; $F_{19,19} = 3.83, 3.85, 6.87$ and $p = 0.005, 0.005, 0.0001$ respectively), while peak power did not ($F_{16,16} = 2.71$, $p = 0.054$).

Lastly, we compared linear curve fits to single level statistical testing. In Table 2, we compared the Hotelling test, as used by (Van Dun et al., 2015) to the four features in this paper using single level detection as described in the section 2.7 of Methods. A direct comparison between thresholds estimated using a growth function and using single level detection is difficult because different stimulus levels and data qualities at each level is required for either method. This difficulty is further explained in the Discussion. Nevertheless, we found no significant difference in the variance of the threshold estimates using the five different forms of single level testing as described in the Methods section (Levene's test, $W = 0.809$, $p = 0.523$).

Table 1: Threshold estimates using all data for each of the features. Mean ± 1 standard deviation and goodness of fit (adjusted r^2 value) shown. The first two rows show the results generated from the procedure as described in the Methods section: linear curve fits, and threshold calculated using the feature value of baseline response. The next two rows show results using an exponential curve fit, and the last row show results using the linear curve fit but the baseline response is assumed to have a feature value of zero.

		Time-domain features		**Time-frequency domain features**	
		Amplitude	**RMS**	**Power**	**PLV**
Proposed Method	Threshold estimate (dB SL +- SD)	5.7 ± 5.9	5.4 ± 7.6	18.5 ± 20.2*	2.7 ± 5.9
	Goodness of fit (adjusted r^2)	0.93 ± 0.07	0.94 ± 0.06	0.64 ± 0.29*	0.91 ± 0.10
Exponential fit	Threshold estimate (dB SL)	5.4 ± 5.8	6.5 ± 5.4	10.4 ± 10.0†	5.18 ± 3.89
	Goodness of fit (adjusted r^2)	0.89 ± 0.17	0.91 ± 0.09	0.61 ± 0.34†	0.88 ± 0.18
Zero for noise floor	Threshold estimate (dB SL)	-17.4 ± 11.7	-21.4 ± 14.8	-38.1 ± 33.3*	-21.1 ± 15.3

*Three subjects excluded due to invalid fit (see text)

†Eight subjects excluded due to invalid fit

Table 2: Threshold estimates generated from single level statistical tests. The Hotelling test is as described in (Van Dun et al., 2015), while the features explored in this study are tested for significance by comparing against their bootstrapped baseline (-20 dB SL epochs). The lowest level at which a cortical response is significantly different from baseline with the level immediately above it also detecting a response is deemed the threshold for each of the five methods.

	Hotelling	**Amplitude**	**RMS**	**Power**	**PLV**
Threshold estimate (dB SL)	17.8 ± 10.8	26.7 ± 15.9	29.5 ± 16.7	48.3 ± 15.9[†]	26.0 ± 14.3

[†]Eight subjects had responses that were not detected at any level

3.4 Peak-to-peak amplitude measures perform better than other features when a limited number of epochs are used

The results in Figure 4 show excellent threshold estimates, particularly for the peak-to-peak amplitude and peak PLV feature; however, up to 50 minutes' worth of data (mean = 38 minutes, std = 9.1 minutes) contributed to each threshold estimate. Therefore, we explored the effect of reducing the number of epochs acquired (and hence the testing time needed) on the threshold estimate accuracy. We expected that the feature noise level determines the accuracy of the threshold estimate, where increased feature noise level corresponds to decreased threshold estimate accuracy, since the feature noise level relates to the uncertainty of the feature points used in the growth function fit. We also expected that the number of epochs determines the feature noise level, where fewer epochs corresponds to increased feature noise level, since we expect the uncertainty of the feature points to decrease as more data is acquired. Thus, the expectation is that the threshold estimate accuracy decreases when using fewer epochs. We quantify the relation between the number of epochs used for feature calculations, the feature noise level and the threshold estimate accuracy in Figure 5.

<<Insert Figure 5 about here, in color>>

Figure 5A shows the relation between threshold estimate accuracy and feature noise level, using the peak PLV feature as an example. 300 feature noise levels, evenly spaced between 0.03 and 0.1 (arbitrary units), were used as feature noise thresholds. For each of the stimulus intensities, epochs were randomly sampled and added to the peak PLV calculation until the feature noise level was below the feature noise threshold. If the feature noise threshold could not be reached, all epochs in the block were used, and any requested feature noise levels

lower than the aforementioned were excluded from further analysis as they were duplicates using the same data. Following this, the growth function fitting and threshold estimation was performed. Thus, 6000 threshold estimates were calculated (300 feature noise levels for each of the 20 subjects) and each threshold estimate is represented by an individual point in Figure 5A. We observed that the standard deviation across subjects of the threshold estimates, shown by the outer black lines, increases with higher feature noise level, indicating a decrease in threshold estimate accuracy.

Figure 5B shows the relation between feature noise level and the number of epochs used, using one subject's data as an example. We observed a clear, monotonic decrease in feature noise level as more epochs were added, regardless of stimulus intensity. In conjunction with Figure 5A, it can be inferred that threshold estimate accuracy decreases with fewer epochs, and accordingly, less recording time.

