The Precision of eCAP Thresholds Derived from Amplitude Growth Functions

Jan Dirk Biesheuvel[1], Jeroen J. Briaire[1], Johan H.M. Frijns[1,2]

[1] ENT-department, Leiden University Medical Centre, Leiden, The Netherlands

[2] Leiden Institute for Brain and Cognition, Leiden, The Netherlands

Conflicts of Interest and Source of Funding

This study was supported by the Dutch Technology Foundation STW and Advanced Bionics.

The authors have no conflicts of interest to disclose.

Correspondence

Prof. Johan H.M. Frijns MD PhD, ENT Department, Leiden University Medical Centre, P.O. BOX 9600,

2300 RC Leiden, the Netherlands, E-mail: j.h.m.frijns@lumc.nl

ABSTRACT

Objective: An amplitude growth function (AGF) shows the amplitude of an electrically evoked compound action potential (eCAP) as a function of the stimulation current. AGFs can be used to derive the eCAP threshold, which represents the minimum amount of current needed to elicit a measurable eCAP. eCAP thresholds have been widely used clinically to, for example, assist with sound processor programming. However, no eCAP precision has been included so far. The aim of this study was to investigate the precision of eCAP thresholds and determine whether they are precise enough for clinical use.

Design: The study is retrospective and the data comprised 826 AGFs, intra-operatively measured in 111 patients implanted with a HiRes90K cochlear implant (Advanced Bionics). For each AGF, the eCAP threshold was determined using two commonly used methods: linear extrapolation (LE) towards the x-axis and detection of the last visible eCAP (LV). Subsequently, the threshold confidence interval (TCI) of each eCAP threshold was calculated to serve as a metric for precision, whereby a larger TCI means a lower precision or reliability. Additionally, the eCAP thresholds results were compared to most recent behavioral fitting thresholds (T-profile) to put the eCAP threshold analysis in clinical context. Thereby, the association between eCAP and behavioral thresholds was calculated, both for all subjects together (group analysis) and, in contrast to previous studies, within individual subjects.

Results: Our data show that the TCIs were larger with the LE method than with the LV method. The eCAP thresholds estimated by the LE method were systematically smaller than those estimated by the LV method, while the LE thresholds with the smallest TCIs correlated best with the LV thresholds. Correlation analysis between eCAP and behavioral thresholds revealed correlation coefficients of $r = 0.44$ and $r = 0.54$ for the group analysis of LE and LV thresholds, respectively. Within individual subjects, however, the correlation coefficients varied from approximately -1 to +1 for both LE and LV

thresholds. Further analysis showed that across subjects the behavioral thresholds fell within the TCIs of the eCAP threshold profiles.

Conclusion: This study shows that eCAP thresholds have an uncertainty that can be estimated using TCIs. The size of the TCI depends on several factors, e.g., the threshold estimation method and measurement conditions, but it is often larger than one would expect when just looking at the threshold values. Given these large TCIs, future research on eCAP thresholds should be accompanied by a measure of precision to correctly apply eCAP thresholds in clinical practice. Comparing our eCAP threshold results with T-profiles indicates that the eCAP thresholds are possibly not precise enough to predict T-profiles.

Keywords: Precision, Amplitude growth function; Cochlear implant; Electrically evoked compound action potential; Threshold; Fitting

INTRODUCTION

The electrically evoked compound action potential (eCAP) of the auditory nerve can be measured using the telemetry function of a cochlear implant (CI). The eCAP is a synchronous response from multiple auditory nerve fibers evoked by electrical stimulation. An eCAP is typically recorded as a negative peak (N_1) followed by a positive peak or plateau (P_1). The amplitude of the response is measured from peak to peak and reaches values of several hundred microvolts (Abbas et al. 1999). Since the eCAP amplitude depends on the applied current, plotting the amplitude as function of stimulus level will result in an amplitude growth function (AGF) (Figure 1A). From this AGF, the eCAP threshold can be derived, representing the minimum amount of current needed to evoke a measurable neural response (Hughes 2013). These thresholds can be used to assist with sound processor programming predicting behavioral threshold levels (T-profiles) and maximum comfort levels (C/M-level). However, no clear correlations between eCAP thresholds and fitting levels have been found (Franck & Norton 2001; Franck 2002; Potts et al. 2007; Holstad et al. 2009). Possibly the mismatch between eCAP thresholds and T-profiles can be explained by an error analysis of eCAP thresholds. Two commonly used terms in error analysis are 'accuracy' and 'precision'. Accuracy describes the difference between the estimated and the true value, whereas the precision describes the error around the estimated value. Since the eCAP threshold is a (best) estimate, we do not know the true value, and the accuracy can hardly be determined. Therefore, the focus of this study is on the precision of the eCAP thresholds as derived using standard (clinical) methodologies and the consequences for clinical practice.

Two methods are commonly used to estimate eCAP thresholds (Figure 1A). The simplest method is detection of the last visible eCAP (LV). This method tracks all recordings of the AGF from high to low current while detecting whether the recording still contains an eCAP. The current level where the eCAP was last visible is defined as the eCAP threshold. The LV threshold (green dot) of the AGF shown in Figure 1A is 200 clinical units (CU). Because eCAP amplitudes below noise level cannot

be detected, the LV threshold reflects the eCAP threshold at noise level. The second method is called linear extrapolation (LE). This method uses a linear regression line (Figure 1A, blue line) through data points representing a true neural response (black dots) to estimate the eCAP threshold at zero amplitude. The intercept of the line with the x-axis is defined as the eCAP threshold (blue dot at 159 CU). In general, eCAP thresholds are determined automatically using smart algorithms followed by visual inspection by an expert (clinician, audiologist). Visual inspection is necessary to verify whether the algorithm has not been misguided by, for example, (stimulus) artefacts or noise. Moreover, one should be aware of bias (e.g., systematically over- or under-estimation of threshold) due to methodology or observer (Glassman & Hughes 2013; Akhoun et al. 2015).

Thus far, AGFs have been measured without any precision analysis; eCAP amplitudes have been used as data points without measurement error (Figure 1A). However, the eCAP amplitude has a measurement error that can easily be included, and it would be worthwhile to include also a measure of the precision of eCAP thresholds. For example, the measurement error of the eCAP amplitude can be derived from a measurement without stimulation or from a section of the recording in which no neural response is expected (e.g., the last samples of the recording). In Figure 1B the measurement error is added to the eCAP amplitude, as shown by the error bars. The method of threshold estimation itself also has uncertainty. Unfortunately, standard clinical AGFs do not contain the required information to calculate the measurement error of the eCAP threshold itself. Alternatively, we defined the threshold confidence interval (TCI) as a metric of threshold precision. The blue and green lines in Figure 1B show how the precision of the LE and LV methods can be determined and this is explained below in the methods section.

The aim of this study was to investigate the precision of the eCAP threshold and the consequences on clinical practice. The study was retrospective and we made use of large dataset of eCAP recordings, which were measured regularly during surgery. Further, the LV and LE methods were evaluated as commonly done in the literature and clinic. The LV threshold was identified by a human observer rather than using an automated computer algorithm. The LE method was based on a linear fit towards the zero-

intercept, though a non-linear fit or the intercept just above noise level would be better. To put the eCAP threshold precision in clinical context, we additionally compared the eCAP threshold analyses to behavioral T-profiles, and the association between both measures was calculated as well. To the best of our knowledge, this study is the first to perform this analysis for such a large population of CI recipients using an Advanced Bionics (AB) implant. In contrast to previous studies, the correlation analysis was not only performed for all subjects together (group analysis), but also within individual subjects. This makes present study one of the few studies examining the association between eCAP and behavioral thresholds within individual subjects (de Vos et al. in press).

METHODS

Patients and Data

The AGFs used in this study originate from intra-operative eCAP measurements sequentially recorded in our hospital from January 2010 to December 2015. AGFs were available from 191 patients. All patients were implanted with a HiRes90K^{TM} device (Advanced Bionics, Sylmar, CA), either with a 1J or Mid-Scala electrode array, both consisting of 16 electrode contacts (1-16 in an apical to basal order). The AGFs were collected as part of our standard clinical eCAP recordings using forward masking and the Research Studies Platform Objective Measures (RSPOM) software from AB. This allowed us to compare the AGF data with the other eCAP recordings, including spread of excitation (Biesheuvel et al. 2016) and refractoriness. The AGFs were measured on all odd electrode contacts using the following parameters: monopolar cathodic first biphasic pulses; pulse duration, 32 µs/phase; masker probe interval, 400 µs; sweep rate, 25 Hz; masker offset, 0%; sampling rate, 56 kHz; gain, 300; 32 averages; recording electrode two electrodes apical to stimulus (except electrode 1 where the recording electrode was two electrodes basal to the stimulus).

