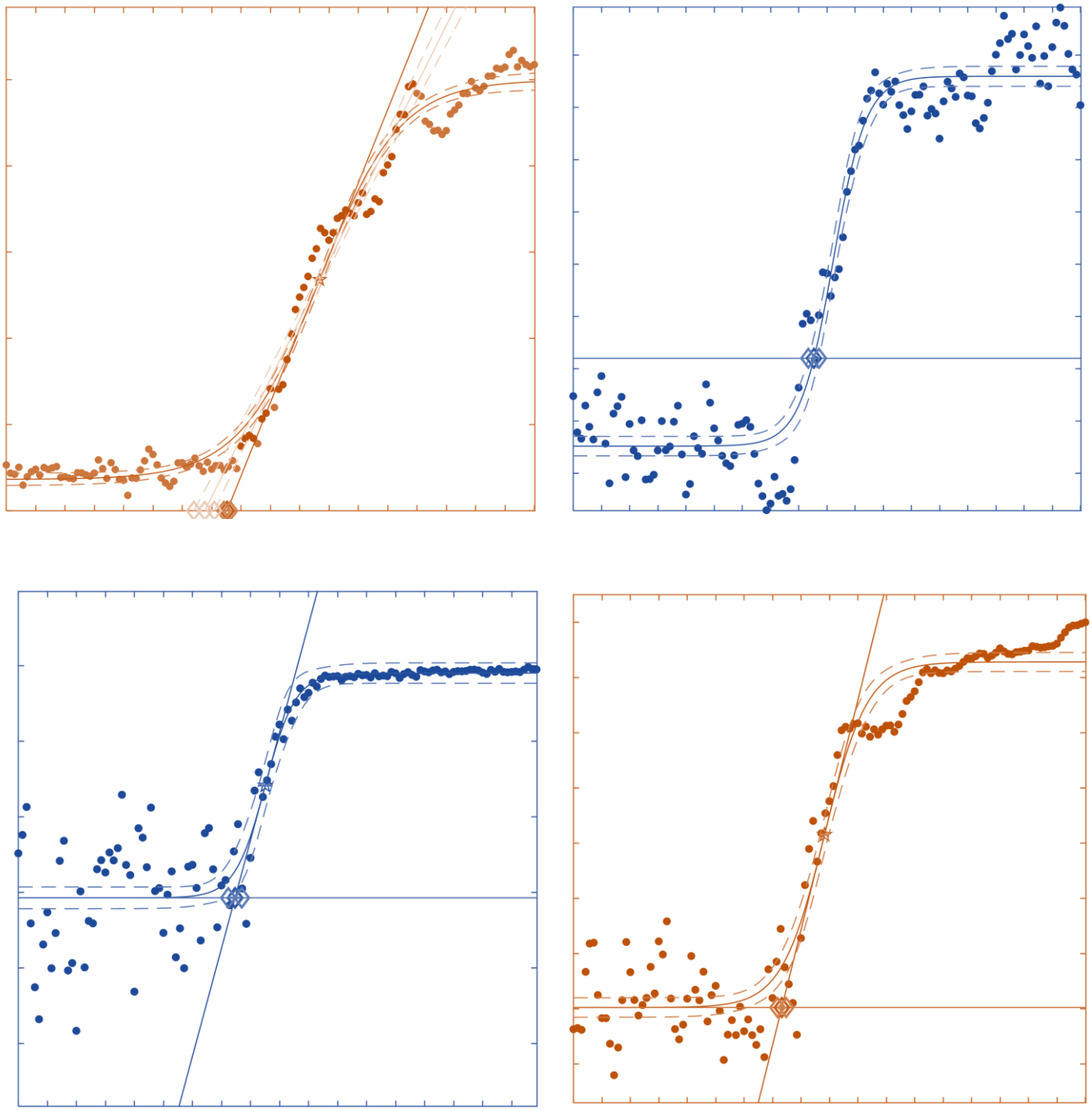
Automatic detection of eCAP thresholds

- Precision and accuracy of different methods -



Eleen Schupp

Master thesis Technical Medicine

April 2023

This page was intentionally left blank.

AUTOMATIC DETECTION OF eCAP THRESHOLDS

- Precision and accuracy of different methods -

Schupp, Eleen

Student number: 4550544

April 2023

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University ; Delft University of Technology ; Erasmus University Rotterdam

Master thesis project (TM30004 ; 35 ECTS)

Dept. of Biomechanical Engineering, TUDELFT

September 2022 – May 2023

Feasibility study (15 ECTS): February 2022 - May 2022

Supervisor(s):

dr. ir. Jeroen Briaire

dr. Dick Biesheuvel

Thesis committee members:

dr. ir. Richard Hendriks, TU Delft (chair)

dr. ir. Jeroen Briaire, LUMC

dr. Dick Biesheuvel, LUMC

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Universiteit Leiden

TU Delft Delft University of Technology

ERASMUS UNIVERSITEIT ROTTERDAM

Preface

It is with great pleasure that I present this thesis report for the Master Technical Medicine, titled *Automatic detection of eCAP thresholds: precision and accuracy of different methods*. The journey of this Master thesis started in March 2022, and it was a journey in which I got to experience the entire process of conducting clinical research.

The start of the project was to gain approval of the medical ethics committee (METC). I learned which documents needed to be composed to obtain their approval and how clinical research can be safely conducted with self-designed software. I especially learned that things do not always go as planned when conducting clinical research, but with the required patience and perseverance, you will get there. Next, I conducted a literature review into different automatic eCAP threshold detection methods. During the final part of this project, I performed eCAP measurements in CI users and analysed these measurements by using different automatic threshold detection methods and compared these threshold detection methods based on their precision and accuracy.

First, I would like to thank my supervisors: Jeroen Briaire and Dick Biesheuvel. They supervised me during the whole project, taught me the ins and outs of eCAP measurements and helped me when I got stuck and could no longer see the end goal. Besides, I would also like to thank Richard Hendriks, for being part of me graduation committee.

From the beginning of this project, Berber Mol, the research coordinator of the ENT-department at the LUMC, helped me with all the administrative aspects of this study and drawing up the necessary documents for the METC approval. The METC approval took longer than expected, but we would have never gotten the approval if it wasn't for the help of Berber. Therefore, I would like to express my gratitude to Berber.

In addition, I would like to thank the other (PhD-) students and colleagues at the ENT-department of the LUMC, for providing me with new insights and dragging me away from my computer (or at least trying to) from time to time.

Furthermore, I want to thank my friends and family for their continuous support and for listening to my complaints or worries, even though they did not always completely understand what I was talking about.

Last but not least, I would like to thank all the CI users who participated in my study. Without them, this study would not have been possible. I hope this study and all future studies concerning CIs can contribute to the further improvement of CIs.

I hope you enjoy reading this Master thesis.

Eleen Schupp
Leiden, April 2023

Summary

When a person suffers from severe to profound hearing loss, a cochlear implant (CI) can aid in restoring auditory perception and speech comprehension. To obtain good speech comprehension, fitting of a CI to the user's specific characteristics is crucial. Fitting can be a time-consuming process which requires the cooperation of a CI user and is dependent on the used methods (e.g., T-level measurements). Besides transmitting electrical stimuli, a CI can also record the response of the auditory nerve fibres to a stimulus. This response is called the electrically evoked compound action potential (eCAP). eCAP responses can be measured objectively without a user's active cooperation and could potentially aid in the fitting of a CI. For this purpose, the eCAP thresholds are of main interest. eCAP thresholds can be determined manually by a clinical specialist, or automatically by an automatic threshold detection method. Automatic eCAP threshold detection can therefore be of aid in a completely objective and uniform CI fitting method. The goal of this study was to compare different automatic eCAP threshold detection methods (in combination with different averaging methods and different artefact reduction methods) based on the precision and accuracy of these methods.

Five different automatic eCAP threshold detection methods have been examined in this study: sigmoid amplitude growth function (AGF), linear AGF, signal-to-noise ratio (SNR), cross-covariance between adjacent levels and cross-covariance with maximum level. The two different averaging methods that have been examined are standard averaging (SA) and FineGrain averaging (FG), the two different artefact reduction methods are alternating polarity (AP) and forward masking (FM). In total, 20 different combinations have been examined. The success rates of these 20 combinations have been determined, threshold confidence intervals (TCIs) were calculated as a measure of precision and the correlations between eCAP thresholds and T-levels were determined as a measure of accuracy of the different (combinations of) methods.

The combination of FG FM resulted in the highest success rates for different threshold detection methods, and the threshold detection method SNR had the overall highest success rates. A two-way ANOVA revealed that both artefact reduction/averaging method and threshold detection method have a significant effect on the TCIs. The combination of FG FM had the best results regarding the TCIs, and the sigmoid AGF threshold detection method was the threshold detection method with the lowest mean TCI. A similar two-way ANOVA was performed for the correlation between eCAP thresholds and T-levels, revealing the same results as for the TCIs that both artefact reduction/averaging method and threshold detection method have a significant effect on the correlation coefficients. FG FM was again the best performing combination, and the sigmoid AGF threshold detection method resulted in the highest correlation coefficients.

Based on these results, it can be stated that the FG FM combination for averaging and artefact reduction was the overall best combination. When comparing the different automatic threshold detection methods, the sigmoid AGF method resulted in eCAP thresholds with the highest precision and accuracy. Future research should focus on obtaining more data, further refinements of the different automatic eCAP threshold detection methods and the use of the determined eCAP thresholds in the clinical fitting of a CI.

List of abbreviations

Abbreviation	Definition
AB	Advanced Bionics
AGF	Amplitude growth function
ANOVA	Analysis of variance
AP	Alternating polarity
ART	Auditory Response Telemetry
BEDCS	Bionic Ear Data Collection System
CI	Cochlear implant
eCAP	Electrically evoked compound action potential
FG	FineGrain averaging
FM	Forward masking
NRI	Neural Response Imaging
NRT	Neural Response Telemetry
SA	Standard averaging
SNR	Signal-to-noise ratio
SW	SoundWave
TCI	Threshold confidence interval

List of figures

List of tables

Contents

1. Introduction

1.1. Cochlear implants

In normal hearing, soundwaves enter the ear through the ear canal leading to the tympanic membrane. The incoming soundwaves make the eardrum vibrate, and these vibrations are propagated through the auditory ossicles up until the oval window, which forms the separation between the ossicles and the cochlea. The vibrations of the oval window set the fluid inside the cochlea in motion. Inside the cochlea, the stereocilia on the inner hear cells will move along with the motion of the fluid. This movement of the stereocilia leads to action potentials, these action potentials are propagated along the auditory nerve to the brain which can in turn lead to auditory perception and speech comprehension (1).

When a person suffers from sensorineural hearing loss, this is often caused by damage or malfunctioning of the cochlea or inner hear cells. Persons with this type of hearing loss can benefit from cochlear implants.

A cochlear implant (CI) is an auditory prosthesis used to restore hearing in people with severe to profound hearing loss. A CI consists of three main outer parts. A microphone to capture the sound, a speech processor to transform the soundwaves of this captured sound into electrical signals and an external coil or transmitter to transmit the electrical signals to the internal coil or receiver of the CI. Internally, the received signals at the coil are sent to the cochlea via a multiple-channel electrode array. Inside the cochlea, these electrical signals are picked up by the auditory nerve fibres which send them to the brain and can lead to auditory perception and speech comprehension as in normal hearing. The different parts of a CI are displayed in Figure 1.1.

In order to obtain good speech comprehension, fitting of a CI to a patient's personal characteristics is crucial. The most important fitting parameters are the minimum and maximum stimulus level per electrode contact. There is currently no uniform fitting method, which could lead to variances in fitting depending on the clinical specialist performing the fitting. Furthermore, fitting is often a time-consuming process which requires the active cooperation of a patient (2). Fitting of a CI is now often based on behavioural measurements. The behavioural threshold, the lowest stimulus level for which a patient indicates it is audible, is called the T-level. The maximum level that is still comfortable for a patient is called the M-level or C-level. It would be beneficial if a CI could be fitted based on objective measurements, this could lead to more uniform fittings across clinical specialists, and it can be performed without a patient's active cooperation (which would be especially useful in for example paediatric patients).

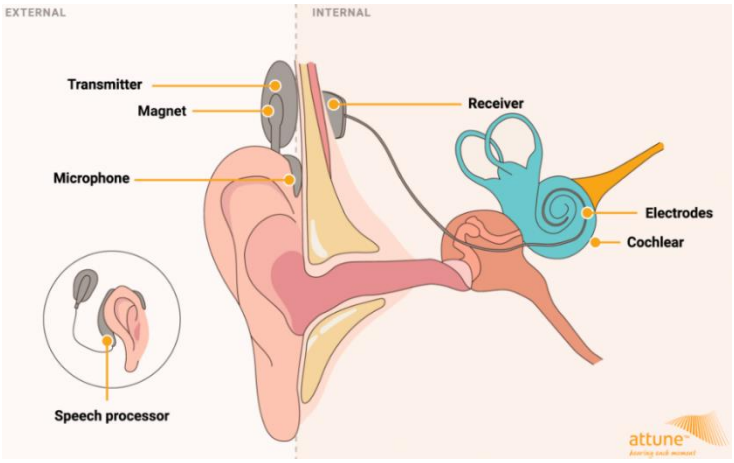


Figure 1.1: Parts of a cochlear implant. Image courtesy of Attune.