We quantified the relation between recording time and threshold estimate accuracy and the results are shown in panel C. A moving window of 30 feature noise levels was used to aggregate the threshold points for each feature (an example of these points estimated with peak PLV is shown in panel A), and the spread of these threshold estimates gave a quantifiable performance measure. For example, using the PLV feature guaranteed with 95% confidence that, at a feature noise threshold of between 0.063 and 0.07, the threshold estimate was within 20 dB of the actual threshold. The number of epochs that were used in generating each threshold estimate was then converted to a recording time by assuming a constant 1.5 second ISI, resulting in 14.0 ± 1.7 minutes of recording time to obtain the threshold estimate. Recording time estimates are summarised in Table 1. These results indicate that when using the peak-to-peak amplitude feature for example, we can be 95% confident that a threshold estimate is within 20 dB of the behavioural threshold after recording for an average of 12.7 minutes.

These results are summarised in Table 3 by converting the feature noise level to recording time across all subjects. A two-way repeated measures ANOVA was performed in Minitab with the General Linear Model function with feature and threshold estimate accuracy restriction both as fixed factors. The model found significant differences in feature ($F_{2,3594}$ = 1343), threshold estimate accuracy restriction ($F_{1,3594}$ = 2819), and the interaction between feature and threshold estimate accuracy restriction ($F_{2,3594}$ = 175.2), with p<0.001 for all three cases. Post-hoc analysis using the Tukey pairwise comparisons revealed that each of the peak

amplitude, RMS and peak PLV features took significantly different recording times to reach the threshold estimate accuracy restriction (p<0.001 for all two feature comparisons across two threshold estimate accuracy restrictions).

Table 3. Recording time needed to reach 95% confidence that one threshold estimate is within 15 dB and 20 dB of the behavioural threshold. The results are expressed as the mean recording time in minutes with one standard deviation across all simulated threshold estimates that was at a 15 dB SL and 20 dB SL 95% confidence interval respectively.

Recording time needed for one threshold estimate to be:	**Time domain features**		**Time-frequency domain features**	
	Amplitude	**RMS**	**Power**	**PLV**
95% CI within ± 15 dB (mean ± std. minutes)	18.0 ± 4.1	28.6 ± 6.0	Never reached	19.4 ± 2.3
95% CI within ± 20 dB (mean ± std. minutes)	12.7 ± 3.1	18.0 ± 5.2		14.0 ± 1.7

4 Discussion

The main aim of this work was to compare features extracted from CAEPs to acoustic stimuli for use in hearing threshold estimation using growth function fitting. The significance of this work lies in its clinical value, where objective and accurate hearing threshold estimates are desired. Although we did not find a significant improvement by using time-frequency domain features, we have shown that growth function fitting of root-mean-square value and peak phase-locking value in addition to traditional amplitude growth functions are a viable method for hearing threshold estimates. In addition, we have validated the linear relationship between CAEP features and stimulus intensity in dB SL. The analyses defining minimum required recording time is the first step towards a fully automated threshold estimation technique. We have also shown that the accuracy of hearing threshold estimates is related to the recording time, allowing comparison of results for hearing threshold estimation using novel growth function algorithms.

4.1 Performance of features for threshold estimation

We observed similar performance between peak-to-peak amplitude, RMS and peak PLV for threshold estimation, while the performance using spectral power is far worse than the other features. For the growth function fit to accurately estimate threshold, the features must be present with lower but still supra-threshold stimuli. However, it has been shown that there is no significant change in spectral power of the CAEP when the stimulus is at a low intensity (around 30 dB SL), whereas significant phase changes are present (Savers et al., 1974).

Furthermore, the similarity in performance between the other three features may be due to event-related dynamics represented by the features. As peak-to-peak amplitude and RMS operate on the averaged waveform, and peak PLV extracts phase locking of responses to individual trials, these methods extract responses conforming to the partial phase resetting (PPR) and the event-related potential (ERP) paradigms. As a change in spectral power from baseline would also be present in the ERP paradigm yet has poor performance in this study, it is possible that PPR is the underlying event-related dynamic being described by amplitude, RMS and PLV features. Harris et al. (2014) showed a correlation between P2 amplitude and PLV, but not spectral power, thus also suggesting the PPR paradigm, in their case for cortical responses to gaps in noise. The event-related dynamics represented by the different features could be further tested by classification using features. By classifying CAEPs from background brain activity using different combinations of features, a mutual information metric between actual response and predicted response could provide further insights into features extracted from EEG and their underlying event-related dynamics.

4.2 Growth of CAEP features with stimulus intensity

We made two assumptions in our method for estimating hearing thresholds from growth functions. First, we assumed that the growth of these features with stimulus intensity is linear, at least in the intensity range of interest (threshold up to 60 dB SL). Although exponential fits have been used previously (Abbas et al., 2015; Ross et al., 1999; Visram et al., 2015), we have shown that the linear curves fit equally well. The differences in the types of functions used in these studies may be due to several factors. Firstly, there is a difference between acoustic and electric hearing, although these differences have not been quantified. One possibility is the much larger dynamic range in decibels of acoustic listeners than electric listeners; this may cause the growth functions in acoustic listeners to have a steeper slope or to change its shape at higher levels. Secondly, different EEG response features were used in

each: global field power gave a mix of linear and exponential fits across individual subjects (Visram et al., 2015) and N1 amplitude gave an exponential fit (Ross et al., 1999). Finally, Ross et al. (1999) used a grand average growth function to examine the characteristics of amplitude feature growth with stimulus intensity, but in our results (Figure 4B) and in those of Visram et al. (2015) and Abbas et al. (2015), individual growth functions differ. Using a grand average growth function may instigate the negative effects of inter-subject variability and give inaccurate growth functions and thus poorer threshold estimates.