The eCAP analysis was performed automatically by RSPOM using the default settings. Raw eCAP recordings were filtered using a low pass filter with a cut-off frequency of 8 kHz. Peak N_1 was detected as the minimum over the time period between 180 and 490 µs and peak P_1 as the

maximum between 470 and 980 µs after the end of stimulation. The eCAP amplitude was calculated as the peak-to-peak voltage between P_1 and N_1. After the automatic analysis, all peak detections and data points were verified. If necessary, corrections were made manually. In addition, the signal-to-noise ratio (SNR) and measurement error of each eCAP were calculated using MATLAB (Mathworks, Inc., Natick, MA). The SNR of the eCAP was calculated as the eCAP amplitude divided by the noise amplitude, which in turn was defined as the maximal amplitude difference in the last 30 samples of the response. We assumed that no possible remaining artefact or neural response would occur over this section. The measurement error of the eCAP was set to the noise amplitude.

The total batch of AGF measurements comprised AGFs measured with six and ten current steps. The functions with six current steps were measured when a full AGF with ten current steps could not be recorded due to time limitations during surgery. These six-point curves were measured in 56 patients and were, unfortunately, not reliable enough for the purpose of this study. Therefore, only the ten-point AGFs, measured in 135 patients, were analyzed. Each AGF was visually inspected for inclusion; an AGF was included when at least one clear eCAP response was available and the AGF was not disrupted, such as due to a stimulus artefact. Finally, a total of 826 AGFs originating from a heterogenic group of 111 patients were included for further analysis (Table 1).

To put the eCAP threshold analysis in clinical context, we additionally compared the eCAP threshold results with behavioral T-profiles. Most recent fitting thresholds were extracted from the SoundWave fitting software, assuming that these T-profiles were stable and most optimal. To ensure the reliability of the T-profiles, we only included profiles measured in subjects with an age above 4 years, resulting in T-profiles from 73 subjects. The threshold levels were measured either for each electrode contact separately or by using streamlined programming (Plant et al. 2005). For these measurements we used tone bursts of biphasic pulse trains of 200 ms, as provided by the clinical speech program of the CI user. The measurement started at a subthreshold stimulus level and the level was increased until the patient heard the sound. After a clearly audible percept was achieved,

the stimulus level was decreased to reach a subthreshold level again. Subsequently the stimulus level was raised to find the final threshold level. The behavioral T-profiles were never set automatically using 10% of most comfortable level (optional in SoundWave).

eCAP Thresholds and Precision

The eCAP thresholds were calculated using a semi-automatic method programmed in MATLAB. The method automatically detects which recordings contain an eCAP using two conditions: the eCAP amplitude exceeds 20 µV and the SNR of the eCAP exceeds +13 dB Subsequently, all eCAPs and AGFs were visually inspected and, if necessary, corrections made manually.

The LE method was based on a linear least square fit through all data points containing a neural response. Data points at the top of the AGF which clearly deviate from linear amplitude growth were excluded from the fit, as common in clinical practice. At least three data points were required because confidence intervals cannot be determined if the number or observations is equal to the number of coefficients. The precision of the LE threshold was calculated using the 95% confidence interval of the linear fit (Figure 1B, blue dashed line); there is a 95% probability that the true best-fit line for the AGF lies within this confidence interval. The confidence interval was extrapolated towards the x-axis in order to estimate the interval at zero-amplitude. For this purpose, the steepest and least steep linear fits within the confidence interval were calculated, and the intercepts of these lines with the x-axis were used to extrapolate the lower and upper boundaries of the confidence interval, respectively. Next, the distance between the two intercepts of the confidence interval with the x-axis was defined as the TCI. This TCI will be used as a measure of precision of the eCAP threshold estimated by the LE method.

For the LV method, the eCAP threshold was set to the current level where the smallest true eCAP could be detected. This threshold can be found by tracking all recordings of the AGF from high to low current. The lowest current whereby an eCAP could be detected was defined as the eCAP

threshold. Estimating the precision of LV thresholds was less straight forward than with the LE method because the LV method does not have any mathematical clues. In our opinion, a good method was to use several guides as depicted in green in Figure 1B. The noise level (horizontal solid line) was estimated using data points not representing an eCAP (grey dots). Subsequently, a guide was drawn from the data point representing the LV (abscissa: 200 CU, ordinate: 68 µV) through the calculated noise level (abscissa: eCAP threshold minus one current step = 150 CU, ordinate: 38 µV). This guide (dashed line) represents the precision of the eCAP threshold towards the left side. If the noise level is zero, the size of the confidence interval at the left side of the threshold will be equal to the step size in current; the true zero-amplitude threshold will be between the estimate threshold and lower current. A high noise level, which is close to the smallest detectable eCAP amplitude, will result in a large(r) left confidence interval. Thus, due to the high noise level, it is uncertain where the zero-amplitude threshold exactly is. In addition, there is a high chance that the detected amplitude is not an eCAP, but still noise. In principle, the zero-amplitude threshold is at a lower current than the LV threshold and, consequently, the confidence interval at the right side could be set to zero. However, it is important to also include the eCAP measurement error, especially in cases in which the noise level is close to the smallest detectable eCAP amplitude. Therefore, the amplitude variation (error bar) in the last visible eCAP was translated into CU using a guideline between the last visible and next eCAP (solid green line). Projecting the top of the error bar on this guide led to the precision at the right side of the threshold, expressed in CU (vertical dashed line). Finally, the total TCI of LV thresholds was calculated as the distance between the two dashed lines at zero-amplitude. Importantly, the TCIs are not the true errors of the eCAP thresholds, but a best estimate of their precision.

Analysis

When investigating the predictive value of eCAP thresholds for fitting behavioral levels, thresholds are frequently processed according to the method of Brown et al. (2000), matching the eCAP threshold profile with the T-profile using the threshold data at one electrode contact. This method overcomes the offset difference between the objective and subjective profiles induced by the effect that a small stimulus may be audible though the eCAP is still too small to measure. This method was applied in our study as well. It is important to realize that the study of Brown included Nucleus® patients in which the currents are presented along a logarithmic scale, while the HiRes90K™ as used in the present study applies linear current units. From a mathematical point of view, a scaling factor was required to replicate the shift as described by Brown. In addition, the scaling was also required to take the pulse width into account when comparing behavioral thresholds (measured in µA) to the objective thresholds (measured in CU). The profiles were matched on the electrode providing the least square error between the two profiles.

Exploration of the TCI values estimated with the LE and LV methods revealed non-normally distributed TCI values and unequal variances for both methods. Therefore, the difference in TCI size between the two methods was investigated using Student's t-tests for unequal variances in combination with a logarithmic transformation of the TCI values. Assuming that within the LE and LV methods the TCIs are equally distributed across the electrode array, we applied a linear mixed model to test the effect of electrode contact on the TCI size for each method separately. The model included electrode as fixed effect and subject as random effect. Straightforward correlation analysis was used to compare the eCAP thresholds obtained by the LE and LV methods.

In addition, the association between eCAP thresholds for behavioral T-profiles was examined using a linear regression model. Pearson's correlation coefficient was calculated within individual subjects as well as for all thresholds together (group analysis). Fisher's z-transformation was used to

summarize the individual correlation coefficients and to enable further analyses on the correlation coefficients.