1.2. Electrically evoked compound action potential

Besides transmitting electrical stimuli, a CI can also record the response of the auditory nerve fibres to a stimulus via its telemetry function. This response is called the electrically evoked compound action potential (eCAP), which represents the neural response from multiple auditory nerve fibres to an electrical stimulus.

An eCAP occurs within roughly one millisecond after the stimulation pulse (latency) and typically contains a negative peak N1 and a positive peak P1. The difference between those two peaks is the eCAP amplitude, which varies between 50 and 1500 µV on average. An example of a simplified eCAP response is shown in Figure 1.2. Besides the eCAP response, a recording also contains signal noise, recording artefacts and stimulus artefacts (3).

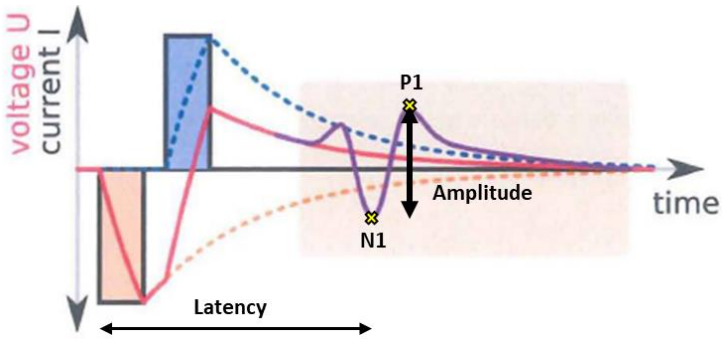


Figure 1.2: Properties of the electrically evoked compound action potential (eCAP). The rectangles represent the stimulus, the purple line represents the eCAP response and the red line represents the artefact of decaying charge. Image courtesy of MED-EL.

Signal noise is considered to be random and can therefore be reduced by averaging. In standard averaging (SA), multiple measurements are made at the same stimulus level and the resulting eCAP response is the average over all these measurements. As multiple measurements prolong the measuring time, measurements are usually restricted to a limited number of stimulus levels with relatively bigger step sizes. An alternative for SA is FineGrain (FG) averaging. In FG averaging, measurements are performed only once at each stimulus level, but more stimulus levels are measured with smaller step sizes. Averaging is then performed over the measurements within a small range of stimulus levels. As a result, during the same measuring time, eCAP measurements can be made at more stimulus levels with a smaller step size. Consequently, this leads to a higher resolution of the amplitude growth function (AGF, see 2.3.2. Amplitude growth function threshold detection), which in turn can improve the accuracy of the threshold determination (4). In Figure 1.3, the difference between SA and FG is displayed.

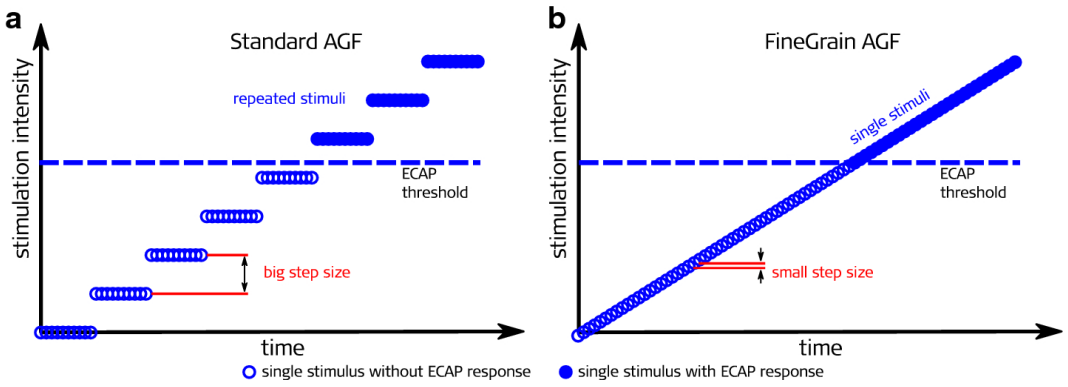


Figure 1.3: Illustrated difference between standard averaging (**a**) and FineGrain averaging (**b**). AGF = amplitude growth function, eCAP = electrically evoked compound action potential. Figure from Gärtner et al. (5).

Recording artefacts in eCAP measurements are induced by switching on the recording and can be eliminated by template subtraction. For this purpose, a recording with a zero-amplitude stimulus level is recorded, which will only contain this recording artefact and no eCAP response, and is subtracted from the other eCAP measurements.

Stimulus artefacts are caused by the tissue which acts as a capacitor. So, after the stimulus there is an artefact of decaying charge. This artefact is shown in Figure 1.2. To reduce stimulus artefacts, different artefact reduction methods exist (6, 7). The most common methods are alternating polarity (AP) and forward masking (FM). In AP, an anodic-leading and a cathodic leading frame are measured and the average over these two frames is the resulting eCAP. Figure 1.4 is a graphical representation of the AP method. There are differences in eCAP response between anodic and cathodic leading stimuli, such as the latency of an eCAP response. Averaging over these two different polarity frames can therefore result in dispersion of the eCAP response. Another stimulus artefact reduction method is forward masking. To perform forward masking, four different frames are measured:

A. Probe stimulus
B. Masker stimulus + probe stimulus
C. Masker stimulus

In frame B, the masker stimulus will set the auditory nerve fibres in a refractory period whereby the probe stimulus will not result in an eCAP response. The resulting eCAP is constructed by A-B+C. Figure 1.5 is a graphical representation of the FM method. Previous studies have demonstrated significant differences in eCAP responses between different artefact reduction methods. There is no clear consensus on which artefact reduction method results in the most precise and accurate eCAP thresholds (3, 6-10).

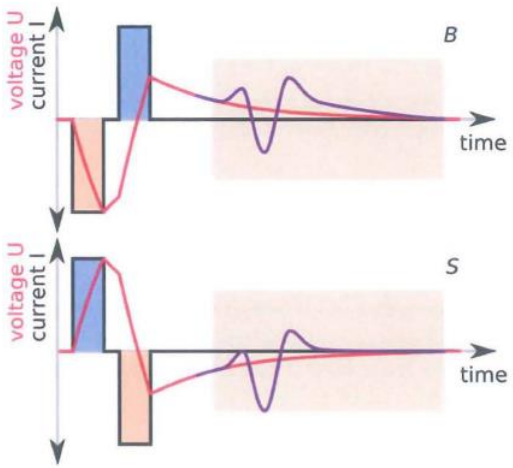


Figure 1.4: Alternating polarity method. Frame B is the cathodic leading frame, frame S is the anodic leading frame. The rectangles represent the stimulus, the purple line represents the eCAP response and the red line represents the artefact of decaying charge. Image courtesy of MED-EL.

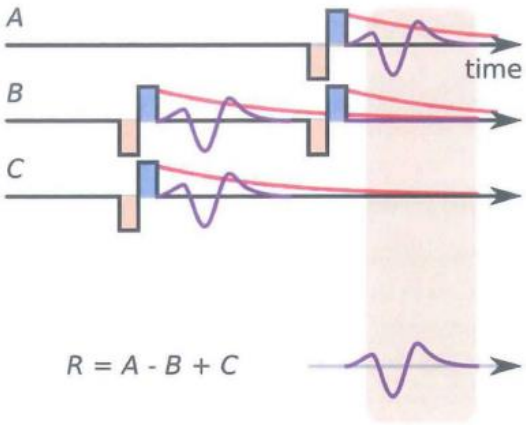


Figure 1.5: Forward masking method. Frame A is the probe frame, frame B is the masker+probe frame, frame C is the masker frame. The rectangles represent the stimulus, the purple line represents the eCAP response and the red line represents the artefact of decaying charge. Image courtesy of MED-EL.

eCAP measurements can be used intra-operatively to ensure a good placement of the electrode, and could post-operatively be used for, among other things, monitoring and fitting (7).

An important advantage of eCAP measurements which make them beneficial for use in clinical practice, is that no extra equipment is involved (apart from the components of the CI that are already accessible). For other types of measurements, such as the auditory brainstem response for example, extra external electrodes are often necessary. eCAP measurements can be performed for all types of CIs, but each manufacturer has its own terminology: it is called Neural Response Imaging (NRI) by Advanced Bionics (AB) (Valencia, CA, USA), Neural Response Telemetry (NRT) by Cochlear (Sydney, Australia) and Auditory Response Telemetry (ART) by MED-EL (Innsbruck, Austria).

1.3. eCAP thresholds

From eCAP measurements, the eCAP threshold can be estimated. The eCAP threshold is defined as the minimum amount of current that evokes a measurable eCAP response. This eCAP threshold could potentially be used in the fitting of CIs, as an addition to behavioural measurements. The correlation between behavioural measurements and eCAP thresholds has been widely studied, but to date no strong correlations are found (7, 11, 12).

There are different eCAP threshold detection methods. Thresholds can be determined by clinical specialists or automatically via algorithms. Determining eCAP thresholds by clinical specialists is more time-consuming and can lead to a wider distribution of the results as each clinical specialists develops his own method and uses different criteria for threshold detection. Therefore, automatic eCAP threshold detection would be advantageous for a fully automatic and objective CI fitting.

In a previous systematic review (see Appendix A), an overview was created of different automatic threshold detection methods found in literature. All threshold detection methods (standard/non-automatic and automatic) could be divided into two categories: last visible technique and amplitude growth function (AGF) technique. A detailed explanation of these techniques and the different automatic threshold detection methods can be found in the review. The automatic methods found in literature were not ranked based on their outcome measures because of differences in methodology (e.g., intra-operative vs. post-operative), biases (e.g., selection bias, analytical bias) and limitations. One of the limitations is that there is currently no golden standard for eCAP threshold detection to compare (new) automatic methods with. As an alternative, many studies determined correlation coefficients between eCAP thresholds and behavioural thresholds. There are however substantial differences between eCAP measurements and behavioural measurements leading to a worse

correlation. Behavioural measurements are usually measured with tone bursts and with a higher stimulation rate compared to eCAP measurements. Therefore, correlations between behavioural and eCAP thresholds are not ideal to evaluate threshold detection methods either. Besides, a lot of studies suffered from analytical bias by only showing grouped correlations and no individual correlations. Grouped correlations can suppress variations in individual correlations, while individual correlations are important for investigating the fitting of CI based on eCAP thresholds. The correlation between eCAP thresholds and behavioural threshold is often used to represent the accuracy of threshold detection methods. Accuracy is a measure for the nearness of the calculated threshold value to the actual value. Another important outcome measure to evaluate the use of eCAP thresholds in clinical fitting of a CI, is the precision. Precision is the nearness of measurements as obtained with repeated measures. Only one study included in the review, the study of Gärtner et al. (13), determined the precision of their eCAP thresholds. Biesheuvel et al. (11), another study not included in the systematic review, created threshold confidence intervals to estimate the precision of eCAP thresholds determined with the AGF method.

1.4. Goals and objectives

To summarise, automatic detection of eCAP thresholds could potentially aid in the uniform and objective fitting of CIs. Different automatic threshold detection methods exist and have been examined as shown previously in a review but could not be compared with each other because of the used outcome measures and differences in methodology. There are measuring differences between behavioural thresholds and eCAP thresholds, which make the correlation between these two not the ideal measure of accuracy. Besides, the precision, an important outcome measure when introducing a method in clinical practice, has only been determined for one automatic threshold detection method. In addition, different averaging and artefact reduction methods exist for threshold detection without a clear consensus on the best method.

Therefore, the primary goal of this study is to determine the precision of different automatic eCAP threshold detection methods in combination with different averaging and artefact reduction methods (SA vs. FG and AP vs. FM). The secondary goal is to determine the accuracy of these different automatic eCAP threshold detection methods. In order to compensate for the measuring differences, behavioural measurements will also be determined with eCAP stimuli.

To achieve these goals, eCAP measurements were performed in CI users, different automatic eCAP threshold detection methods were constructed and used in combination with different averaging and artefact reduction methods to analyse these eCAP measurements and finally, the precision and accuracy of these combinations were determined.

2. Methods

2.1. Demographics

Participants were recruited from the ENT department of the LUMC. Participants were selected based on the following criteria: at least 18 years old, postlingually implanted with an AB Hires 90K HiFocus Mid-Scala or AB HiRes Ultra 3D HiFocus Mid-Scala CI and 1-3 years of experience with their CI (calculated from their first fitting session). Finally, 14 participants were included in the study. Written informed consent was obtained from each participant. This study was approved by the Medical Ethics Review Committee Leiden Den Haag Delft. The demographics of the included participants are shown in Table 2.1.

2.2. Measurements

Each participant underwent five measurements:

1. Impedance measurements. The impedances were used to exclude electrode contacts exceeding 10 kOhm and to determine the compliances of the electrode contacts.
2. T-level measurements using tone bursts with pulse properties as used in the daily program of the participant.
3. T-level measurements with eCAP stimuli. T-levels were determined using the FM eCAP stimuli with 10 averages.
4. M-level measurements with eCAP stimuli. M-levels were determined through the 8-points loudness scale (M-level is equal to 8, see Appendix B) using the FM eCAP stimuli with 10 averages. No analysis is performed on these measurements, M-levels are only used for the eCAP measurements.
5. eCAP measurements.