It must also be noted that even though the growth function fit gives accurate hearing threshold estimates, the shape of the true underlying growth function is difficult to determine. A linear function would not be realistic at higher intensity levels because the features cannot continue growing; likewise, an exponential function would not be realistic at near- or below-threshold levels as the features are expected to be near the baseline value and cannot be negative. The accurate threshold estimates, however, suggest both functions are a good approximation of the shape of the underlying growth function, at least for stimulus levels within the person's dynamic range of hearing.

Second, we found that inferring behavioural threshold from the noise floor and fitted growth function gave better hearing threshold estimates than using a noise floor of zero (Table 1). This contrasts with both Visram et al. (2015) and Ross et al. (1999), who assumed that the threshold is the stimulus intensity at which the fitted function value is zero. Our results provide evidence that the CAEP features for the supra-threshold levels contain an amount of noise in addition to that of the underlying response. For the peak-to-peak amplitude, peak PLV and peak spectral power features, choosing the peak value introduces a positive bias to the feature; moreover, for the RMS and PLV, there is an inherent positive bias due to the way they are calculated. RMS includes the spontaneous EEG activity, while the baseline PLV is distributed based upon the number of epochs in its calculation. Performing the same calculations on the baseline level improves threshold estimation performance and provides evidence that it at least partially accounts for the aforementioned biases.

4.3 The effect of recording time on threshold estimation accuracy

We have defined the notion of feature noise level, which describes the uncertainty in the features, becomes smaller as more data is acquired. We have subsequently shown that objective hearing threshold estimation accuracy is related to the feature noise level. This is useful knowledge for clinical applications as the trade-off between recording time and

accuracy can now be quantified, and a more informed choice can be made when designing a data acquisition paradigm.

Hoth (1993) and Van Dun et al. (2015) found a standard deviation of 11.4 and 10.2 dB difference from the behavioural threshold respectively, with a small but considerable number of outliers that had a difference of greater than 30 dB. In this study, we found a standard deviation of 5.9 dB using both the amplitude and the PLV feature with no differences greater than 20 dB. This came at the expense of recording for much longer than the previous studies. While Hoth (1993) did not give an indication of test time as more stimulus intensities than needed were presented, similar recording time per level was used compared to Van Dun et al. (2015) (50 sweeps with 2.5s ISI, 120 sweeps with 1.175s ISI respectively). Van Dun et al. used 10.6 ± 1.2 minutes to reach standard deviation of the estimate of 10 dB; in this study, we achieved a similar result (95% within 20 dB, equal to a standard deviation of 10.2 dB) this in 12.7 ± 3.1 minutes using the peak-to-peak amplitude feature. Our study shows that the test time is slightly longer, meaning further improvements can consolidate growth function fitting as the preferred method for objective threshold estimation.

4.4 Comparison between growth functions and statistical testing for threshold estimates

Growth function fitting is advantageous over statistical testing in two ways. Firstly, a set of measured responses which do not have significant amplitude change or phase locking does not imply that those epochs do not contain a response to the sound; rather, the model that has been built may not have sufficient epochs or statistical power to detect a significant response. The growth function fitting method can still make use of the information in epoch blocks in which the chosen feature is not statistically significant. Secondly, in single-level statistical testing, the accuracy of the threshold estimate is limited by the spacing of the stimulus levels tested. We have shown this in Table 2 where the hearing threshold estimates using single level testing are shown to have far greater variance across subjects than when using the growth function method. As previously mentioned, direct comparison between these methods is biased, as a true single-level detection algorithm would adaptively test levels (Van Dun et al. 2015) instead of larger, 20 dB steps.

The stopping criterion for acquiring epochs is a possible disadvantage of the growth function method compared to statistical testing. This is important as recording time is a key factor for whether a method is clinically viable. For statistical testing, false detection rates can be controlled, and acquisition of an epoch block can be terminated as soon as significance is

detected. In this work, we remedied this by showing that the feature noise level, and the number of epochs used, is a useful measure for the accuracy of the threshold estimate. The feature noise level can be used as a stopping criterion, whereas the number of epochs used gives an indication of how quickly the threshold estimate can be obtained.

4.5 Unexplored parameter space for growth function fitting

There is a large, unexplored parameter space with the growth function method which may impact the recording time and accuracy of the threshold estimate. We will consider clinical relevance when discussing the following points. Firstly, the number of stimulus levels cannot be pre-determined in a clinical setting. For this method to be clinically applicable, an algorithm that will determine the best stimulus levels and dynamically test a range of levels is needed. Further work is needed on exploring the optimal number of points to fit the growth function: to optimise recording time, the relation between threshold estimate accuracy, number of stimulus levels and feature noise level needs to be identified. Secondly, the feature baselines were calculated from a well-below-threshold level in this study. It is possible to explore the use of the EEG response in the silent region between stimuli to calculate the baseline instead, saving further time. Lastly, audiograms or cochlear implant fitting generally require finding more than one hearing threshold at a time; testing multiple levels where different tones are mixed can shorten recording time further (Lightfoot et al., 2006).