RESULTS

Figure 2 shows four illustrative examples of AGFs to demonstrate the effect of AGF morphology on the eCAP threshold and TCI. In panel A, the AGF measured at electrode 7 in patient S124 is shown. The AGF contains clear and large eCAPs, leading to a clear distinction of measurements containing an eCAP. Consequently, the TCI is small for both LE and LV methods, and the thresholds of both methods are close to each other. In Figure 2B, an AGF with much smaller eCAPs is shown. Because the eCAPs are closer to the noise level, it is more difficult to distinguish eCAPs from noise. Due to the lower SNR, the precision of eCAP amplitudes is less, resulting in a less reliable fit for the LE method and larger TCIs for both LE and LV methods. The examples in Figure 2A and 2C clearly show that an AGF is not necessarily linear; a shoulder near the noise level or a roll over at the top of the AGF was observed frequently. When data points at the top of the AGF clearly deviated from the linear amplitude growth, they were excluded from the linear fit (C), considerably improving the precision of the eCAP threshold. The AGF in panel D illustrates the effect of using a weighted linear fit (red) instead of the default normal fit (blue). The weighted fit took into account the measurement error of the eCAP amplitude. Consequently, the fit best reflects the most precise data points. In this example, weighting the measurement errors in the fit did not markedly affect the eCAP threshold itself, but the TCI became smaller. Finally, all AGFs were classified on eCAP amplitude to indicate how many AGFs will be represented by the examples shown. It turned out that 11% of the AGFs have an amplitude larger than 900 μV (A), 62% between 300 μV and 900 μV (C), 23% between 150 μV and 300 μV (D) and 5% smaller than 150 μV (B).

The boxplots in Figure 3 show the absolute TCIs grouped per electrode contact for the LE (A) and LV (B) methods. For each AGF, the absolute size of the confidence interval with respect to the

corresponding eCAP threshold was calculated separately for the left and right side. The intervals of all subjects were grouped per electrode contact and they are shown in boxplots. For visibility, the limits of the current-axis were set to -750 and 250 CU. Consequently, for the LE method thirteen outliers fell outside the axes limits of the figure: eleven on the left side (< -750 CU) and two on the right side (> 250 CU). For the LV method, all data points were within the axes limits. Comparing the TCIs of both methods revealed that the TCIs of the LE method were significantly larger at the contacts 3-15 ($p < 0.005$ for contact 5, $p < 0.001$ for all other contacts), while no difference in TCI size was found for the most apical contact 1 ($p = 0.28$). Further analysis of the effect of electrode contact on the TCI size revealed that for both the LE and LV methods the TCI is significantly larger ($p < 0.001$) at the base than at the apex. Across the electrode array, larger eCAP amplitudes were observed at the apex than at the base.

In Figure 4A, the absolute difference between the LE and LV thresholds was plotted as a function of TCI size for the LE method. The LE thresholds are divided into four groups, each containing 25% of the sorted TCIs. The figure shows that the LV thresholds are larger than the corresponding LE thresholds and that the difference between the LV and LE threshold tends to increase with increasing TCI for the LE method. In addition, for each category, the LE thresholds were plotted against the LV thresholds using the same colors (Figure 4B). Significant correlations were found between the LE and LV thresholds for all groups ($p < 0.001$). The group with the most reliable LE threshold correlated best with the LV thresholds ($r = 0.91$ versus $r = 0.71$ for the most extreme pairs).

The scatterplots in Figure 5A and 5B show the relationship between eCAP thresholds and behavioral thresholds for the LE and the LV methods, respectively. Only subjects who had a complete eCAP threshold profile were included in the analysis. This resulted in 48 subjects for the LE method and 50 for the LV method, whereby only 5 subjects had an 1J array and the others a Mid-Scalar. The thresholds are plotted in grey and the correlations within individual subjects are

illustrated by the black regression lines. Below the scatterplots, accompanying individual correlation coefficients are presented rank-ordered from low to high (panel C and D). The coefficients vary from -0.96 to 0.82 for the LE thresholds and from -0.81 to 0.92 for the LV thresholds. Mean individual Pearson's correlation coefficients (calculated using Fisher's z-transformation) for the LE and LV methods were r = 0.16 and r = 0.34, respectively. When analyzing all threshold values as one dataset, significant correlations were found between T-profiles and LE threshold profiles ($p < 0.001$, $r = 0.44$) and between T-profiles and LV threshold profiles ($p < 0.001$, $r = 0.54$). Additionally, Figure 6 compares the behavioral thresholds with the eCAP thresholds and their TCIs. The TCIs are visualized as in Figure 3 but rotated 90 degrees clockwise. Remember from Figure 3 that the (normalized) eCAP thresholds are represented by the line at 0 CU. In panel A, the data are shown for the LE method (blue) and in panel B for the LV method (green). The behavioral thresholds (T, red) were plotted relative to the eCAP threshold profiles, showing the absolute difference between the eCAP and behavioral thresholds per electrode contact. Across all patients, the behavioral thresholds fell within the TCIs of the eCAP thresholds.

DISCUSSION

This study focuses on the precision of eCAP thresholds. In order to estimate precision, TCIs were assigned to eCAP thresholds estimated using the LE and LV method. eCAP thresholds estimated by the LE method were systematically smaller than the thresholds estimated by the LV method. The TCIs of the LE method were larger than those of the LV method, while the LE thresholds with the smallest TCIs correlated best with the LV threshold. Comparing the eCAP threshold profiles to optimally scaled T-profiles, we found that the T-profiles fell within the TCIs of the eCAP threshold profiles. This finding is one of the potential explanations why the literature provides mixed results for correlations between behavioral and objective profiles; the eCAP threshold precision is too low.

Correlation between eCAP and behavioral thresholds

The correlation between eCAP thresholds and behavioral T- and C/M-levels has frequently been investigated to determine whether eCAP can be used to assist with sound processor programming and verify questionable behavioral responses. However, the outcomes have been moderate at best (Brown et al. 2000; Hughes et al. 2000; Franck & Norton 2001; Smoorenburg et al. 2002; Franck 2002; Di Nardo et al. 2003; Polak et al. 2006; Potts et al. 2007; Holstad et al. 2009; Mittal & Panwar 2009; Botros & Psarros 2010; Muhaimeed et al. 2010). There are several possible causes of the mismatch between objective and behavioral thresholds.

First, there is an offset difference between behavioral and objective thresholds; a small stimulus may be audible though the eCAP is still too small to measure. This offset is patient-dependent and can even vary between electrode contacts, e.g., due to the electrode impedances. The difference between eCAP thresholds and behavioral levels can be reduced by matching the two profiles using the threshold data at one electrode contact (Brown et al. 2000), which was applied in this study as well. “Additionally, there are known variations, between clinics and brands, in the way T-profiles are set in clinical routine. Next to the commonly used 10% of M-level for AB recipients, there are various ways to determine the threshold levels (e.g., first hearing thresholds, counted thresholds, single-electrode stimulation, interleaved bursts). Up till now, there has been no consensus which of these approaches lead to the best subjective map and should serve as the reference for eCAP-based fitting. The actual way of determining thresholds across the various studies using eCAP-based fittings is highly underreported (de Vos et al., in press), making it difficult to determine the best methods. Further, it is well known that pulse rate and duration of the stimulus affect both eCAP and behavioral thresholds, e.g., due to temporal integration (McKay et al. 2005; Lai & Dillier 2007; McKay et al. 2013; Hughes et al. 2014). Previous research found that the relationship between eCAP and behavioral thresholds becomes stronger when associated stimulation rates were equal or close to each other ((Brown et al. 1998; Franck & Norton 2001). However, equalizing the stimulation rates of

eCAP and behavioral measurements, routinely measured with single pulses (30 Hz) and pulse trains (>500 Hz) respectively, is not ideal. Lowering the stimulation rate for behavioral measurements makes the thresholds inadequate with respect to the high-rate speech strategy, while increasing the rate of eCAP measurements has a degrading effect on the quality and amplitude of the eCAP response (Charasse et al. 2004; Hughes et al. 2014). Therefore, the correlation between eCAP and behavioral thresholds is likely affected by differences in stimulation rate. This is probably also the case in this study.

Because our study was retrospective, there were some non-ideal conditions for comparing the eCAP outcomes to behavioral data. First, there was a time lag between the intra-operative eCAP measurements and the T-profiles obtained from regular fitting sessions. We decided to use most recent T-profiles, because we assumed that these were stable and most optimal. However, because of physiologic changes within the cochlea in the first months after surgery (Hughes et al. 2001; Spivak et al. 2011), it is likely that the association between eCAP thresholds and T-profile presented in this paper is affected by the time lag between the two measurements. To evaluate this potential effect on our results, we re-analyzed the correlation using T-profiles from the first regular fitting after surgery, which were available in our SoundWave database as well. Using the Fisher r-to-z transformation, we tested whether the correlation changes if first fitting data (on average 2 months after surgery) was used instead of the most recent fitting profiles (on average 11 months after surgery). However, no significant differences were found, neither for all threshold data together (LE thresholds: $p = 0.94$, LV thresholds: $p = 0.56$) nor for the individual correlations (LE thresholds: $p = 0.79$, LV thresholds: $p = 0.50$).