Measurements were performed on all odd electrode contacts, unless an odd electrode contact is switched off in a participant's daily used program or when the impedance of an odd electrode contact exceeds 10 kOhm. When an odd electrode contact was excluded, the nearest basal electrode contact was included. All measurements of one participant were performed on the same day. The complete measurement protocol can be found in Appendix C.

Impedance measurements and T-level measurements using tone bursts were performed using the SoundWave (SW) software from AB.

T- and M-level measurements with eCAP stimuli and eCAP measurements were recorded using the Bionic Ear Data Collection System (BEDCS) research software from AB, controlled by a custom-made MATLAB (Natick, MA, USA) interface.

Table 2.1: Demographics of included participants. CI = cochlear implant.

Participant ID	Age (years)	Gender	CI	Implantation side	Experience with CI (months)	Aetiology	Measured electrode contacts
A_001	75	Male	HiRes Ultra 3D	Right ear	27	Familial progressive hearing impairment, possibly combined with Meniere	1, 3, 5, 7, 9, 11, 13, 15
A_002	76	Male	HiRes 90K	Left ear	35	Unknown	1, 3, 5, 7, 9, 11, 13, 15
A_003	75	Male	HiRes Ultra 3D	Right ear	19	Familial progressive hearing impairment	1, 3, 5, 7, 9, 11, 13, 15
A_004	82	Male	HiRes 90K	Right ear	20	Unknown	1, 3, 5, 7, 9, 11, 13
A_005	68	Female	HiRes Ultra 3D	Right ear	20	Possibly familial	1, 3, 5, 7, 9*, 11, 13
A_006	58	Male	HiRes 90K	Left ear	28	Unknown	1, 3, 5, 7, 9, 11, 13, 15
A_007	74	Male	HiRes Ultra 3D	Right ear	13	Possibly familial, gene mutation	1, 3, 5, 7, 9, 11, 13, 15
A_008	70	Female	HiRes Ultra 3D	Right ear	29	Possibly familial (DFNA9)	1, 3, 4, 13, 16
A_009	84	Male	HiRes 90K	Right ear	16	Unknown	1, 3, 5, 7, 9, 11, 13, 15
A_010	73	Female	HiRes 90K	Left ear	20	Pendred syndrome, enlarged vestibular aqueducts on both sides	1, 3, 5, 7, 9, 11, 13, 15
A_011	63	Female	HiRes Ultra 3D	Left ear	17	DFNA9	1, 3, 5, 7, 9, 11, 13, 15
A_012	64	Male	HiRes Ultra 3D	Left ear	16	DFNA9	1, 3, 5, 7, 9, 11, 13, 15
A_013	41	Female	HiRes Ultra 3D	Right ear	12	Unknown	1, 3, 5, 7, 9, 11
A_014	60	Male	HiRes Ultra 3D	Left ear	31	Glomus tumours	1, 3, 5, 7, 9, 11, 13, 15
	Mean: 69				**Mean: 22**		**Total: 105 (104 FineGrain)**

* Only eCAP measurements with standard averaging for this electrode contact (FineGrain measurements could not be performed due to software malfunctioning)

For the eCAP measurements, two different averaging methods were used: SA and FG. For SA, eCAPs were recorded in a range from 0 µA to M-level (or compliance limit if compliance limit is reached before the M-level), divided in 10 linearly spaced stimulus levels, with 32 averages at each stimulus level. In FG, eCAPs were recorded in the same range (0 µA to M-level/compliance limit), but with a fixed step size of 7.5 µA instead of dividing this range in a fixed number of steps. The number of stimulus levels in FG averaging therefore varied depending on the M-level. In FG, 2 averages were recorded at each stimulus level. Each eCAP measurement consisted of five measured frames:

A. Masker
B. Masker + probe
C. Probe
D. Anodic leading probe
E. Zero-amplitude template

In frames A, B and C, eCAPs were measured with biphasic, cathodic leading pulses. In frame D, anodic leading biphasic pulses were used. For all frames, a phase width of 32 µs, a sampling rate of 56 kHz and a gain of 300 were used. For frame D, a masker probe interval of 400 µs was used. The recording electrode was two electrodes apical to the stimulus electrode, except for electrodes 1 and 2, where the recording electrode was two electrodes basal to the stimulus electrode. The recording was started before the start of the stimulus.

For the AP stimulus artefact reduction, the resulting eCAP response was constructed by: (C+D)/2-E (subtraction of zero-amplitude template for recording artefact reduction) and for the FM stimulus artefact reduction, the resulting eCAP response was constructed by: A-B+C-E (subtraction of zero-amplitude template for recording artefact reduction).

2.3. eCAP analysis

2.3.1. Pre-processing

FG eCAP measurements were averaged over seven adjacent stimulus levels, based on the FG method used by Estienne et al. (14). Next, for all measurements, the stimulus was removed. For the FM method, this was done by removing samples 1 to 35 and for the AP method samples 1 to 38. Then, the measurements were filtered using a low-pass filter with a cut-off frequency of 8 kHz. The next step was the automatic threshold detection. The different automatic threshold detection methods that have been used in this study are described in the sections 2.3.2. to 2.3.4.

2.3.2. Amplitude growth function threshold detection

To create an AGF, a peak-picking algorithm was used to detect the N1 and P1 peaks. N1 was defined as the minimum over the first 20 samples (after removing the stimulus) and P1 as the maximum over the first 60 samples after the N1 peak. The eCAP amplitude was calculated as the difference in voltage between the P1 and N1 peak. Next, the eCAP amplitudes were plotted in function of the stimulus levels and a sigmoid and linear function were fitted to this data.

For the sigmoid function, a Boltzmann function in the form of

$$y = a + \frac{b}{1 + e^{-\left(\frac{x-c}{d}\right)}}$$

Equation 1

was used, in which y was the eCAP amplitude, x the stimulus levels and a, b, c and d were the parameters to be fitted. Fitting was performed in a least-squares sense using the Levenberg-

Marquardt algorithm. Subsequently, the inflection point of this sigmoid function was determined, which was equal to the fitted parameter *c*. The tangent line at this inflection point was calculated, with the slope of the tangent line equal to *b*/(4*d*). Lastly, this tangent line was linearly extrapolated and the intersection with the x-axis is labelled as the eCAP threshold. Unrealistic fittings or fittings of poor quality with thereby unrealistic thresholds were prevented by setting requirements for calculating and marking a threshold as valid. These requirements were as follows: the inflection point of the fitted sigmoid function should fall within the range of measured stimulus levels, the threshold should be lower than the inflection point and fall within the range of measured stimulus levels and the upper confidence bound of the threshold (see 2.4.2. Precision of thresholds) can be determined within the measured range of stimulus levels. When these requirements were not met, an eCAP threshold was not calculated or marked as invalid.

For the linear AGF, a function of the form $y = a * x + b$ was used, in which *y* was the eCAP amplitude, *x* the stimulus levels and *a*, *b*, *c* and *d* were the parameters to be fitted. Whereas for the sigmoid AGF all measurements were used to construct the fit, for the linear AGF a selection of the measurements was used. The first selection was done by only selecting measurements of which the eCAP amplitude was at least 20 µV and the signal-to-noise ratio (SNR) was at least 6 dB. This SNR was defined as the ratio of the eCAP amplitude divided by the noise amplitude. This noise amplitude was in turn defined as the difference between the maxim and the minimum of the last 30 samples of an eCAP measurement (no eCAP response was expected in these samples). Further selection differed between SA and FG (but were the same for FM and AP). This further selection always started with the measurement with the lowest index that passed the first selection, and measurements were selected from that point on if they met the predetermined requirements. In SA, measurements were only selected if they met the two following criteria:

1. Amplitude(*i*) > Amplitude(*i-1*)
2. Amplitude(*i+1*) – amplitude(*i*) ≥ 0.8*(amplitude(*i*) – amplitude(*i-1*))

with *i* is the index of the measurement. If these criteria were not fulfilled, the measurement with index *i* and all the following measurements were not selected for the fit.

In FG measurements, further selection was performed by looking at six consecutive measurements (measurements with indexes *i*, *i+1*, *i+2*, *i+3*, *i+4* and *i+5*). If the eCAP amplitudes of all of the measurements with indexes *i+1*, *i+2*, *i+3*, *i+4* and *i+5* were lower than the eCAP amplitude of the measurement with index *i*, the measurement with index *i+1* and all the following measurements were not selected.

Same as for the sigmoid function, fitting of the linear AGF was performed in a least-squares sense using the Levenberg-Marquardt algorithm. This fitted function was then linearly extrapolated and the intersection with the x-axis is labelled as the eCAP threshold. Requirements to prevent invalid thresholds were set as well. For the linear AGF, thresholds were only calculated and marked as valid when parameter *a* of the fitted linear function was positive and the threshold and the upper confidence bound of the threshold (see 2.4.2. Precision of thresholds) fell within the range of measured stimulus levels.

2.3.3. Signal-to-noise ratio threshold detection

The SNR threshold detection method was based on the method by Hoth et al. (15) and Gärtner et al. (13). The first step of this method was to split the measurement into two equal time windows: the first half of the measurement is the time window with an eCAP response and noise, and the second half is the time window with noise only. Secondly, the variance is calculated in both of these time

windows, with $variance = \frac{1}{t2-t1}\int_{t1}^{t2}(y - y_{ref})^2 dt$, with *t1* being the first and *t2* being the last point of the time window, *y* is the measured signal and y_{ref} is a reference line of the form $y_{ref} = a * \left(1 - b * e^{-\frac{t}{\tau}}\right)$ fitted to the data to compensate for the stimulus artefact and possible DC components. Subsequently, the quotient $SNR = \frac{variance_{response+noise\ window}}{variance_{noise\ window}}$ was derived and converted to decibels and plotted in function of the stimulus levels. Similar as for the AGF threshold detection, the sigmoid Boltzmann function (Equation 1) is fitted. The intersection of this sigmoid fit with a horizontal line at 6 dB is labelled as the eCAP threshold. At 6 dB, the signal amplitude exceeds the background by at least a factor of 2 and is therefore a common choice in many applications of signal detection (15). In order to prevent unrealistic fittings and thereby unrealistic thresholds, thresholds were only calculated and marked as valid when the threshold lies within the measured range of stimulus levels and when the upper confidence bound of the threshold (see 2.4.2. Precision of thresholds) could be determined within the measured range of stimulus levels.

2.3.4. Cross-covariance threshold detection

Cross-covariance is a measure of similarity between two signals. The reasoning behind this technique for using it as a threshold detection method is as follows. eCAP measurements contain an eCAP response and noise. The noise is random and thus differs between stimulus levels, a possible eCAP response on the other hand is the same in all stimulus levels (except for the eCAP amplitude that can differ between stimulus levels). Therefore, if an eCAP measurement at a higher stimulus level contains an eCAP response, an eCAP measurement at a lower stimulus level contains an eCAP response as well only if it sufficiently resembles the eCAP measurement at the higher stimulus level, thus if the cross-covariance between the two stimulus levels is high enough. Normalised cross-covariance is used for the eCAP threshold detection to compensate for the possible difference in eCAP amplitudes between stimulus levels. Two forms of cross-covariance were used for threshold detection: the cross-covariance between adjacent stimulus levels and the cross-covariance between all stimulus levels and the maximum stimulus level. The processing of these two forms is the same. Cross-covariance is always calculated for different time lags (shifting the two signals that are being compared in time relative to each other). As differences in eCAP latencies between stimulus levels is possible, further calculations are performed with the maximum cross-covariances (independent at which time lags these were). These maximum cross-covariances are plotted in function of the stimulus levels (since you are comparing two stimulus levels with each other, the stimulus levels on the x-axis are always the lowest of the two stimulus levels). Next, same as for the sigmoid AGF threshold detection method, a sigmoid Boltzmann function (Equation 1) is fitted to this plot and the inflection point and tangent line at this inflection point are determined. As described above, an eCAP response is present if the cross-covariance is high enough. There are no previously published studies using cross-covariance as eCAP threshold detection method to define how high the cross-covariance should be. There are studies using cross-covariance as threshold detection method for auditory brainstem responses (16), but is remains unclear whether the same threshold of cross-covariance can be used for eCAP measurements. Therefore, there is chosen not to work with one predefined value of the cross-covariance to determine the threshold. Instead, the lower asymptote of each sigmoid fit is determined, which is equal to value of fitted parameter *a* of the Boltzmann function. Finally, the eCAP threshold is defined as the intersection of this lower asymptote with the linear extrapolation of the tangent line at the inflection point of the sigmoid fit. To prevent unrealistic fittings and thresholds, the same requirements as for the AGF sigmoid method were used for both methods of cross-covariance: the inflection point of the fitted sigmoid function should fall within the range of measured stimulus levels, the threshold should be lower than the inflection point and fall within the range of measured

stimulus levels and the upper confidence bound of the threshold (see 2.4.2. Precision of thresholds) can be determined within the measured range of stimulus levels.

2.3.5. Combinations of artefact reduction methods and automatic threshold detection methods

Two different methods of signal noise reduction were performed: SA and FG, and two different methods of stimulus artefact reduction: AP and FM. The four different combinations of these methods: SA AP, SA FM, FG AP and FG FM will be called the artefact reduction methods from now on for simplicity reasons (but the SA and FG are strictly speaking no artefact reduction methods but averaging methods). Besides the artefact reduction methods, there were five different methods of automatic threshold detection methods: AGF sigmoid, AGF linear, SNR, cross-covariance between adjacent levels and cross-covariance with maximum level. All possible combinations of the artefact reduction methods with the threshold detection methods were performed, thus in total 20 different combinations of methods were tried.