5 Conclusion

In this study, we have shown for the first time the performance of growth function fitting to objectively estimate hearing thresholds in normal hearing adults, using the CAEP recorded from one active electrode. The results show that linear growth functions produce accurate threshold estimates for all subjects with any of peak-to-peak amplitude, RMS or peak PLV features. In addition, growth function hearing threshold estimates are comparable to existing methods for fully objective hearing threshold estimate using the cortical auditory evoked response. Future work to explore other parameters of growth function fitting is necessary to further improve hearing threshold estimation.

6 Acknowledgements

Darren Mao was supported by an Australian Government Research Training Program Scholarship. Colette McKay was supported by a Veski innovation fellowship. Hamish Innes-Brown was supported by a National Health and Medical Research Council (Australia) Early-Career Research Fellowship. The Bionics Institute acknowledges the support it receives from the Victorian Government through its Operational Infrastructure Support Program.

7 References

Abbas, P.J., Brown, C.J. 1991. Electrically evoked auditory brainstem response: growth of response with current level. Hear Res 51, 123-37.

Abbas, P.J., Brown, C.J. 2015. Assessment of responses to cochlear implant stimulation at different levels of the auditory pathway. Hear Res 322, 67-76.

Brown, C.J., Lopez, S.M., Hughes, M.L., Abbas, P.J. 1999. Relationship between EABR thresholds and levels used to program the CLARION® speech processor. Ann Otol Rhinol Laryngol 108, 50-57.

Brown, C.J., Hughes, M.L., Luk, B., Abbas, P.J., Wolaver, A., Gervais, J. 2000. The relationship between EAP and EABR thresholds and levels used to program the nucleus 24 speech processor: Data from adults. Ear Hear 21, 151-163.

Coles, R.R., Mason, S.M. 1984. The results of cortical electric response audiometry in medico-legal investigations. Br J Audiol 18, 71-8.

Davis, H. 1965. Slow cortical responses evoked by acoustic stimuli. Acta Oto-laryngologica 59, 179-185.

Efron, B., Tibshirani, R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci, 54-75.

Elberling, C., Don, M. 1984. Quality estimation of averaged auditory brainstem responses. Scand Audiol 13, 187-97.

Galambos, R., Makeig, S., Talmachoff, P.J. 1981. A 40-Hz auditory potential recorded from the human scalp. Proc Natl Acad Sci U S A 78, 2643-7.

Golding, M., Dillon, H., Seymour, J., Carter, L. 2009. The detection of adult cortical auditory evoked potentials (CAEPs) using an automated statistic and visual detection. Int J Audiol 48, 833-42.

Goldstein Jr, M.H., Kiang, N.Y.S. 1958. Synchrony of neural activity in electric responses evoked by transient acoustic stimuli. J Acoust Soc Am 30, 107-114.

Harris, K.C., Vaden, K.I., Jr., Dubno, J.R. 2014. Auditory-evoked cortical activity: contribution of brain noise, phase locking, and spectral power. J Basic Clin Physiol Pharmacol 25, 277-84.

Hotelling, H. 1992. The generalization of Student's ratio, Breakthroughs in Statistics. Springer. pp. 54-65.

Hoth, S. 1993. Computer-aided hearing threshold determination from cortical auditory evoked potentials. Scand Audiol 22, 165-77.

Jewett, D.L., Romano, M.N., Williston, J.S. 1970. Human auditory evoked potentials: possible brain stem components detected on the scalp. Science 167, 1517-8.

Lightfoot, G., Kennedy, V. 2006. Cortical electric response audiometry hearing threshold estimation: accuracy, speed, and the effects of stimulus presentation features. Ear Hear 27, 443-56.

Makeig, S. 1993. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroen Clin Neuro 86, 283-93.

Mardia, K.V. 2014. Statistics of directional data Academic press.

McKay, C.M., Smale, N. 2017. The relation between ECAP measurements and the effect of rate on behavioral thresholds in cochlear implant users. Hear Res 346, 62-70.

McKay, C.M., Chandan, K., Akhoun, I., Siciliano, C., Kluk, K. 2013. Can ECAP Measures Be Used for Totally Objective Programming of Cochlear Implants? J Assoc Res Otolaryngol 14, 879-890.

Miller, C.A., Brown, C.J., Abbas, P.J., Chi, S.L. 2008. The clinical application of potentials evoked from the peripheral auditory system. Hear Res 242, 184-97.