Further, our eCAPs were measured using RSPOM with the forward masking artifact rejection method, because they were a part of our clinical eCAP recording protocol. In contrast, eCAPs for fitting AB recipients will likely be measured via the clinical SoundWave software which by default uses alternating polarity. This difference in artefact rejection method potentially affects the clinical

applicability of our results as some studies observed an effect of method on the eCAP (Frijns et al. 2002; Baudhuin et al. 2016). However, Hughes et al. (2016) thoroughly investigated the effect of artefact rejection on eCAP and they found no significant difference between forward masking and alternating polarity for the amplitude and threshold of the eCAPs measured in AB recipients. For Cochlear devices, they conclude that the forward masking paradigm was even advantageous over alternating polarity. Lastly, the majority of the available AGF literature is based on the Cochlear system in which the forward masking technique is the default (de Vos et al. in press). Considering these aspects, we think that the effect of these measurement settings is negligible in light of the TCI size, the main topic of the study.

Looking at the methods for calculating the correlation between objective and behavioral thresholds, only four studies perform correlation analysis on individual eCAP and T-profiles (Holstad et al. 2009; Franck 2002; Franck & Norton 2001; Potts et al. 2007) while the majority of the studies perform the analyses on grouped data (Brown et al. 2000; Cullington 2000; Smoorenburg et al. 2002; Di Nardo et al. 2003; Kiss et al. 2003; Morita et al. 2003; Kaplan-Neeman et al. 2004; Cafarelli Dees et al. 2005; King et al. 2006; Polak et al. 2006; Pedley et al. 2007; Lai et al. 2009; Mittal & Panwar 2009; Alvarez et al. 2010; Botros & Psarros 2010; Hughes & Stille 2010; Muhaimeed et al. 2010). However, group analysis does not provide any information about the correlation between profiles of individual subjects. The group correlation will be driven by the inter-patient variation, especially when the eCAP thresholds were matched to the T-profiles using one behavioral point (Brown et al. 2000; Smoorenburg et al. 2002; Willeboer & Smoorenburg 2006). Group analyses of our data revealed moderate correlation coefficients of $r = 0.44$ for T-profiles versus LE thresholds and $r = 0.54$ for T-profiles versus LV thresholds. Despite the non-ideal conditions for comparing eCAP and behavioral thresholds, these correlation coefficients were in accordance with other (large) studies in which the grouped correlation coefficient was calculated for T-profiles versus eCAP. Mittal & Panwar (2009) found $r = 0.33$ based on 90 subjects, Brown et al. (2000) found $r = 0.55$ with 44 subjects and

Muhaimeed et al. (2010) found r = 0.57 with 47 subjects. However, in order to examine the predictive value of eCAP thresholds for the fitting of individual subjects, it is preferable to compare the eCAP and behavioral thresholds for individual patients (de Vos et al. in press). Calculating the correlation within individuals resulted in correlation coefficients varying from strongly negative to strongly positive (Figure 5C and 5D). Compared to the three studies reporting within subject correlation coefficients, our results were comparable with these of Holstad et al. (2000). The study of Holstad reported coefficients ranging from -0.67 to 0.99 with a mean of 0.24. Note that these results were obtained in children using Cochlear devices. In the studies of Franck et al. (2001/2002) and Potts et al. (2007), the individual correlation coefficients ranged between 0 and +1. Together with the finding that the behavioral threshold levels fell within the (relative large) TCIs of the eCAP thresholds (Figure 6), these results indicate that the eCAP thresholds are possibly not precise enough to predict T-profiles at the level of individual subjects.

eCAP Threshold Precision

To the best of our knowledge, no study on eCAP thresholds has included error analysis for the eCAP amplitude. The data points of AGFs are commonly processed without error bar, and the estimated eCAP threshold is assumed to be good enough for application. However, to ensure that measurements are providing data that actually is clinical applicable its precision should be evaluated. For the LE method, eCAP amplitudes have a measurement error that can be used to improve the reliability of the linear fit. Data points with a small measurement error will have a higher weight, contributing more to the fit than points with a larger error. The effect of including the measurement error is shown in Figure 2D. Here, the TCI becomes smaller when a weighted linear fit is used. To investigate how a weighted fit influences the eCAP threshold and its TCI on group level, a correlation analysis was performed on both thresholds and TCIs estimated by LE and weighted LE. This post-hoc analysis revealed no significant difference between both LE and weighted LE

thresholds ($p < 0.001$, $r = 0.99$) and between corresponding TCIs ($p < 0.001$, $r = 0.97$). Nevertheless, including the measurement error increases the reliability of the eCAP threshold and TCI, because the weighted linear fit better reflects the measured data points.

In addition, the goodness of fit (GOF) could be evaluated to estimate the reliability of the LE threshold. A worse fit would lead to a less reliable eCAP threshold and vice versa. Concerning the GOF, the question arose whether the linear fit is ideal, because we recurrently observed non-linear AGFs (e.g. Figure 2C). Therefore, we are working on a follow-up-study whereby we investigate the shape of the AGF and the GOF of several mathematical functions, e.g., a sigmoid fit (Ramekers et al. 2014). Because the GOF is an indicator for the error of the linear fit (expressed as R^2), it does not provide a valid measure of precision for the extrapolated threshold value (expressed in CU). Therefore, the TCI was developed to estimate the eCAP threshold precision in CU. Note, that the TCI is strongly associated with the GOF ($p < 0.001$, $r = -0.79$), because TCI of the LE method is based on 95% confidence bounds which also is measure of GOF.

It would be worthwhile to (re)consider what is actually measured by the LE and LV methods. For example, what kind of information is in an AGF, and how precisely can eCAPs and their thresholds be measured? Figure 4 shows that there is a systematic difference in the eCAP thresholds estimated by the two commonly used methods, thresholds estimated by LE being systematically lower than those estimated by the LV method. This difference can be explained by the fact that LE thresholds are estimated at zero-amplitude, where LV thresholds are estimated just above noise level. Consequently, the LV thresholds are probably overestimated compared to the real eCAP threshold. Furthermore, the success rate of the LV method is higher than that of LE. To estimate the eCAP threshold, one eCAP response is sufficient for the LV method, while at least two responses are required for the LE method. This explains why 48 subjects were included in Figure 5A compared to 50 in 5B. To estimate threshold precision, one additional data point at a higher current is required for the LV method. For the LE method, at least three data points are required because the TCI

cannot be computed if the number of data points is equal to the number of coefficients in the linear fit. The size of the TCI considerably reduces when 4, or even more, data points are used. The step size or applied current scale affects the TCI as well, e.g., using a logarithmic current scale (Nucleus) would result in different TCIs than a linear scale (this study). When relatively large current steps are used, the LV threshold can be estimated less precisely, probably resulting in a larger overestimation of the threshold. In this study, a default step size of 50 CU was used, which was not ideal for the LV method. Generally, the LV method is performed with smaller current steps, especially around the threshold level (Botros et al. 2007; Glassman & Hughes 2013; Baudhuin et al. 2016). The use of smaller current steps in this study would have led to lower LV thresholds, and possibly smaller TCIs. The measure of how well the LV threshold represents the real eCAP threshold also depends on the noise level. The higher the noise level, the larger the eCAP must be to be visible and, in turn, the larger the deviation from the true eCAP threshold. Thus, when a measurement system has a relatively high noise floor, the LE method provides a better approximation of the eCAP threshold than the LV method. For our data, a mean systemic noise of 31.4 μV with a standard deviation of 14.4 μV was found. Compared to other studies, this noise level is relatively high (Glassman & Hughes 2013). Therefore, performing the analysis on an AGF measured by a CI with lower systemic noise would reduce the eCAP thresholds and TCIs, especially with the LV method. Finally, note that presented method of estimating eCAP precision using TCIs is generally applicable, but the presented quantities are specific for our data. Interpretation of eCAP thresholds and their precision should always be performed in light of the measurement conditions. Furthermore, the TCI reflects a best guess of the eCAP threshold precision rather than the true threshold error. In the follow-up-study, we want to investigate the error of the eCAP threshold induced by methodology or the random noise component as well. This error could be estimated using a varying number of averages or repeated measures.