2.4. Outcome measures

2.4.1. Success rates of methods

The success rate is defined as the percentage of the total measured electrode contacts over all participants in which an eCAP threshold could be defined and was marked as valid.

2.4.2. Precision of thresholds

All automatic threshold detection methods used in this thesis, are based on a fitted function. Therefore, the MATLAB-function 'nlpredci' can be used to determine the estimated function values and their corresponding 95% upper and lower confidence bounds. For the linear AGF and the SNR method, these bounds were extrapolated and the intersections of these bounds with the x-axis or the 6 dB line respectively were defined as the lower and the upper threshold confidence bounds. For the other automatic eCAP threshold detection methods, the distance at the inflection point between the lower and upper confidence bounds was calculated, this distance or width was projected on the level of the determined threshold and the lower and upper threshold confidence bounds were determined at this projected width on the level of the threshold.

When a lower threshold confidence bound could not be determined within the measured range of stimulus levels, the lower threshold confidence bound was set at 0 µA (as a threshold could never be below 0 µA). When an upper threshold confidence bound could not be determined within the measured range of stimulus levels, the determined eCAP threshold was marked invalid (as explained before). No threshold confidence bounds were determined for invalid thresholds.

After the lower and upper threshold confidence bounds were determined, the difference between these bounds was calculated as a measure for the threshold confidence interval (TCI). Next, these TCIs were normalised by dividing them with their respective determined eCAP thresholds and the normalised TCIs were converted to decibels (dB). Normalisation was performed to correct for the dynamic range of a CI user. The dynamic range of a CI user is the difference between T- and M-level. The same difference between lower and upper threshold confidence bounds would have a bigger impact in a CI user with a smaller dynamic range compared to a CI user with a bigger dynamic range. As the TCIs are only divided by the threshold and not by the dynamic range itself, it is assumed that there is a correlation between eCAP thresholds and the dynamic range. The TCIs in dB are a measure of precision of eCAP thresholds. TCIs from all participants over all electrode contacts are analysed together, as it is desirable that an automatic eCAP threshold detection method has a high precision for all participants and all electrode contacts.

Analysis of the TCIs (in dB) is performed by a two-way analysis of variance (ANOVA). The two-way ANOVA was constructed to test the effects of the artefact reduction methods, the threshold detection methods and the combinations of the artefact reduction methods and threshold detection methods on the (mean of the) TCIs.

2.4.3. Accuracy of thresholds

As mentioned before, there are measuring differences between standard measurements of T-levels and eCAP measurements. Therefore, T-levels were determined using tone bursts in SW (standard method) and with eCAP stimuli. Correlations were calculated between eCAP thresholds and the T-levels measured in SW and between eCAP thresholds and T-levels measured with eCAP stimuli. The Pearson correlation coefficient was calculated over the electrode array of each participant individually and was only calculated when there were valid eCAP thresholds on at least three electrode contacts. These correlation coefficients can serve as a measure for the accuracy of the eCAP thresholds.

To enable further analyses on the correlation coefficients, Fisher's z-transformation was performed on the correlation coefficients. Similarly as the analysis of the precision, ANOVA models were used. First, a one-way ANOVA was conducted to test the effect of T-level measurement method on the correlation. Next, two separate two-way ANOVAs were constructed, one for the correlation between T-levels measured in SW and eCAP thresholds, and one for the correlation between T-levels measured with eCAP stimuli and eCAP thresholds. Analogous to the two-ANOVA of precision, for both two-way ANOVAs of accuracy, the effects of the artefact reduction methods, the threshold detection methods and the combinations of the artefact reduction methods and threshold detection methods on the (mean) correlation coefficients were tested.

3. Results

3.1. Threshold detection methods

Figures 3.2 to 3.5 display the different automatic eCAP threshold detection methods in combination with the different artefact reduction methods on electrode contact 7 of participant A_005 (this electrode contact is chosen as an example). Figure 3.6 and Figure 3.7 display the different automatic eCAP thresholds detection methods in combination with FG FM on electrode contact 13 of participant A_008 and on electrode contact 5 of participant A_013 respectively.

As mentioned before, the cross-covariance method has no earlier (published) studies for eCAP threshold detection and therefore it was chosen not to work with a fixed value for the threshold determination, but to calculate the intersection with the lower asymptote of the fitted sigmoid function. In theory, every electrode contact had thus a different value for the threshold determination. Therefore, the mean values and the accompanying 95% confidence intervals of the asymptote for the different combinations of threshold detection methods were determined and displayed in Table 3.1.

Table 3.1: Asymptote values of fitted cross-covariance functions. SA = standard averaging, AP = alternating polarity, FM = forward masking, FG = FineGrain averaging.

	Cross-covariance between adjacent levels	Cross-covariance with maximum level
SA AP	0.30 (0.27-0.33)	0.30 (0.29-0.32)
SA FM	0.31 (0.30-0.32)	0.29 (0.28-0.30)
FG AP	0.87 (0.87-0.88)	0.31 (0.29-0.33)
FG FM	0.86 (0.85-0.86)	0.31 (0.30-0.32)

3.2. Success rates

The success rates of the 20 different combinations of artefact reduction methods and automatic threshold detection methods over all electrode contacts and participants are graphically represented in Figure 3.1. For all automatic threshold detection methods and for both SA and FG, the success rates of FM were equally high or higher than the success rates of AP. Besides from the SNR method, the combinations with the FG method had higher success rates than the combinations with the SA method. The combination of the FG and FM methods resulted in the highest success rates for the automatic threshold detection methods AGF sigmoid, AGF linear, cross-covariance between adjacent levels and cross-covariance with maximum level. For the SNR method, the combination of SA and FM delivered the highest success rate.

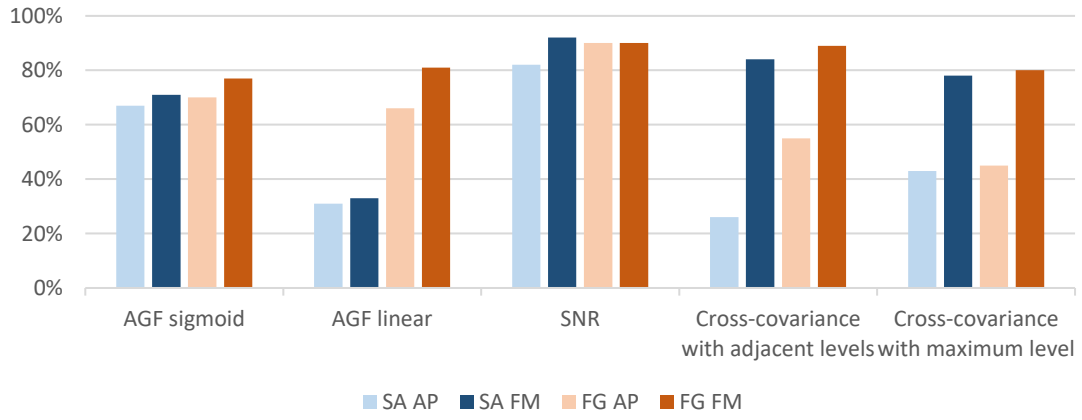


Figure 3.1: Success rates of different methods. AGF = amplitude growth function, SNR = signal-to-noise ratio, SA = standard averaging, AP = alternating polarity, FM = forward masking, FG = FineGrain.

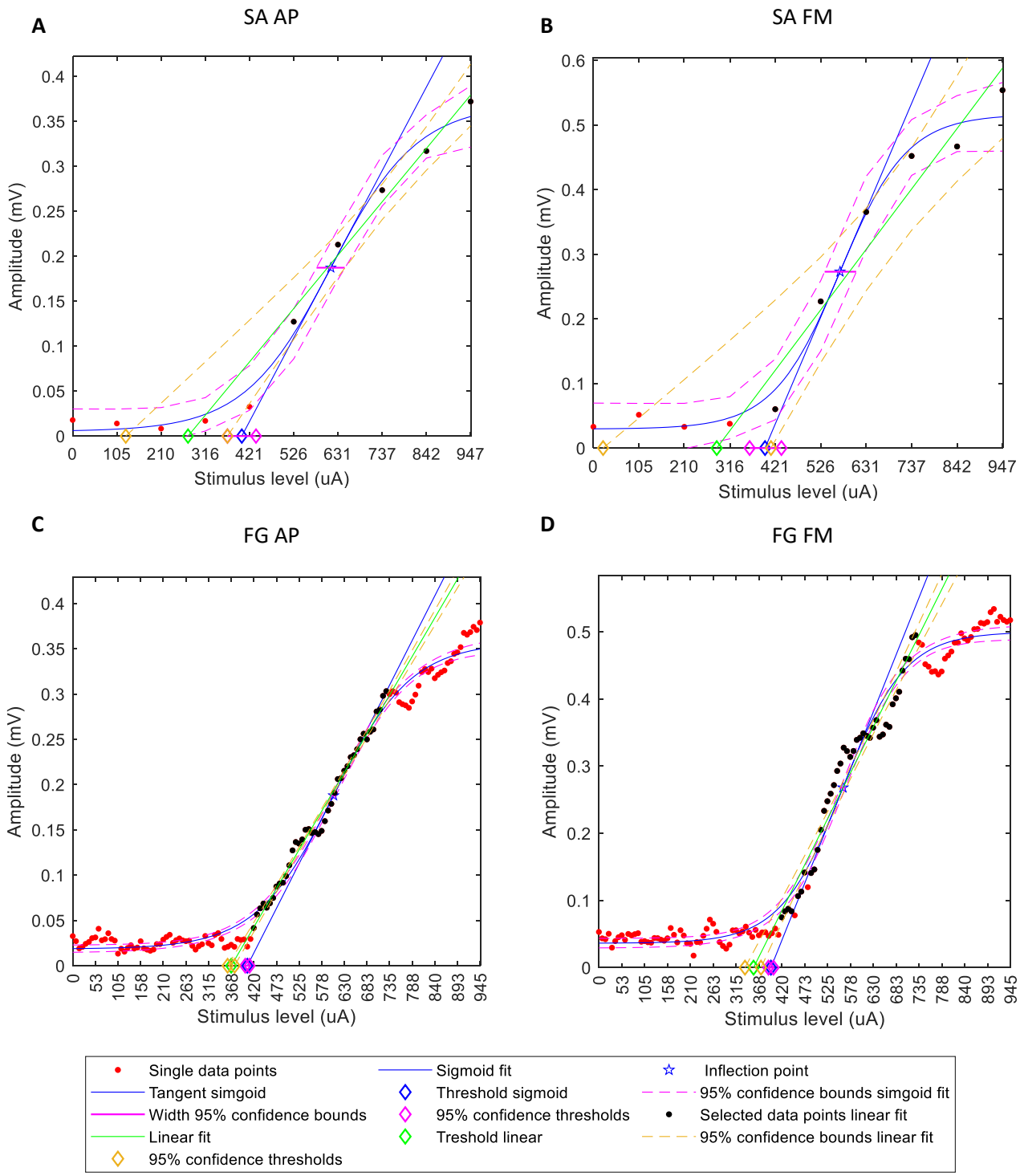


Figure 3.2: Linear and sigmoid amplitude growth function method on electrode contact 7 of participant A_005. **A:** Standard averaging (SA) and alternating polarity (AP). **B:** SA with forward masking (FM). **C:** FineGrain averaging (FG) with AP. **D:** FG with FM.