622 Pearce, W., Golding, M., Dillon, H. 2007. Cortical auditory evoked potentials in the
623 assessment of auditory neuropathy: two case studies. J Am Acad Audiol 18, 380-90.
624 Picton, T.W., Woods, D.L., Baribeau-Braun, J., Healey, T.M. 1976. Evoked potential
625 audiometry. J Otolaryngol 6, 90-119.
626 Picton, T.W., Vajsar, J., Rodriguez, R., Campbell, K.B. 1987. Reliability estimates for
627 steady-state evoked potentials. Electroencephalogr Clin Neurophysiol 68, 119-31.
628 Pratt, H., Sohmer, H. 1978. Comparison of hearing threshold determined by auditory pathway
629 electric responses and by behavioural responses. Audiology 17, 285-92.
630 Rickards, F.W. 1996. Cortical evoked response audiometry in noise induced hearing loss
631 claims. Aust J Otolaryng 2, 237.
632 Ross, B., Lutkenhoner, B., Pantev, C., Hoke, M. 1999. Frequency-specific threshold
633 determination with the CERAgram method: basic principle and retrospective
634 evaluation of data. Audiol Neurootol 4, 12-27.
635 Savers, B.M., Beagley, H., Henshall, W. 1974. The mechanism of auditory evoked EEG
636 responses. Nature 247, 481-483.
637 Starr, A., Picton, T.W., Sininger, Y., Hood, L.J., Berlin, C.I. 1996. Auditory neuropathy.
638 Brain 119 (Pt 3), 741-53.
639 Tallon-Baudry, C., Bertrand, O., Delpuech, C., Pernier, J. 1996. Stimulus specificity of
640 phase-locked and non-phase-locked 40 Hz visual responses in human. J Neurosci 16,
641 4240-9.
642 Tomlin, D., Rance, G., Graydon, K., Tsialios, I. 2006. A comparison of 40 Hz auditory
643 steady-state response (ASSR) and cortical auditory evoked potential (CAEP)
644 thresholds in awake adult subjects. Int J Audiol 45, 580-588.
645 Van Dun, B., Dillon, H., Seeto, M. 2015. Estimating hearing thresholds in hearing-impaired
646 adults through objective detection of cortical auditory evoked potentials. J Am Acad
647 Audiol 26, 370-83.
648 Visram, A.S., Innes-Brown, H., El-Deredy, W., McKay, C.M. 2015. Cortical auditory evoked
649 potentials as an objective measure of behavioral thresholds in cochlear implant users.
650 Hear Res 327, 35-42.
651 Yeung, K.N., Wong, L.L. 2007. Prediction of hearing thresholds: Comparison of cortical
652 evoked response audiometry and auditory steady state response audiometry
653 techniques. Int J Audiol 46, 17-25.

654

Figure 1: Representation of the threshold estimation method used in this study. Multiple sound intensity levels are presented, labeled as L_1, L_2, L_3 as well as a sub-threshold intensity level, L_{sub}; multiple epochs of responses to each are recorded. Features are then extracted from each block of recorded EEG activity, labeled as f_{L1}, etc. The baseline feature estimate is the feature extracted from baseline EEG activity, as there should be no response to a sub-threshold stimulus. For growth funtion threshold estimation, a linear curve is fit to all EEG features except the baseline feature estimate, and the sound stimulus intensity that corresponds to the sub-threshold response feature is the hearing threshold estimate.

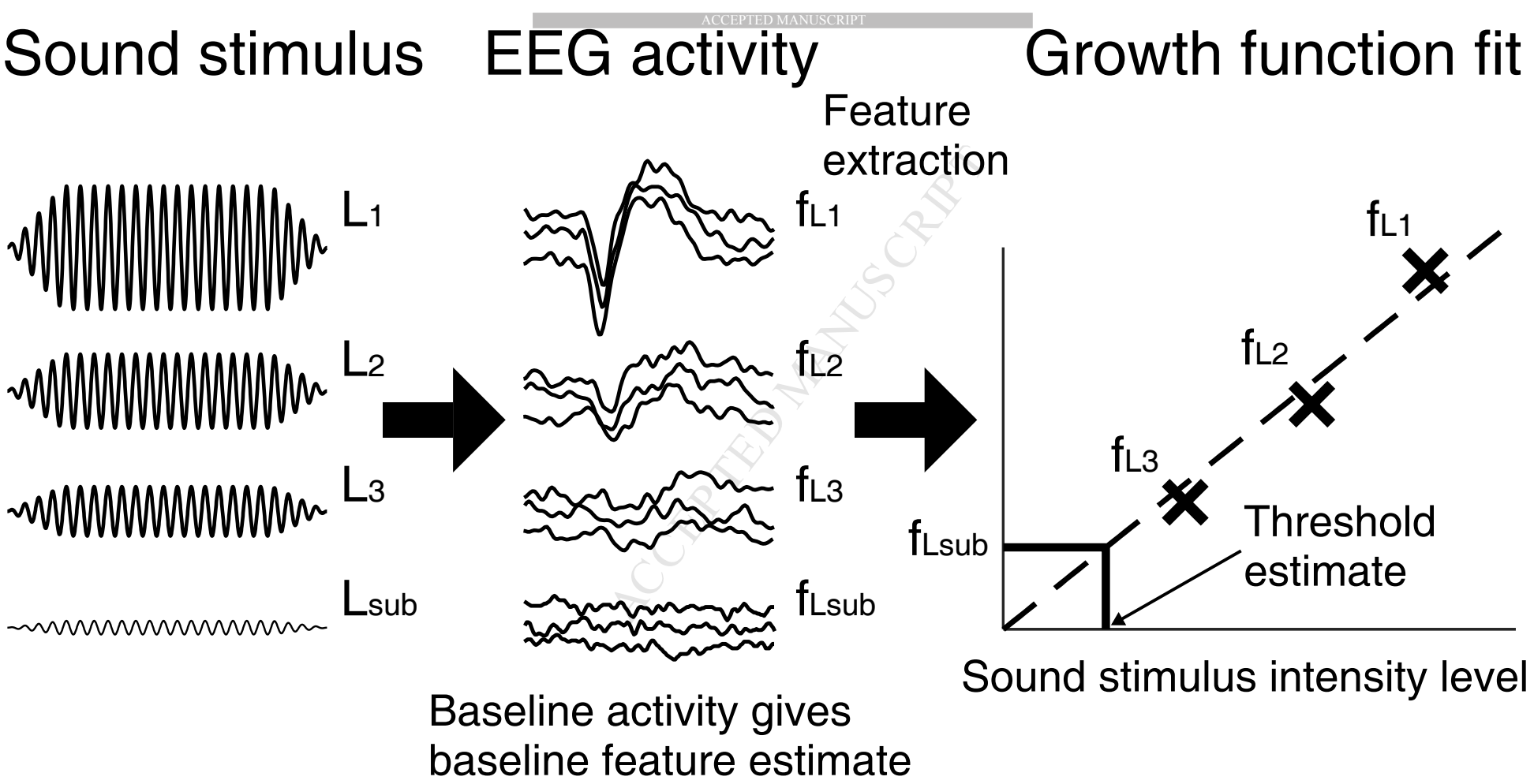
Figure 2: Top row: Responses to 60 dB SL stimuli for subject 14, represented in time and time-frequency domain, with illustration of the four features extracted for growth function fitting. A: Time-domain responses at 60 dB SL. Shaded errorbars enclose three standard errors of the mean (n=151 epochs). The peak to peak amplitude and the RMS value are extracted from a 50ms to 500ms post-stimulus window, **B:** Spectra of power change from baseline (arbitrary units, A.U.); black outline enclose the significant region of power change (z>3, permutation test with baseline at -300ms). Black dot indicates the time-frequency location of the peak power, and thus the extracted feature value. **C:** Spectra of phase locking value, significant region and peak location as with panel B, in arbitrary units (A.U.). **Bottom row: all subject responses to 60 dB SL stimuli. D:** All subjects' averaged time domain responses to 60 dB SL stimuli. Shaded errorbars enclose one standard deviation (n=20 subjects). **E and F:** Power and phase locking spectra averaged across all subjects respectively. Black points signify peak location for each of the twenty subjects in the time-frequency domain and demonstrate that the spread of peak location across subjects is smaller for PLV.