In addition to Glassman et al. (2013), who already comprehensively compared the LE and LV methods, we wanted to highlight that the relationship between these two methods depends on the threshold precision. Figure 4 shows that the correlation between LE and LV thresholds is best if the TCI is small, originating from clear, i.e., relatively large eCAPs with a low measurement error (e.g. Figure 2A). Considering the points of the most precise thresholds (blue and red points), the correlation is close to that found by Glassman et al. (2013), suggesting that the eCAPs and AGFs used by Glassman et al. in their analysis were very clear. Additionally, we found that the TCIs were smaller at the apex than at the base of the cochlea, and that the TCIs of the LE and LV methods did not differ for the most apical electrode contact (Figure 3). This can be explained by the larger eCAPs at the apex than at the base of the cochlea.

CONCLUSIONS

This study showed that eCAP thresholds estimated by LE and LV have an uncertainty that can be estimated using TCIs as measure of precision. The size of the TCI depends on several factors, e.g. the threshold estimation method and measurement conditions, but it is often larger than one would expect when just looking at the threshold values. Given the relatively large TCIs, we recommend that future research on eCAP thresholds should be accompanied by a measure of precision to correctly apply eCAP thresholds in clinical practice. Comparing our eCAP outcomes with behavioral fitting levels, we found that the T-profiles fell within the TCI of the LE and LV thresholds. Further, although our conditions for comparing eCAP and behavioral thresholds were non-ideal, our findings were in line with the literature: significant correlations between the two parameters at the level of grouped data and correlation coefficients almost homogeneously ranging from -1 to +1 for individual subjects. Therefore, not only the relative large TCIs, but also the poor individual correlations of eCAP and behavioral thresholds indicate that the eCAP thresholds are possibly not precise enough to predict T-profiles.

REFERENCES

Abbas, P.J., Brown, C.J., Shallop, J.K., et al. (1999). Summary of results using the nucleus CI24M implant to record the electrically evoked compound action potential. *Ear Hear.*, 20, 45–59.

Akhoun, I., Bestel, J., Pracht, P., et al. (2015). Automated classification of electrically-evoked compound action potentials. In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*. (*pp. 687–690*). IEEE.

Alvarez, I., de la Torre, A., Sainz, M., et al. (2010). Using evoked compound action potentials to assess activation of electrodes and predict C-levels in the Tempo+ cochlear implant speech processor. *Ear Hear.*, 31, 134–45.

Baudhuin, J.L., Hughes, M.L., Goehring, J.L. (2016). A Comparison of Alternating Polarity and Forward Masking Artifact-Reduction Methods to Resolve the Electrically Evoked Compound Action Potential. *Ear Hear.*, 1–9.

Biesheuvel, J.D., Briaire, J.J., Frijns, J.H.M. (2016). A Novel Algorithm to Derive Spread of Excitation Based on Deconvolution. *Ear Hear.*, 37, 572–81.

Botros, A., van Dijk, B., Killian, M. (2007). AutoNR: an automated system that measures ECAP thresholds with the Nucleus Freedom cochlear implant via machine intelligence. *Artif. Intell. Med.*, 40, 15–28.

Botros, A., Psarros, C. (2010). Neural response telemetry reconsidered: I. The relevance of ECAP threshold profiles and scaled profiles to cochlear implant fitting. *Ear Hear.*, 31, 367–79.

Brown, C.J., Abbas, P.J., Gantz, B.J. (1998). Preliminary experience with neural response telemetry in the nucleus CI24M cochlear implant. *Am. J. Otol.*, 19, 320–7.

Brown, C.J., Hughes, M.L., Luk, B., et al. (2000). The relationship between EAP and EABR thresholds and levels used to program the nucleus 24 speech processor: data from adults. *Ear Hear.*, 21, 151–63.

Cafarelli Dees, D., Dillier, N., Lai, W.K., et al. (2005). Normative findings of electrically evoked

compound action potential measurements uing the neural response telemetry of the nucleus CI24M cochlear implant system. *Audiol. Neurotol.*, 10, 105–116.

Charasse, B., Chanal, J.M., Berger-Vachon, C., et al. (2004). Influence of stimulus frequency on NRT recordings. *Int. J. Audiol.*, 43, 236–44.

Cullington, H. (2000). Preliminary neural response telemetry results. *Br. J. Audiol.*, 34, 131–40.

Franck, K.H. (2002). A model of a nucleus 24 cochlear implant fitting protocol based on the electrically evoked whole nerve action potential. *Ear Hear.*, 23, 67S–71S.

Franck, K.H., Norton, S.J. (2001). Estimation of psychophysical levels using the electrically evoked compound action potential measured with the neural response telemetry capabilities of Cochlear Corporation's CI24M device. *Ear Hear.*, 22, 289–99.

Frijns, J.H.M., Briaire, J.J., de Laat, J.A.P.M., et al. (2002). Initial evaluation of the Clarion CII cochlear implant: speech perception and neural response imaging. *Ear Hear.*, 23, 184–197.

Glassman, E.K., Hughes, M.L. (2013). Determining electrically evoked compound action potential thresholds: a comparison of computer versus human analysis methods. *Ear Hear.*, 34, 96–109.

Holstad, B. a, Sonneveldt, V.G., Fears, B.T., et al. (2009). Relation of electrically evoked compound action potential thresholds to behavioral T- and C-levels in children with cochlear implants. *Ear Hear.*, 30, 115–27.

Hughes, M.L. (2013). *Objective Measures in Cochlear Implants*, San Diego: Plural Publishing.

Hughes, M.L., Baudhuin, J.L., Goehring, J.L. (2014). The relation between auditory-nerve temporal responses and perceptual rate integration in cochlear implants. *Hear. Res.*, 316, 44–56.

Hughes, M.L., Brown, C.J., Abbas, P.J., et al. (2000). Comparison of EAP thresholds with MAP levels in the nucleus 24 cochlear implant: data from children. *Ear Hear.*, 21, 164–74.

Hughes, M.L., Goehring, J.L., Baudhuin, J.L. (2016). Effects of Stimulus Polarity and Artifact Reduction Method on the Electrically Evoked Compound Action Potential. *Ear Hear.*, 1.

Hughes, M.L., Stille, L.J. (2010). Effect of stimulus and recording parameters on spatial spread of

excitation and masking patterns obtained with the electrically evoked compound action potential in cochlear implants. *Ear Hear.*, 31, 679–692.

Hughes, M.L., Vander Werff, K.R., Brown, C.J., et al. (2001). A longitudinal study of electrode impedance, the electrically evoked compound action potential, and behavioral measures in nucleus 24 cochlear implant users. *Ear Hear.*, 22, 471–486.

Kaplan-Neeman, R., Henkin, Y., Yakir, Z., et al. (2004). NRT-based versus behavioral-based MAP: a comparison of parameters and speech perception in young children. *J. Basic Clin. Physiol. Pharmacol.*, 15, 57–69.

King, J.E., Polak, M., Hodges, A. V, et al. (2006). Use of neural response telemetry measures to objectively set the comfort levels in the Nucleus 24 cochlear implant. *J. Am. Acad. Audiol.*, 17, 413–31;

Kiss, J.G., Tóth, F., Nagy, A.L., et al. (2003). Neural response telemetry in cochlear implant users. *Int. Tinnitus J.*, 9, 59–60.

Lai, W.K., Dillier, N. (2007). Comparing neural response telemetry amplitude growth functions with loudness growth functions: preliminary results. *Ear Hear.*, 28, 42S–45S.

Lai, W.K., Dillier, N., Weber, B.P., et al. (2009). TNRT profiles with the nucleus research platform 8 system. *Int. J. Audiol.*, 48, 645–54.

McKay, C.M., Fewster, L., Dawson, P. (2005). A different approach to using neural response telemetry for automated cochlear implant processor programming. *Ear Hear.*, 26, 38S–44S.

McKay, C.M.C.M., Chandan, K., Akhoun, I., et al. (2013). Can ECAP Measures Be Used for Totally Objective Programming of Cochlear Implants? *J. Assoc. Res. Otolaryngol.*, 14, 879–90.

Mittal, R., Panwar, S.S. (2009). Correlation between intra-operative high rate neural response telemetry measurements and behaviourally obtained threshold and comfort levels in patients using Nucleus 24 cochlear implants. *Cochlear Implants Int.*, 10, 103–11.

Morita, T., Naito, Y., Hirai, T., et al. (2003). The relationship between the intraoperative ECAP

threshold and postoperative behavioral levels: the difference between postlingually deafened adults and prelingually deafened pediatric cochlear implant users. *Eur. Arch. Otorhinolaryngol.*, 260, 67–72.