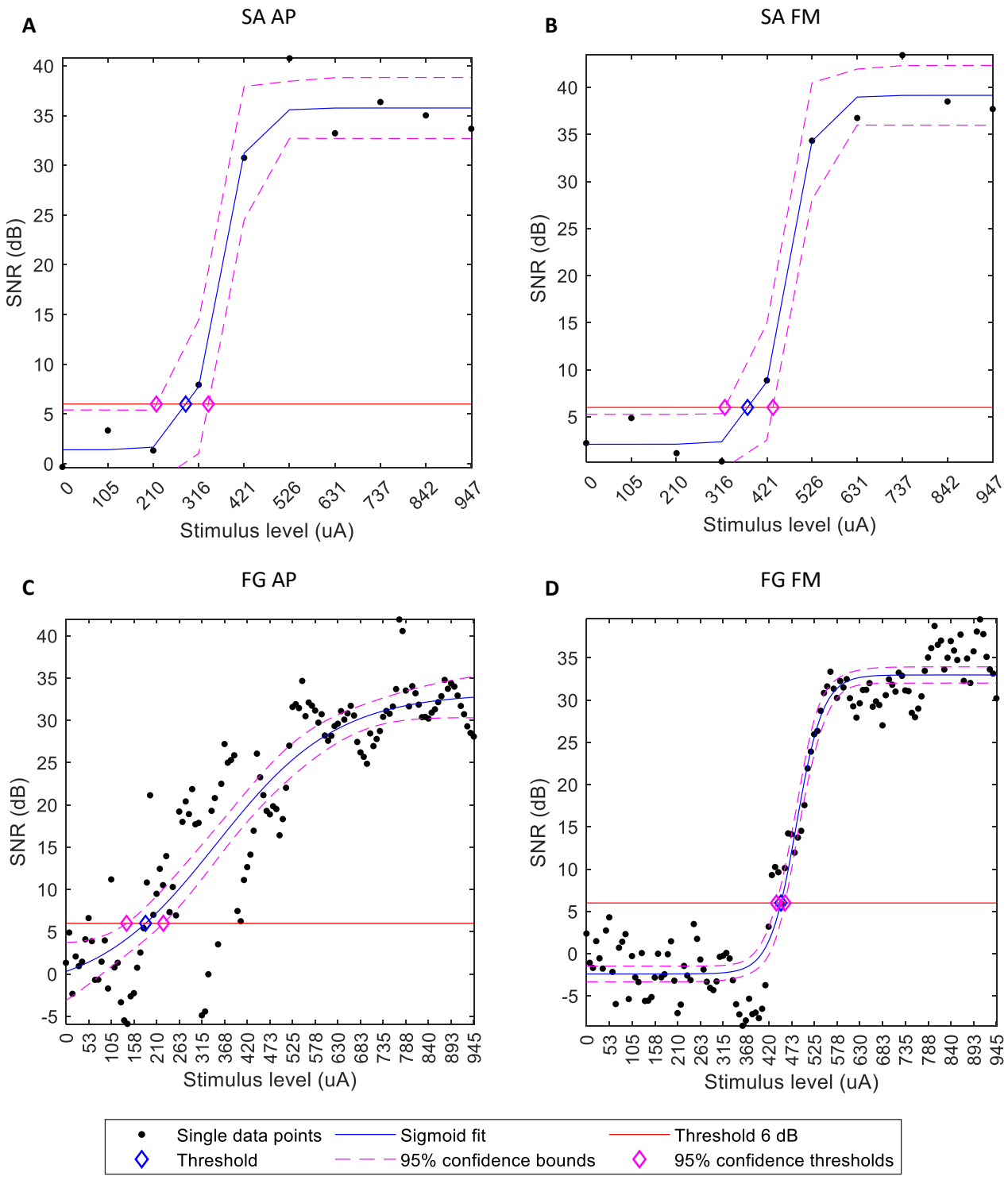


Figure 3.3: Signal-to-noise ratio method on electrode contact 7 of participant A_005. **A:** Standard averaging (SA) and alternating polarity (AP). **B:** SA with forward masking (FM). **C:** FineGrain averaging (FG) with AP. **D:** FG with FM.

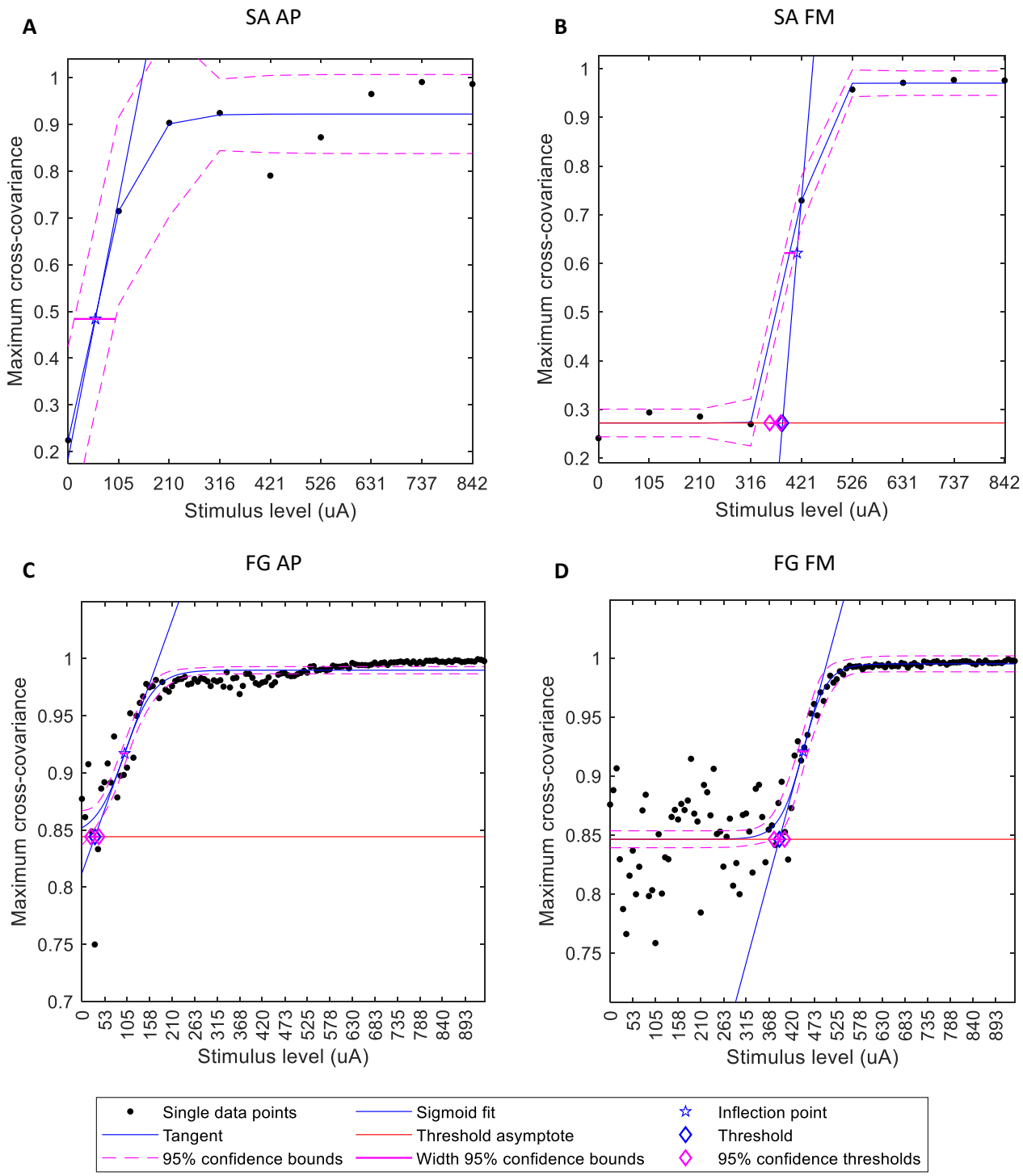


Figure 3.4: Cross-covariance between adjacent levels method on electrode contact 7 of participant A_005. **A:** Standard averaging (SA) and alternating polarity (AP). **B:** SA with forward masking (FM). **C:** FineGrain averaging (FG) with AP. **D:** FG with FM.

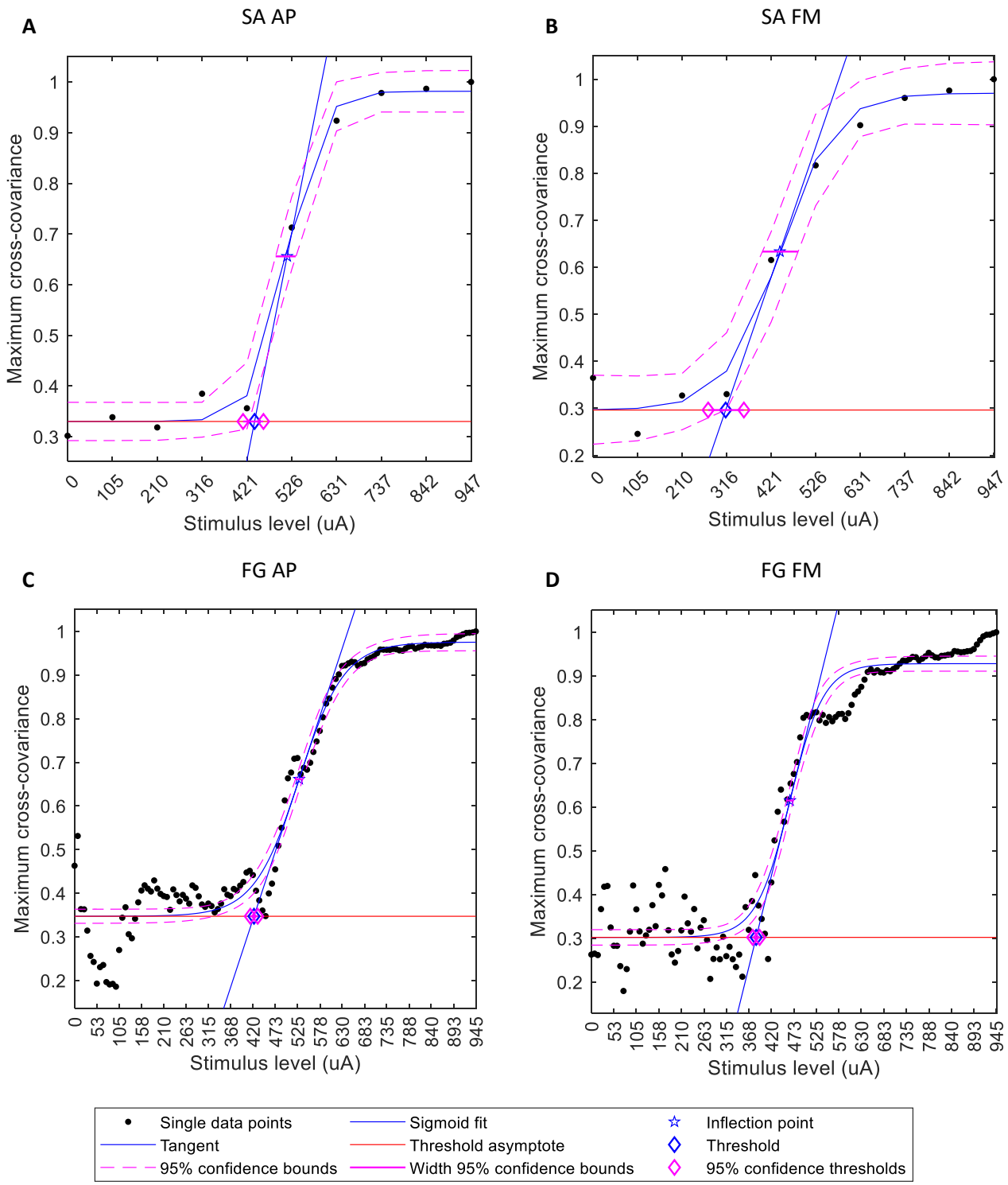


Figure 3.5: Cross-covariance with maximum level method on electrode contact 7 of participant A_005. **A:** Standard averaging (SA) and alternating polarity (AP). **B:** SA with forward masking (FM). **C:** FineGrain averaging (FG) with AP. **D:** FG with FM.

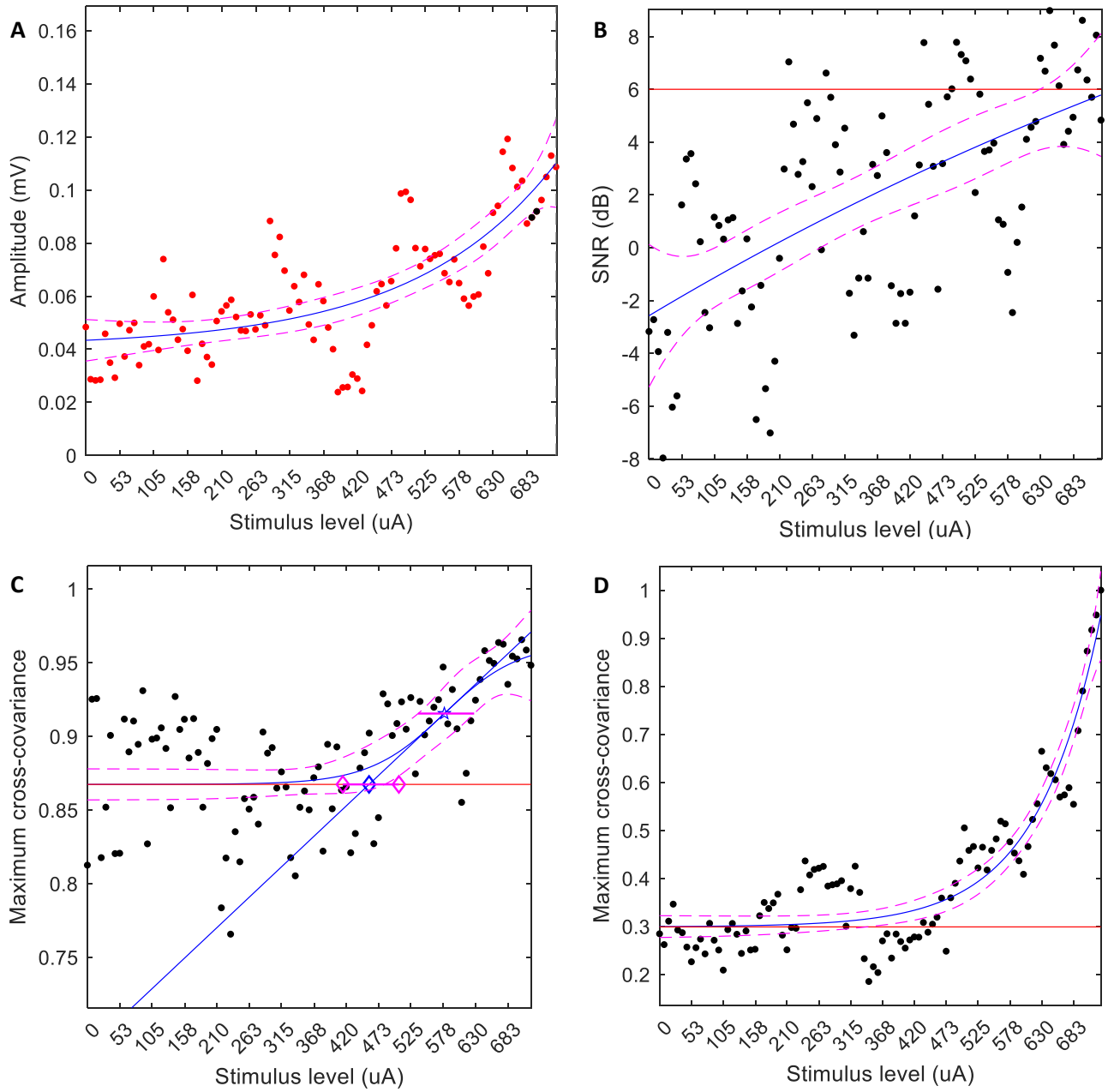


Figure 3.6: Different threshold detection methods with FineGrain averaging and forward masking on electrode contact 13 of participant A_008. **A:** Sigmoid and linear amplitude growth function method. **B:** Signal-to-noise ratio method. **C:** Cross-covariance between adjacent levels method. **D:** Cross-covariance with maximum level method.