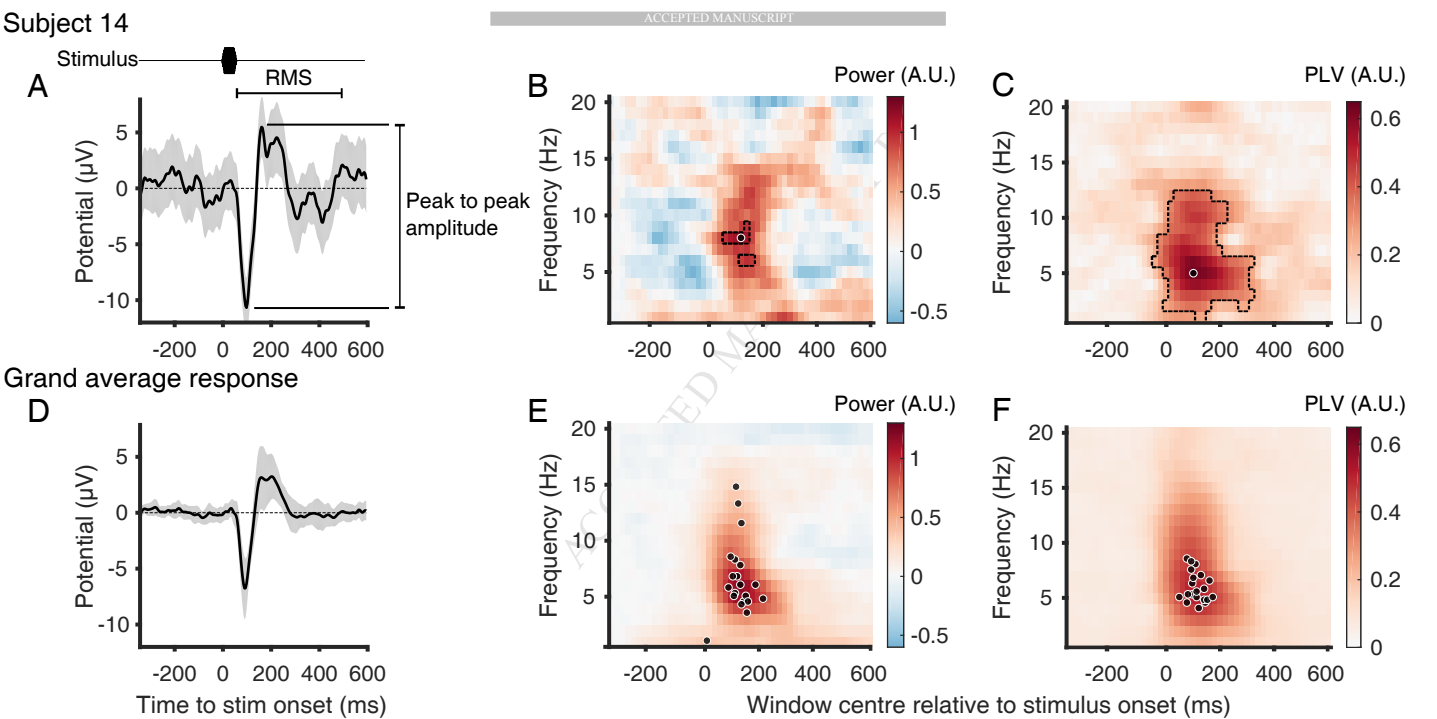
Figure 3: Each feature plotted against stimulus intensity over all subjects. The filled circle and error bars represent the feature mean and one standard deviation respectively. Panels A to D indicate the peak-to-peak amplitude, root-mean-square, peak power and peak phase-locking-value features respectively.