Muhaimeed, H. Al, Anazy, F. Al, Hamed, O., et al. (2010). Correlation between NRT measurement level and behavioral levels in pediatrics cochlear implant patients. *Int. J. Pediatr. Otorhinolaryngol.*, 74, 356–60.

Di Nardo, W., Ippolito, S., Quaranta, N., et al. (2003). Correlation between NRT measurement and behavioural levels in patients with the Nucleus 24 cochlear implant. *Acta Otorhinolaryngol. Ital.*, 23, 352–5.

Pedley, K., Psarros, C., Gardner-Berry, K., et al. (2007). Evaluation of NRT and behavioral measures for MAPping elderly cochlear implant users. *Int. J. Audiol.*, 46, 254–62.

Plant, K., Law, M.-A., Whitford, L., et al. (2005). Evaluation of streamlined programming procedures for the Nucleus cochlear implant with the Contour electrode array. *Ear Hear.*, 26, 651–68.

Polak, M., Hodges, A. V, King, J.E., et al. (2006). Objective methods in postlingually and prelingually deafened adults for programming cochlear implants: ESR and NRT. *Cochlear Implants Int.*, 7, 125–41.

Potts, L.G., Skinner, M.W., Gotter, B.D., et al. (2007). Relation between neural response telemetry thresholds, T- and C-levels, and loudness judgments in 12 adult nucleus 24 cochlear implant recipients. *Ear Hear.*, 28, 495–511.

Ramekers, D., Versnel, H., Strahl, S.B., et al. (2014). Auditory-nerve responses to varied inter-phase gap and phase duration of the electric pulse stimulus as predictors for neuronal degeneration. *J. Assoc. Res. Otolaryngol.*, 15, 187–202.

Smoorenburg, G.F., Willeboer, C., van Dijk, J.E. (2002). Speech perception in nucleus CI24M cochlear implant users with processor settings based on electrically evoked compound action potential thresholds. *Audiol. Neurootol.*, 7, 335–47.

Spivak, L., Auerbach, C., Vambutas, A., et al. (2011). Electrical compound action potentials recorded with automated neural response telemetry: threshold changes as a function of time and electrode position. *Ear Hear.*, 32, 104–13.

de Vos, J.J., Biesheuvel, J.D., Briaire, Jeroen, J., et al. Use of Electrically Evoked Compound Action Potentials for Cochlear Implant Fitting: A Systematic Review. *Ear Hear.*

Willeboer, C., Smoorenburg, G.F. (2006). Comparing cochlear implant users' speech performance with processor fittings based on conventionally determined T and C levels or on compound action potential thresholds and live-voice speech in a prospective balanced crossover study. *Ear Hear.*, 27, 789–98.

FIGURE AND TABLE LEGENDS

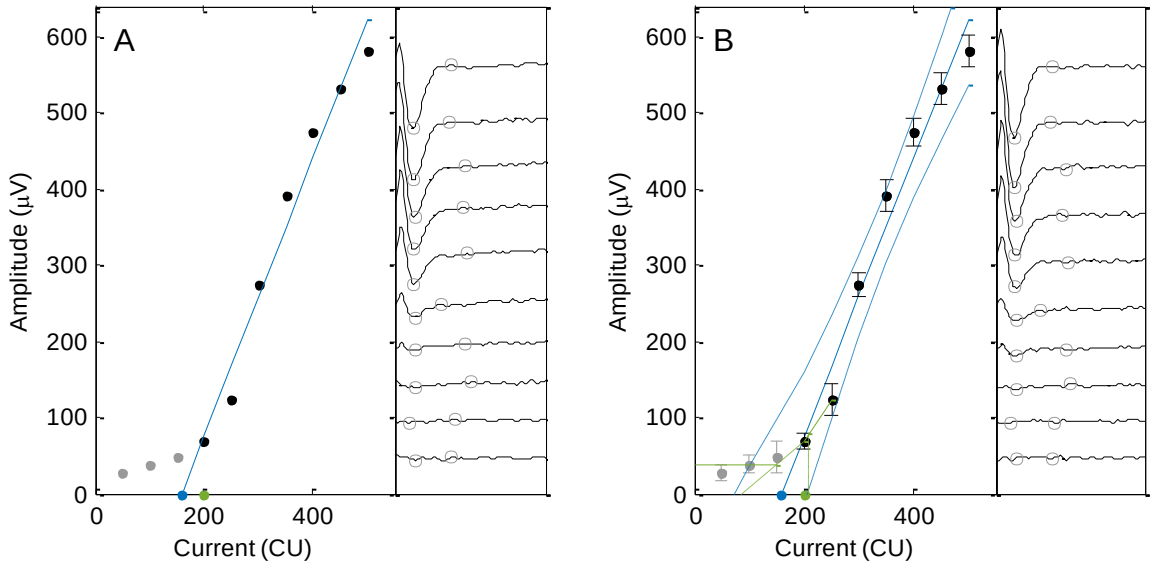

Figure 1. Example of an AGF shown as commonly used (A), and with additional information about the precision of eCAP and eCAP threshold (B). The AGF shows the eCAP amplitude as function of stimulus level, whereby corresponding eCAPs are shown at the right side, plotted from high (top) to low (bottom) stimulus level. Data points not representing true eCAPs are shown in grey and points reflecting eCAP responses (black) were used to estimate the eCAP thresholds for the LE (blue dot) and LV (green dot) methods. In (B), error bars are added reflecting the variation in eCAP amplitude. The guides were used to estimate the precision of the LE (blue) and LV (green) thresholds. AGF indicates amplitude growth function; eCAP, electrically evoked compound action potential; LE, linear extrapolation; LV, last visible.