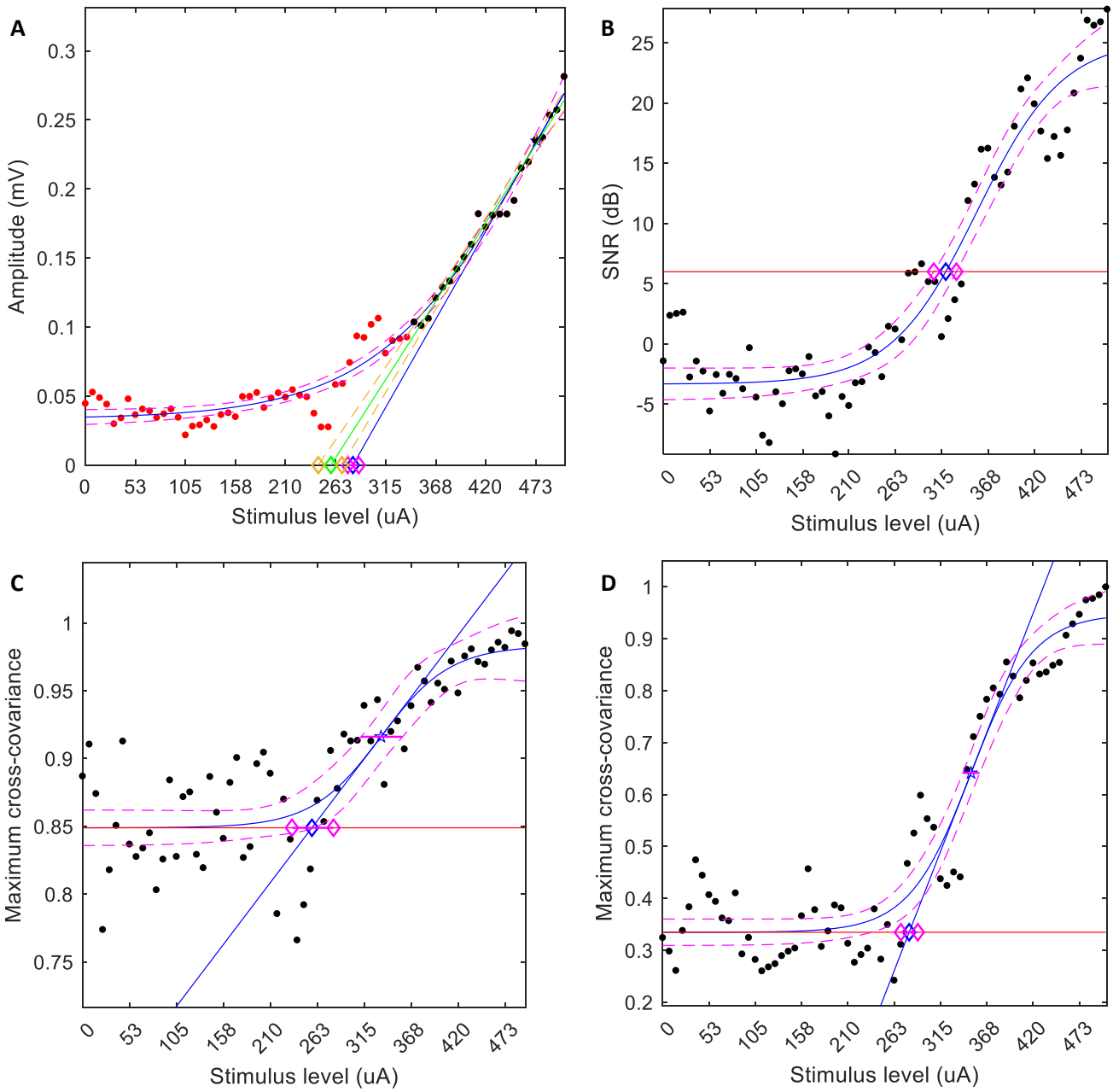


Figure 3.7: Different threshold detection methods with FineGrain averaging and forward masking on electrode contact 5 of participant A_013. **A:** Sigmoid and linear amplitude growth function method. **B:** Signal-to-noise ratio method. **C:** Cross-covariance between adjacent levels method. **D:** Cross-covariance with maximum level method.

3.3. Precision of thresholds

Figure 3.8 displays the distribution of the normalised TCIs in dB for the different combinations of artefact reduction methods and automatic eCAP threshold detection methods.

The two-way ANOVA revealed that there were significant main effects of artefact reduction methods ($F(3,1412)=200.18$, $p<0.001$) and threshold detection methods ($F(4,1412)=141.34$, $p<0.001$) on the normalised TCIs (in dB) and that there was a significant interaction between the effects of artefact reduction methods and threshold detection methods ($F(12,1412)=24.09$, $p<0.001$). For the further analysis of these results, a multiple comparison test (with the Tukey's honest significant difference criterion) was performed. This test showed that all four different combinations of artefact reduction methods have significantly different population marginal means for the TCIs, with the combination of FG FM having the lowest mean TCI (-18,.98 dB). The population marginal means of the five different threshold detection method did not all differ significantly, but the threshold detection method with the lowest mean TCI, the sigmoid AGF method (-19.60 dB), differed significantly from the other four threshold detection methods. Lastly, the multiple comparison test was used to compare the population marginal means of the different combinations of artefact reduction methods and threshold detection methods. The combination of FG AP with the sigmoid AGF method had the lowest mean TCI (-25.29 dB), and this mean was significantly different from all other combinations, except from the combinations: FG FM with sigmoid AGF, FG AP with cross-covariance with maximum level and FG FM with cross-covariance with maximum level. The results of the multiple comparison tests are also displayed in Figures 3.9 to 3.11. The circles represent the population marginal means of the normalised TCIs (in dB) for the different methods, the lines represent the 95% confidence intervals of these means. The best performing methods (or combinations of methods) are shown in blue, the methods in red are significantly different from the best method and the methods in grey are not significantly different from the best method. The difference between two methods is significantly different if their lines of the 95% confidence intervals do not overlap. The exact numbers of the means and their standard deviations can be found in Appendix C.

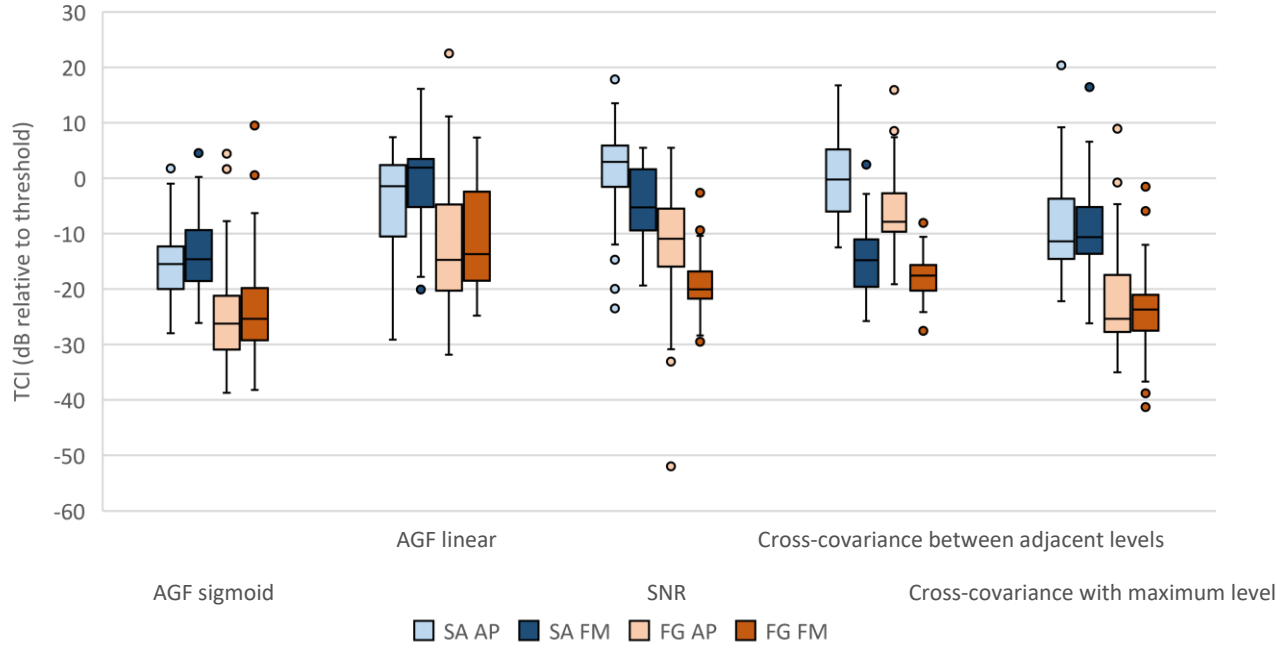


Figure 3.8: Normalised threshold confidence intervals (TCIs) in dB of different methods. AGF = amplitude growth function, SNR = signal-to-noise ratio, SA = standard averaging, AP = alternating polarity, FM = forward masking, FG = FineGrain averaging.

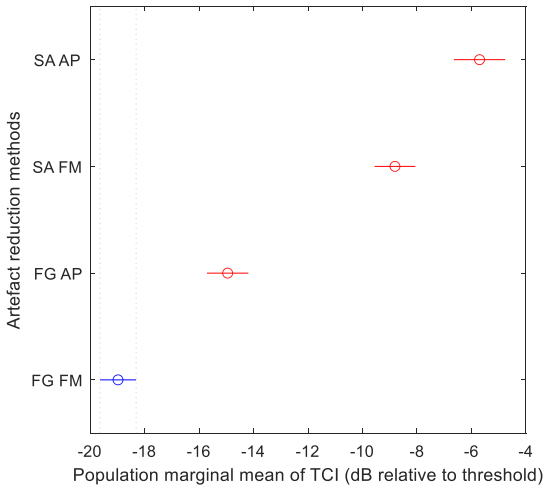


Figure 3.9: Results of the multiple comparison test of the effect of artefact reduction method on the normalised threshold confidence intervals (TCIs) (in dB). SA = standard averaging, AP = alternating polarity, FM = forward masking, FG = FineGrain averaging.

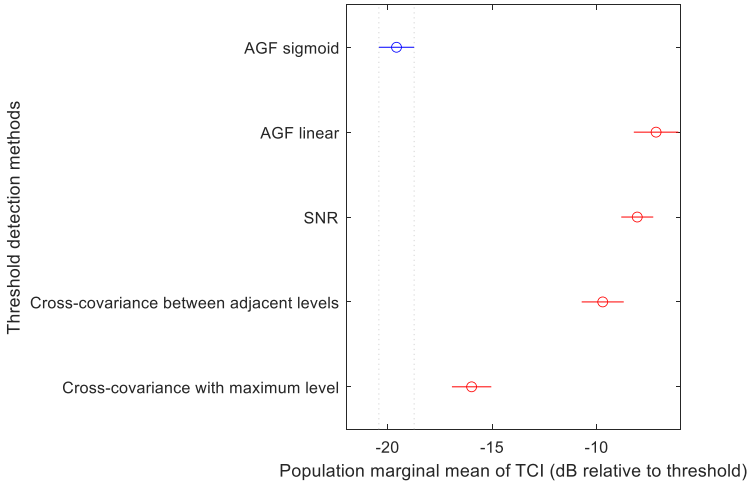


Figure 3.10: Results of the multiple comparison test of the effect of threshold detection method on the normalised threshold confidence intervals (TCIs) (in dB). AGF = amplitude growth function, SNR = signal-to-noise ratio.