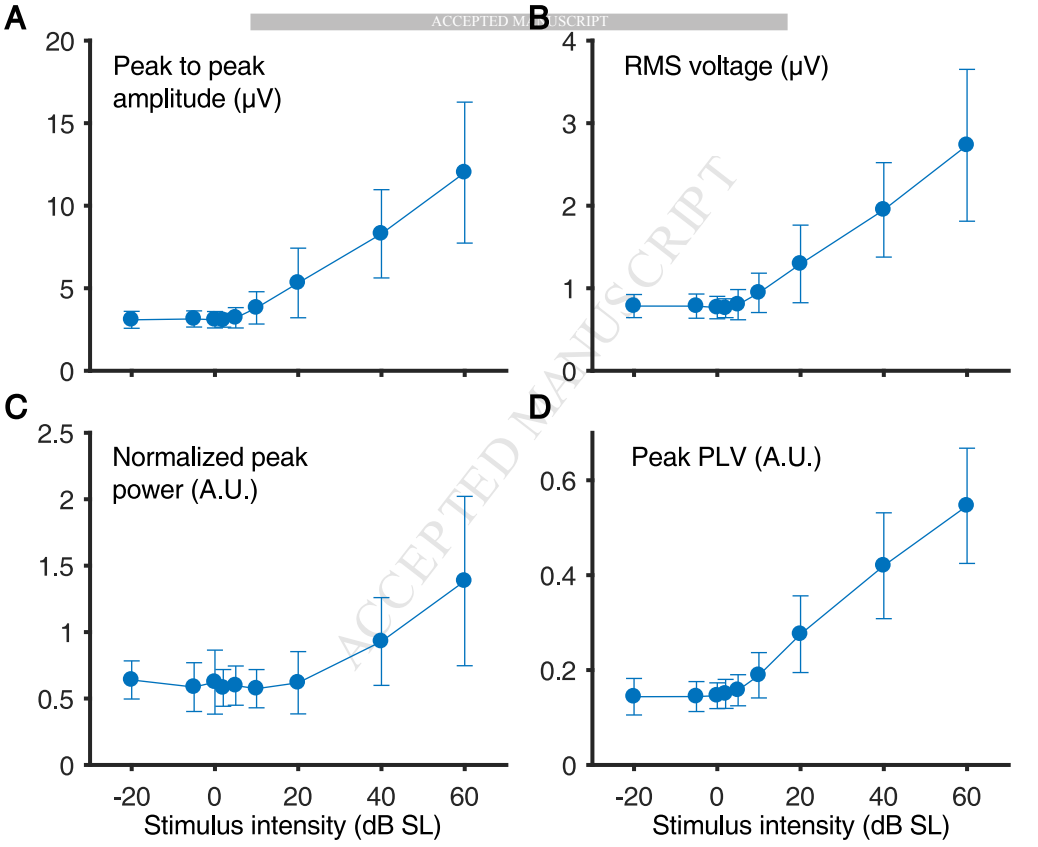
Figure 4: Growth function fitting procedure and threshold estimates using all epochs. A. An example in subject 14 using the peak PLV to find the threshold estimate. **B.** Growth function fits (straight lines) and threshold estimates (circles) for all subjects. The black circles

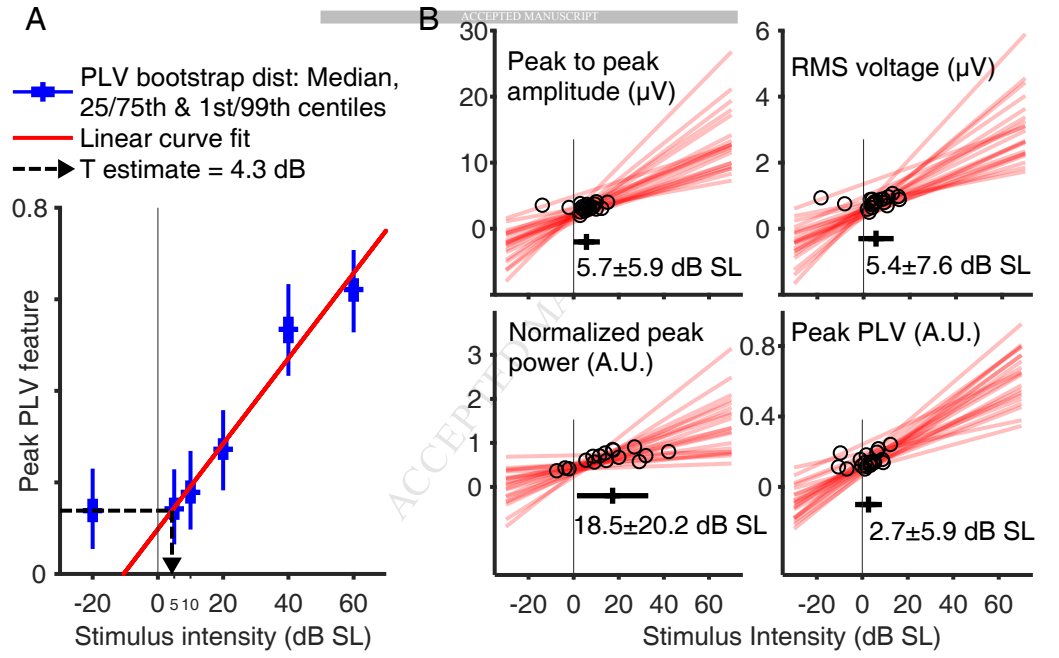
show the spread of threshold estimates across each subject. A point to the right of the vertical behavioral threshold line is a positive threshold estimate and indicates an over-estimation of the threshold, and vice versa for a negative estimate corresponding to an under-estimation. The horizontal bar indicates the mean and one standard deviation of threshold estimates across subjects. Four subjects were omitted from fits for peak power due to invalid fits (see text).