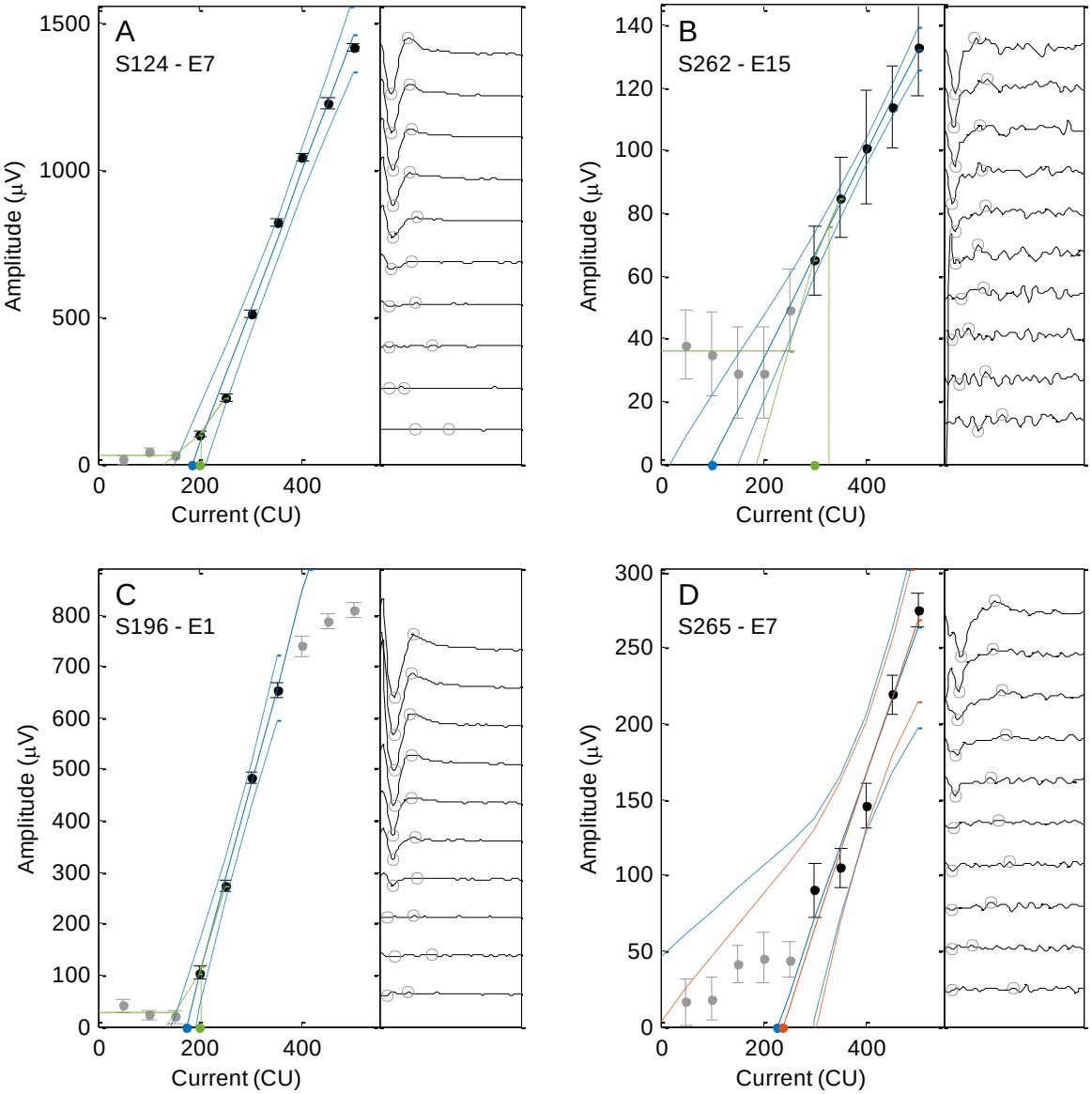

Figure 2. Four examples of AGFs, illustrating different curve morphologies and their consequences for eCAP threshold and precision. The AGFs were plotted in the same way as in Figure 1. In addition, the LE method based on a weighted linear fit was drawn in panel D (red), while the LV threshold was omitted for the sake of visibility. AGF indicates amplitude growth function; LE, linear extrapolation; LV, last visible.

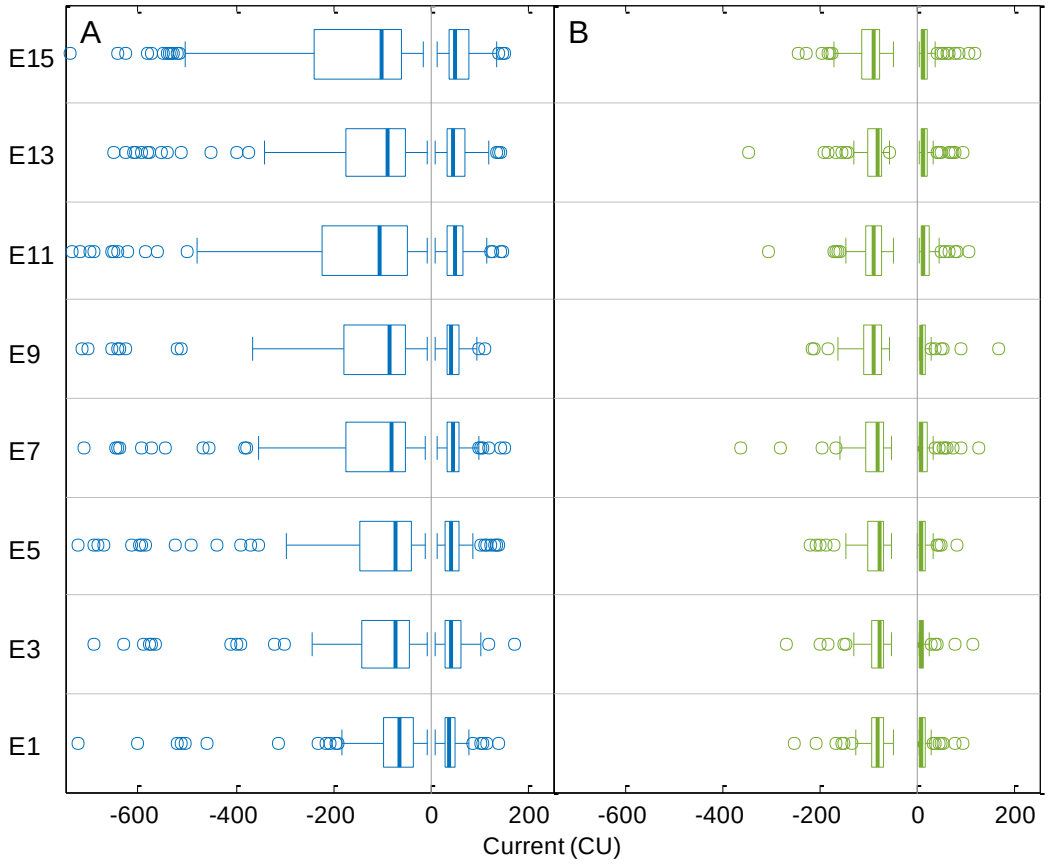

Figure 3. Illustration of the absolute TCI size (expressed in CU) across all subjects. The TCIs are grouped per electrode contact for the LE (A, blue) and LV (B, green) methods separately. Box boundaries represent the 25th and 75th percentiles, whiskers represent the most extreme data points not considered outliers, open circles represent outliers, and solid line within the box represent median. TCI indicates threshold confidence interval; LE, linear extrapolation; LV, last visible.

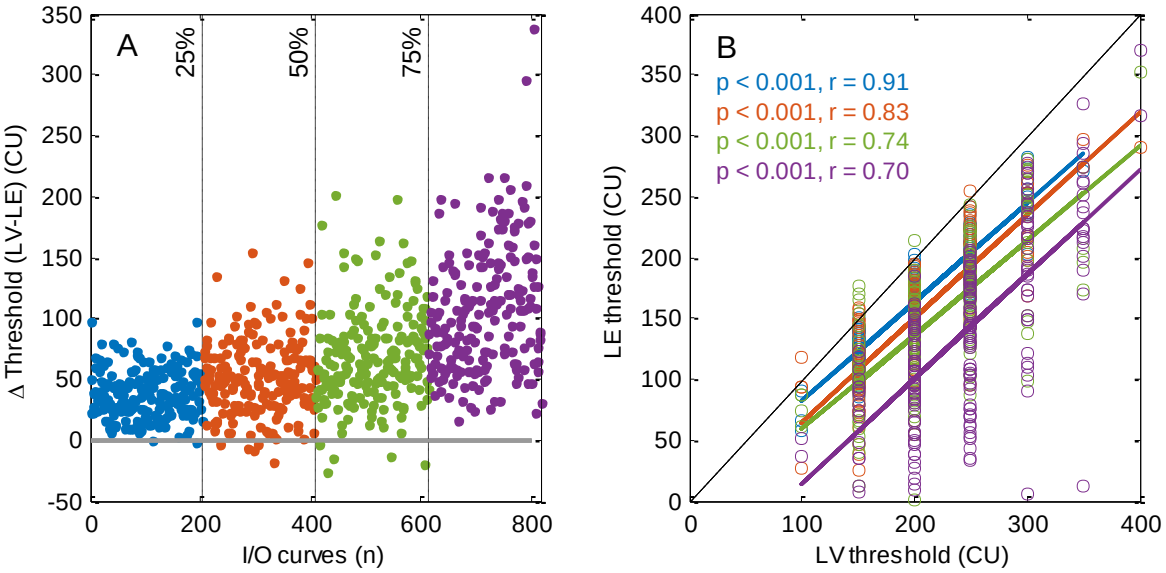

Figure 4. Difference between thresholds estimated using the LE and LV methods. The LE thresholds are sort from small to large TCI and divided in four groups, each group containing 25% of the AGFs (blue, red, green, purple). The left plot shows the absolute difference between LV and LE thresholds for each group separately. In the right panel a scatterplot between LV and LE thresholds is shown for the same groups. TCI indicates threshold confidence interval; LE, linear extrapolation; LV, last visible; AGF, amplitude growth function.