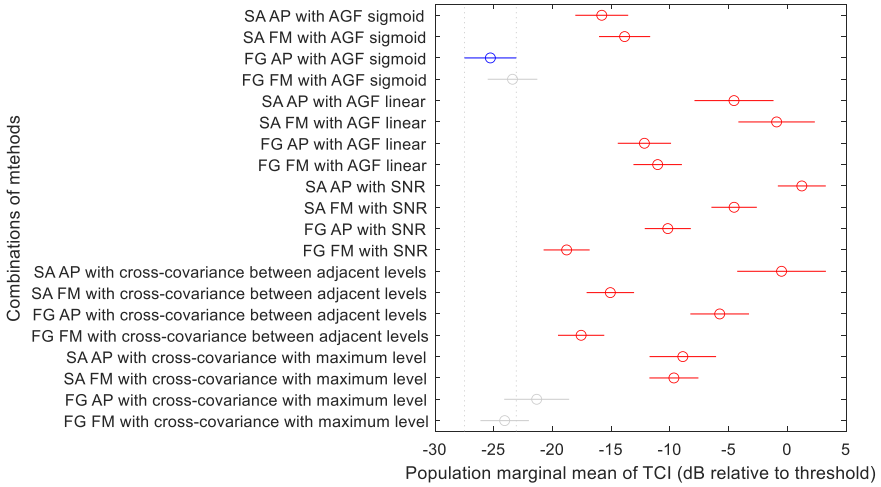


Figure 3.11: Results of the multiple comparison test of the effect of combinations of methods on the normalised threshold confidence intervals (TCIs) (in dB). SA = standard averaging, AP = alternating polarity, AGF = amplitude growth function, FM = forward masking, FG = FineGrain averaging, SNR = signal-to-noise ratio.

3.4. Accuracy of thresholds

Figures 3.12 to 3.16 graphically represent the ranges of the Pearson correlation coefficients of the individual participants of the correlation between the T-levels measured in SW and the eCAP thresholds determined with the different combinations of methods in the solid box plots, and of the correlation between the T-levels measured with eCAP stimuli and the eCAP thresholds determined with the different combinations of methods in the striped box plots.

The one-way ANOVA revealed that there was a significant difference in Fisher transformed correlation coefficients between the correlation with T-levels measured in SW and the correlation with T-levels measured with eCAP stimuli ($F(1,452)=49.8$, $p<0.001$). The correlation between T-levels measured with eCAP stimuli and eCAP thresholds resulted in the highest (Fisher transformed) correlation coefficients (population marginal mean of 0.71), significantly different from the population marginal mean of the correlation between T-levels measured in SW and eCAP thresholds (0.12).

The two-way ANOVA of the correlation between T-levels measured in SW and eCAP thresholds demonstrated that neither of the main effects artefact reduction methods and threshold detection method had a significant impact on the Fisher transformed correlation coefficients ($F(3,225)=0.38$, $p=0.77$ and $F(4,225)=0.81$, $p=0.52$ respectively), nor was the interaction between them significant ($F(12,225)=0.79$, $p=0.66$). Further multiple comparison testing (with the Tukey's honest significant difference criterion as has been used for the precision) revealed that not a single artefact reduction method, threshold detection method or a combination of those had a population marginal mean Fisher transformed correlation coefficient significantly different from another.

The similar two-way ANOVA of the correlation between T-levels measured with eCAP stimuli and eCAP thresholds showed completely opposite results. The single main effects artefact reduction methods and threshold detection methods and the interaction between these two were significant ($F(3,226)=3.6$, $p=0.01$, $F(4,226)=5.88$, $p<0.001$ and $F(12,226)=3.59$, $p<0.001$ respectively). Additional multiple comparison tests for the single main effect artefact reduction methods displayed that only the population marginal means of the Fisher transformed correlation coefficients from the artefact reduction methods with the highest and lowest population marginal mean differed significantly. The highest mean was found for the FG FM method (0.87), and the lowest mean was found for the FG AP method (0.49). Analogous testing of the single main effect threshold detection method showed that the threshold detection method sigmoid AGF had the highest population marginal mean of the Fisher transformed correlation coefficient (1.09), this mean was significantly different from the means of the threshold detection methods: cross-covariance between adjacent levels (0.31) and cross-covariance with maximum level (0.54). Lastly, multiple comparison testing of the interaction between artefact reduction methods and threshold detection methods revealed that the combination of the FG AP artefact reduction method with the sigmoid AGF threshold detection method resulted in the highest population marginal mean of the Fisher transformed correlation coefficient (1.49). This mean was significantly different from the combinations: FG FM with linear AGF (0.31), SA AP with SNR (0.17), FG AP with SNR (0.33), SA AP with cross-covariance between adjacent levels (0.03), FG AP with cross-covariance between adjacent levels (-0.25) and FG AP with cross-covariance with maximum level (-0.07). The results of the multiple comparison tests of the correlation between eCAP thresholds and T-levels measured with eCAP stimuli are also displayed in Figures 3.17 to 3.19. The circles represent the population marginal means of Fisher transformed correlation coefficients for the different methods, the lines represent the 95% confidence intervals of these means. The best performing methods (or combinations of methods) are shown in blue, the methods in red are significantly different from the best method and the methods in grey are not significantly different from the best method. The difference between two methods is significantly different if their lines of the 95%

confidence intervals do not overlap. The exact numbers of the means and their standard deviations can be found in Appendix E.

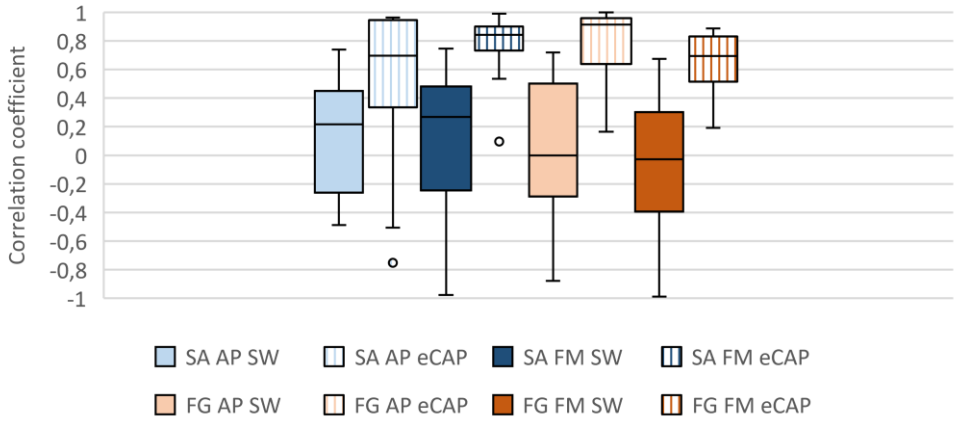


Figure 3.12: Correlation coefficients of correlation between T-levels and eCAP thresholds determined by the amplitude growth function sigmoid method. SA = standard averaging, AP = alternating polarity, SW = SoundWave, eCAP = electrically evoked compound action potential, FM = forward masking, FG = FineGrain averaging.

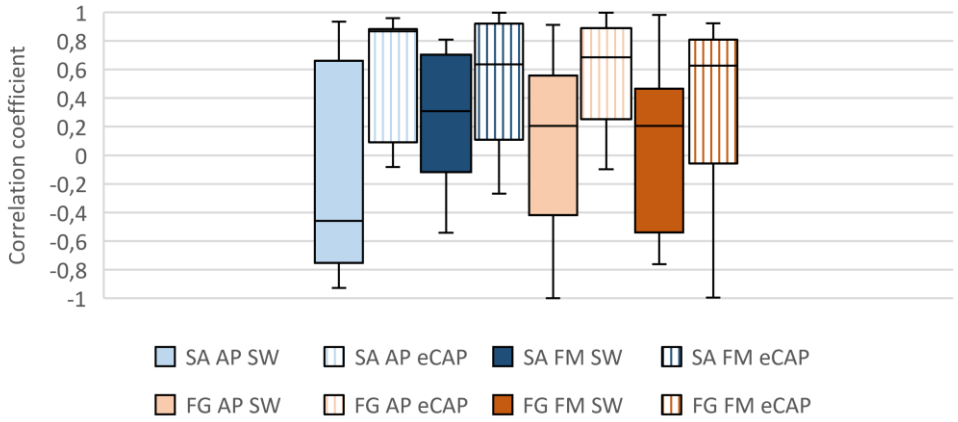


Figure 3.13: Correlation coefficients of correlation between T-levels and eCAP thresholds determined by the amplitude growth function linear method. SA = standard averaging, AP = alternating polarity, SW = SoundWave, eCAP = electrically evoked compound action potential, FM = forward masking, FG = FineGrain averaging.

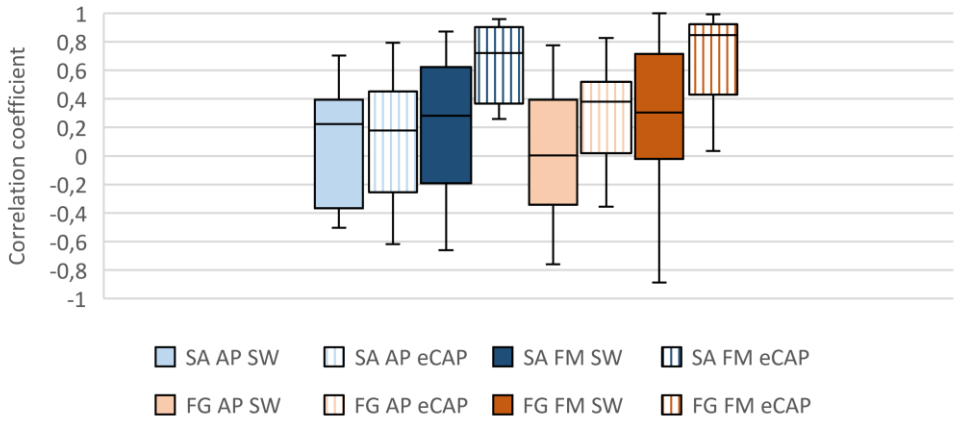


Figure 3.14: Correlation coefficients of correlation between T-levels and eCAP thresholds determined by the signal-to-noise ratio method. SA = standard averaging, AP = alternating polarity, SW = SoundWave, eCAP = electrically evoked compound action potential, FM = forward masking, FG = FineGrain averaging.

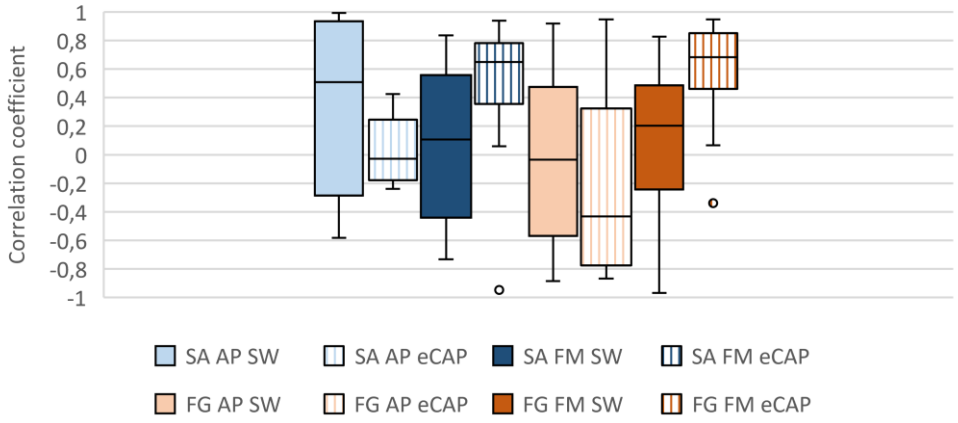


Figure 3.15: Correlation coefficients of correlation between T-levels and eCAP thresholds determined by cross-covariance between adjacent levels method. SA = standard averaging, AP = alternating polarity, SW = SoundWave, eCAP = electrically evoked compound action potential, FM = forward masking, FG = FineGrain averaging.

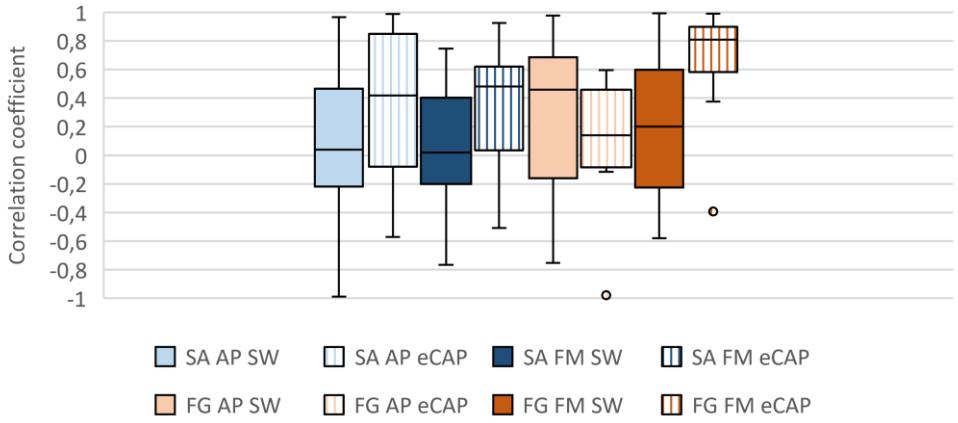


Figure 3.16: Correlation coefficients of correlation between T-levels and eCAP thresholds determined by the cross-covariance with maximum level method. SA = standard averaging, AP = alternating polarity, SW = SoundWave, eCAP = electrically evoked compound action potential, FM = forward masking, FG = FineGrain averaging.