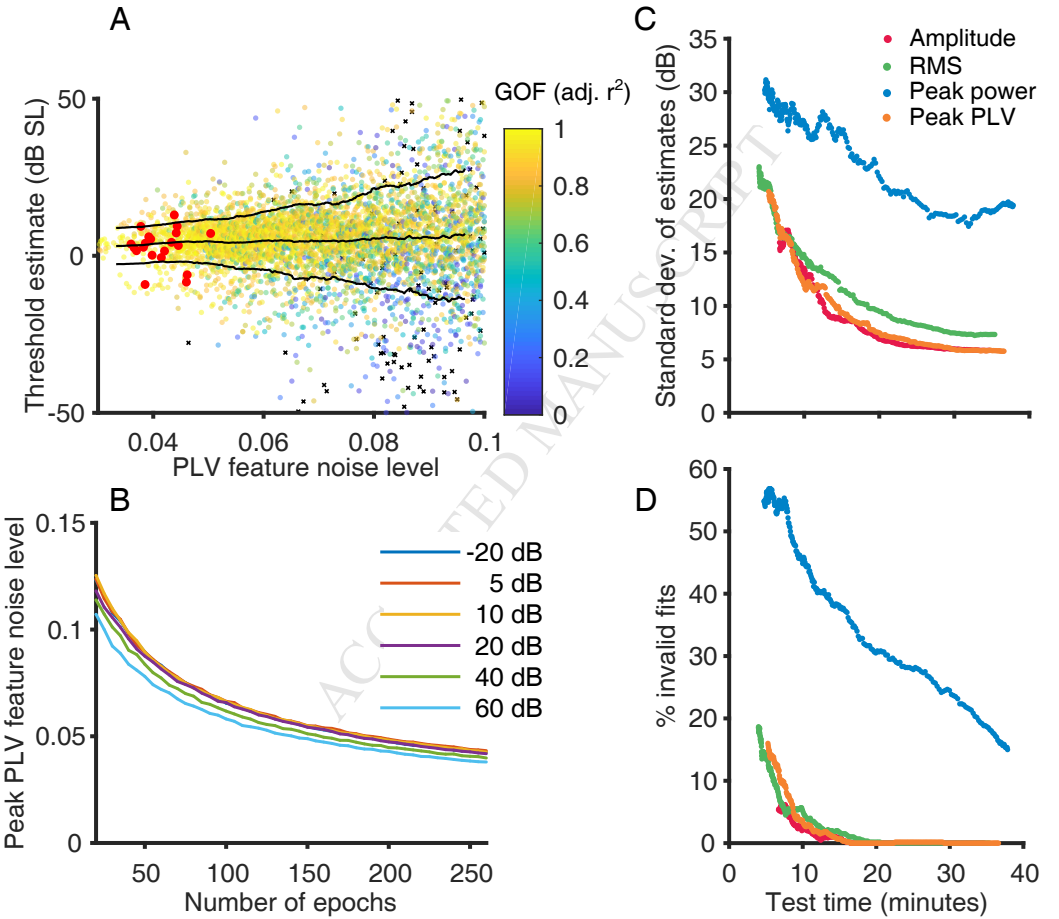
Figure 5. Deterioration of threshold estimate accuracy when decreasing testing time. A. Threshold estimate accuracy (quantified as standard deviation of threshold estimate) decreased with increasing PLV feature noise level, which is inversely related to acquisition time (Panel B). Each point is one threshold estimate and color indicates goodness of the growth function fit (Adjusted r^2 value); red points indicate threshold estimate when using all epochs; black crosses are invalid fits. Black lines indicate threshold estimate mean and one standard deviation respectively. **B.** PLV feature noise decreases with increasing number of epochs and is largely independent of level. **C.** (Top) Scatterplot of threshold estimate standard deviation for each feature against their respective feature noise levels, an example of which is shown in panel A for peak PLV, before converting feature noise level to testing time in minutes. **D.** The percentage of invalid fits decreased when increasing test time.











Highlights

- Growth functions provide clinically viable objective estimates of hearing threshold
- Cortical response peak-to-peak amplitudes and phase locking values are suitable
- Best threshold estimates are consistently within 6 dB of behavioural threshold
- 13 minutes testing time for ±20 dB accuracy with single electrode recording