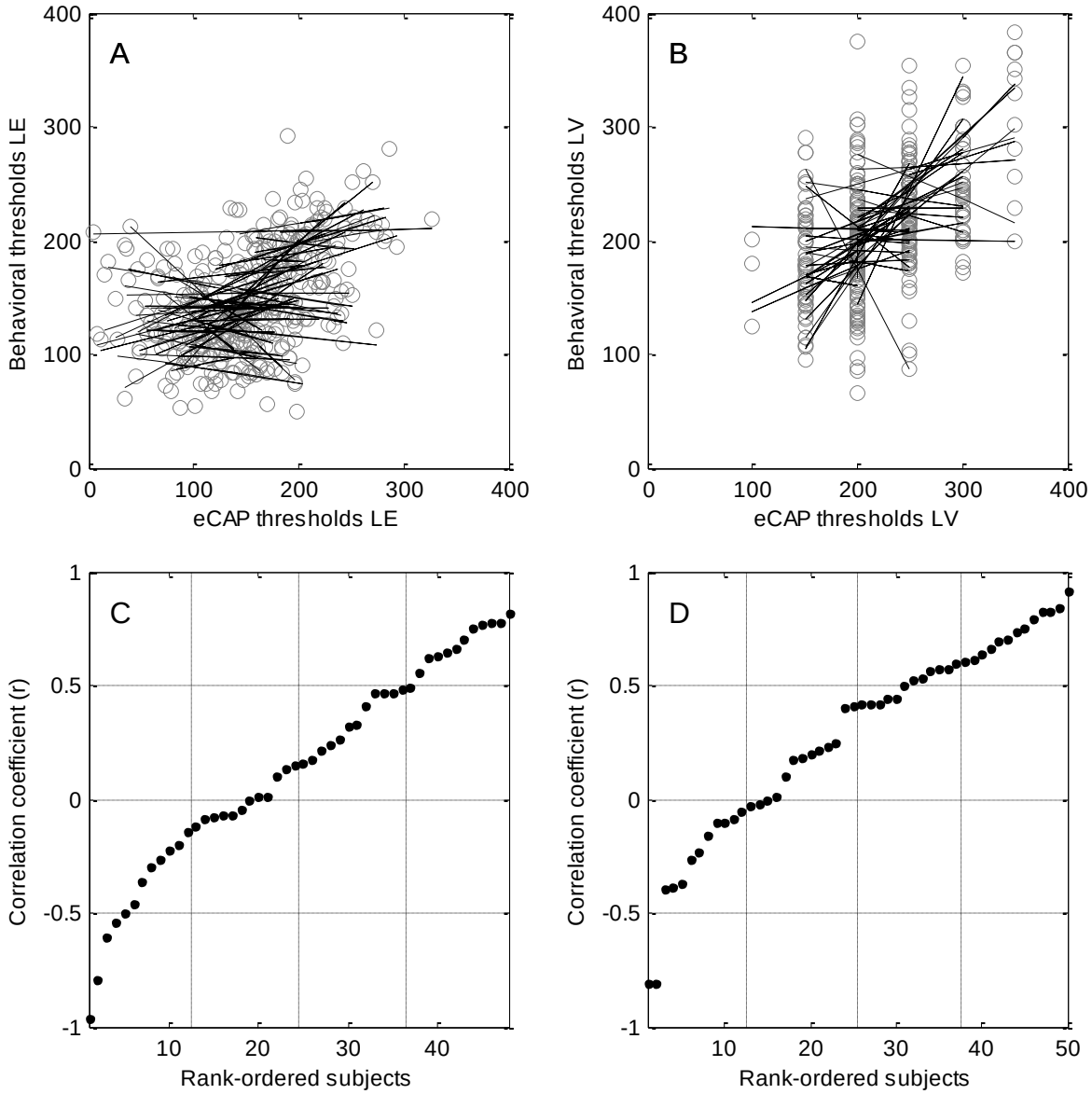

Figure 5. Scatterplots showing the correlation between eCAP thresholds (x-axis) and behavioral thresholds (y-axis) for the LE method (A) and the LV method (B). The thresholds are plotted in grey and the correlations within individual subjects are depicted by the black lines. Below the scatterplots (panel C and D), accompanying Pearson's correlation coefficients were presented rank-ordered from low to high (black line). eCAP indicates electrically evoked compound action potential, LE, linear extrapolation; LV, last visible.

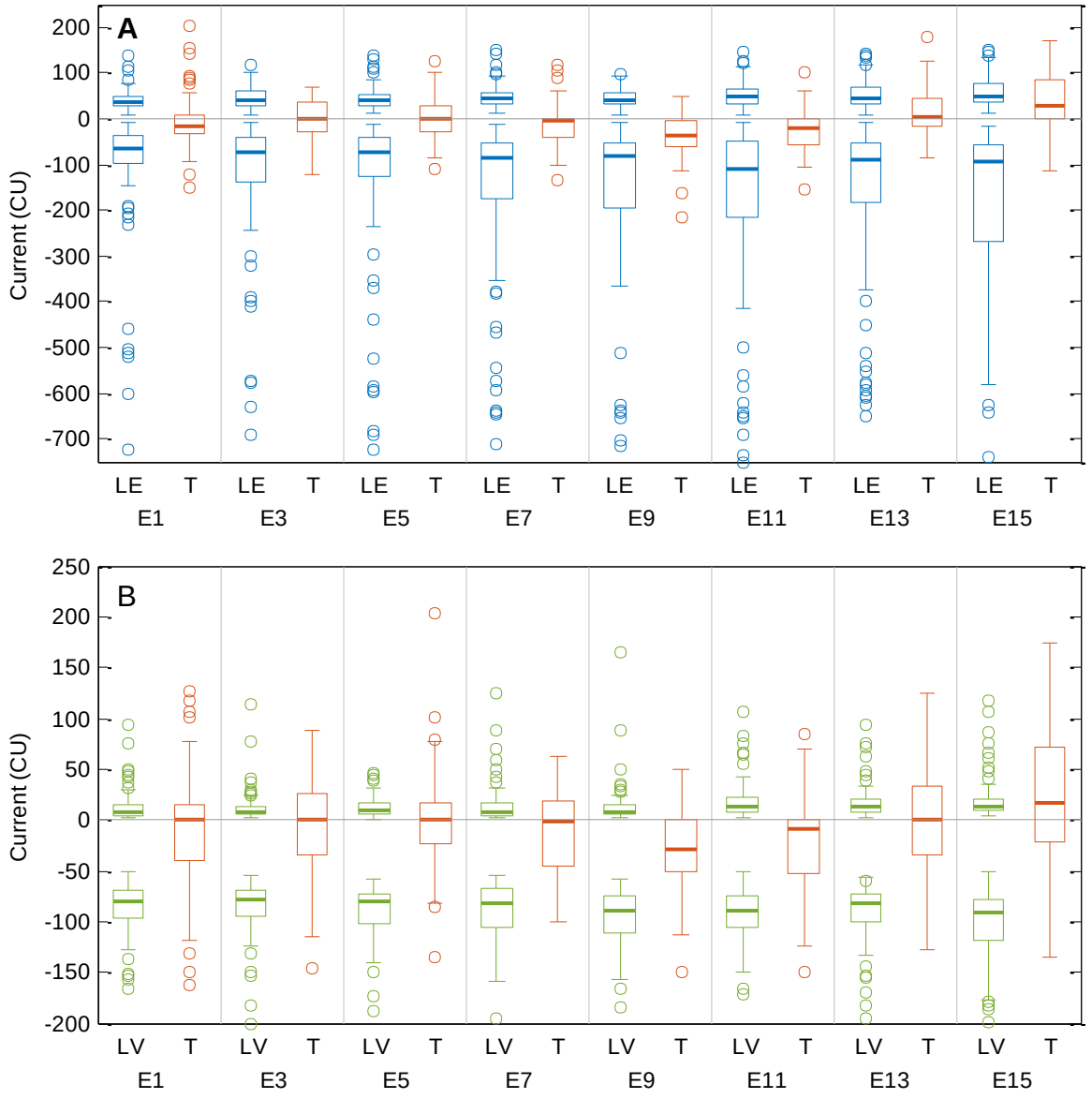

Figure 6. Comparison of the behavioral threshold profiles (T, red) with the objective eCAP thresholds and TCIs obtained with the LE (part A, blue) and LV (part B, green) methods. The eCAP thresholds and TCIs are presented as in Figure 3, whereby the (normalized) eCAP thresholds are represented by the line at 0 CU. The threshold values are grouped per electrode (horizontal) and expressed in CU (vertical). Box boundaries represent the 25th and 75th percentiles, whiskers represent the most extreme data points not considered outliers, open circles represent outliers, and solid line within the box represent median. eCAP indicates electrically evoked compound action potential, LE, linear extrapolation; LV, last visible.

Number of patients	111
Gender (n)	
Male	45
Female	66
Implant type (n)	
HiRes90K 1J	16
HiRes90K Mid-Scala	95
Age (years)	
Children (< 12)	38
Adults	73
Mean ± SD	39 ± 30
Etiology (n)	
Medication	1
Meniere	0
Meningitis	5
Otosclerosis	3
Trauma	3
Usher	3
Rubella	3
Congenital/Hereditary	37
Enlarged vestibular aqueduct	1
Cytomegalovirus	4
Other/Unknown	51

Table 1. Patient demographics.