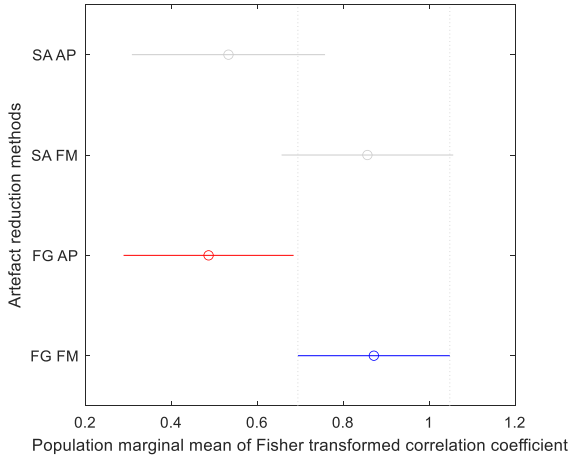


Figure 3.17: Results of the multiple comparison test of the effect of artefact reduction method on the Fisher transformed correlation coefficients for the correlation between electrically evoked compound action potential (eCAP) thresholds and T-levels measured with eCAP stimuli. SA = standard averaging, AP = alternating polarity, FM = forward masking, FG = FineGrain averaging.

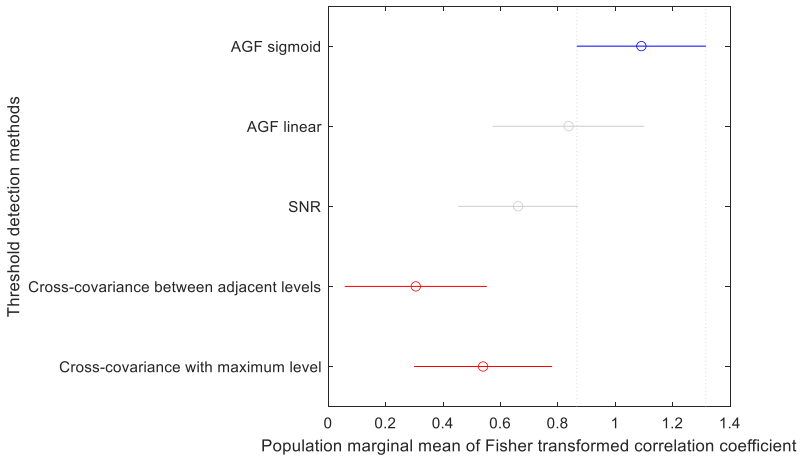


Figure 3.18: Results of the multiple comparison test of the effect of threshold detection method on the Fisher transformed correlation coefficients for the correlation between electrically evoked compound action potential (eCAP) thresholds and T-levels measured with eCAP stimuli. AGF = amplitude growth function, SNR = signal-to-noise ratio.

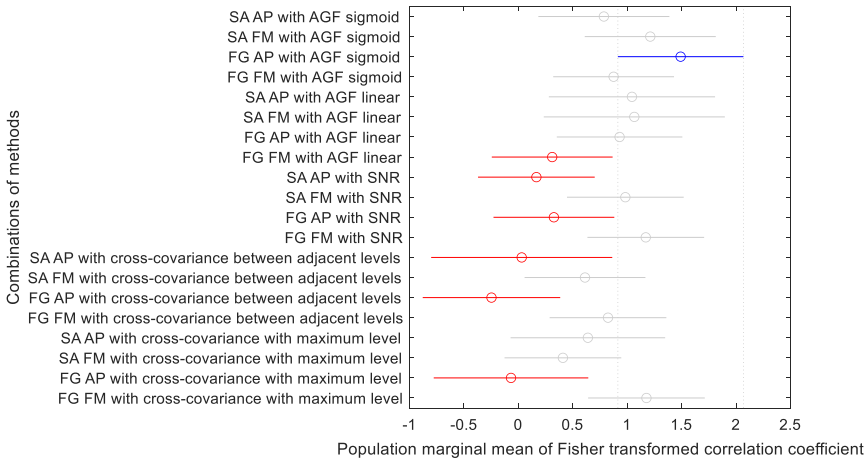


Figure 3.19: Results of the multiple comparison test of the effect of combinations of methods on the Fisher transformed correlation coefficients for the correlation between electrically evoked compound action potential (eCAP) thresholds and T-levels measured with eCAP stimuli. SA = standard averaging, AP = alternating polarity, AGF = amplitude growth function, FM = forward masking, FG = FineGrain averaging, SNR = signal-to-noise ratio.

4. Discussion

4.1. General limitations and recommendations

The inclusion criteria for the participants of this study were limited, this was partially done to create as homogeneous a group as possible. This way, differences in outcome measures due to differences between participants were limited. There were already a lot of other factors that could influence the outcome measures (differences in averaging, differences in stimulus artefact reduction) and thus the choice was made to limit other possible influences. However, if the goal is to use eCAP thresholds in the clinical fitting of CIs, it would be beneficial to study this on the whole patient population of CIs (e.g., different types and brands of CIs, pre- and postlingually implanted patients).

The inclusion criteria of 1 to 3 years of experience with their CI, was specifically set to ensure stable T-levels and stable eCAP thresholds. van der Beek et al. (17) have shown an increase in T-levels over the first year after implantation. Telmesani et al. (18) and Spivak et al. (19) have demonstrated a significant decrease in eCAP thresholds between intraoperative measurements and measurements at first fitting (usually one month after implantation). In years following the first fitting, small increases in eCAP thresholds have been observed by Brown et al. (20) and Molisz et al. (21). In this study, it was chosen to perform eCAP measurements at the same time as the T-levels measurements to eliminate possible differences of time on the outcome measures. Nevertheless, in the current clinical practice, eCAP measurements are often routinely performed intra-operatively. It would be interesting to study the impact on the outcome measures when this study was to be performed with intra-operative eCAP measurements or with eCAPs measured at the first fitting session and whether these early eCAP measurements can be used for CI fitting. Another interesting topic would be the change of eCAP thresholds over time and if eCAP thresholds can be used to follow-up the fitting of a CI over time.

Apart from AP and FM for stimulus artefact reduction, different other methods exist, such as triphasic pulses. As the name suggests, in this artefact reduction method, triphasic pulses instead of biphasic pulses are used. The third phase has a small portion of the charge of the first phase and this third phase can possibly resolve the artefact arising from decaying charge of the stimulus faster (22). Triphasic pulses have not been used in this study due to software malfunctioning, but it would be interesting to study whether triphasic pulses can further improve the determination of eCAP thresholds.

Due to limited time and data, eCAP threshold detection methods based on machine learning models were not examined. Different methods with machine learning models have been tried before (as can be seen from the review in Appendix A), such as NRT artificial neural network (23), automatic ART (24) and automatic NRT (25). For none of these methods, the precision of the determined thresholds has been studied and this could be a topic for further research and how the precision of these machine learning methods relates to the precision of the threshold detection methods used in this study. Besides, many of these machine learning methods make use of adaptive measurements. In adaptive measurements, measurement parameters are adapted depending on whether an eCAP threshold is detected or not. In automatic NRT for example, the step size of the stimulus levels is decreased once a possible eCAP threshold has been detected (25). The SA and FG methods used in this study had different step sizes and the smaller step sizes of the FG methods resulted in smaller TCIs. It is possible that adaptive measurements can even further improve the precision of eCAP thresholds.

4.2. Threshold detection methods

All automatic eCAP thresholds detection methods used in this study, apart from the linear AGF method, used a sigmoid fit. In order to obtain a good sigmoid fit, a plateau phase should be reached at the highest measured stimulus levels. In this study, eCAPs were measured from 0 µA to M-level for

all included electrode contacts, unless the compliance limit was reached before the M-level could be reached. It often occurred that the compliance limit was reached before the M-level, this compliance limit was then usually scored a 7 on the 8-points loudness scale (see Appendix B). In most of these cases, the plateau phase of the sigmoid fit was still reached at the highest stimulus levels, as was the case for electrode contact 7 of participant A005 of which the results of the different threshold detection methods were shown in Figure 3.2 to 3.5. On the other hand, for electrode contact 13 of participant A008, the compliance limit was only scored a 4 on the 8-points loudness scale. Figure 3.6 shows the results of the different threshold detection methods (in combination with FG FM) for this specific electrode contact. As can be seen from this figure, all threshold detection methods, except for the cross-covariance between adjacent levels method, failed to reach the plateau phase of the sigmoid fit, and subsequently failed to determine a valid eCAP threshold. Even the linear AGF method was unsuccessful for this electrode contact. In clinical practice however, a clinical specialist would still want to be able to make a fitting for this electrode contact and the compliance limit of this electrode contact cannot be changed. Additional research is therefore necessary into other, non-sigmoidal fits and/or other automatic eCAP threshold detection methods when the compliance limit of an electrode contact is way below the actual M-level.

Figure 3.7 displays the different automatic threshold detection methods (in combination with FG FM) on electrode contact 5 of participant A_013. For this specific electrode contact, eCAP measurements were performed until the M-level. The plateau phase of the sigmoid fit was reached for all threshold detection methods, except for the AGF method. Even though the plateau phase was not reached, the sigmoid AGF threshold detection method still succeeded in determining a valid eCAP threshold (as did the linear AGF method). This example demonstrates that a full sigmoid function is not always necessary to be able to determine a valid eCAP threshold and raises the question whether other fits are possible for the AGF as well and are broadly applicable. Skidmore et al. (26) did not study the different fits of AGFs to determine eCAP thresholds, but they did study different methods to calculate the slope of an AGF. They developed a new method to determine the slope of an AGF based on sliding window linear regression. This newly developed method was less influenced by the morphology of the eCAP AGF, and reaching the plateau phase was thus not a prerequisite to determine a valid slope. Future research should focus on how this newly developed method for slope determination of AGFs can be used for eCAP threshold determination based on AGFs.

The cross-covariance methods were not published before for automatic detection of eCAP thresholds and therefore the asymptote method was chosen to determine the threshold. The cross-covariance method between adjacent levels has been used for automatic threshold detection in auditory brainstem responses. In the study by Suthakar et al. (16), a cut-off value of 0.35 was chosen for threshold determination by comparing visual thresholds to automatically determined thresholds. This value is comparable with the asymptote values of the cross-covariance between adjacent levels in combination with SA and with the asymptote values of all combinations of the cross-covariance with the maximum level method, but not with the asymptote values from the FG combinations of the cross-covariance between adjacent levels method (which are higher) as can be seen from Table 2.1. Furthermore, in the study from Suthakar et al. (16), the chosen value of 0.35 is higher than the asymptote value of their fitted sigmoid functions. Further research into the optimal and most correct cut-off value for cross-covariance threshold detection methods in eCAP measurements is recommended.

The five different automatic threshold detection methods used in this study only present a part of all the possible automatic threshold detection methods and can probably be optimised further.

4.3. Success rates

For all automatic threshold detection methods, the combinations with the FM artefact reduction method had higher success rates than the combinations with the AP artefact reduction method. This difference became especially clear in both cross-covariance threshold detection methods. As can be seen from Figure 3.4, for the AP combinations of the cross-covariance between adjacent levels threshold detection method, the fitted sigmoid function does not seem to reach a lower asymptote (entirely), which in turn complicated the threshold determination. This effect was observed in the majority of electrode contacts/participants. For the cross-covariance with maximum level threshold detection method, this effect is not visible in Figure 3.5, but it was visible in other electrode contacts/participants (but less frequent than for the cross-covariance between adjacent levels method). The missing of the lower asymptote can be a possible explanation for the lower success rates of the AP combinations for the cross-covariance methods. The mean asymptote values for the electrode contacts in which a (valid) threshold could be determined are however comparable between the AP and FM combinations as can be seen in Table 2.1. Using a fixed asymptote value will therefore not result in higher success rates per se. A different fit and/or threshold detection method for the combinations of AP and cross-covariance could possibly lead to higher success rates, or the combination of AP and cross-covariance is just a worse performing combination, independently of the fit method.

The combinations with the FG averaging method had higher success rates than the combinations with the SA averaging method for all automatic threshold detection methods except for the SNR. This can partially be explained by the fact that the FG combination had more stimulus levels than the SA combinations (which had only 10 stimulus levels), and thus the functions of the FG combinations were fitted on more measurements and more measurements could potentially lead to better fits.

For the AGF threshold detection methods, there was a discrepancy in success rates between the linear and sigmoid method. The success rates of the SA AP, SA FM and FG AP combinations were higher for the sigmoid AGF method than for the linear AGF method, this difference is particularly clear for the SA combinations. A possible explanation for this difference is the selection of measurements to construct the fit for the linear AGF, whereas for the sigmoid AGF all measurements are used to construct the fit. This selection for the linear AGF fit is not yet perfect and could be optimized further. Besides, as explained before, the SA method has less measurements that the FG method and to construct a linear AGF at least three measurements should fulfil the selection criteria. As the SA method only has ten measurements, the chances of three measurements that fulfil the selection criteria are lower than for the FG method which has more measurements.

Overall, the SNR threshold detection method seems to have the highest success rates for all combinations of artefact reduction methods, with the highest success rate of 92% for the combination of SA and FM. This success rate is comparable to the success rate that was found in the study of Hoth et al. (15). They found a success rate of 95% for the SNR method in combination with SA and AP. The success rates of the other automatic threshold detection methods are somewhat lower than those of other automatic threshold detection methods that were found in the review. There were however studies included in this review suffering from selection bias, by selecting patients, ears and/or electrode contacts on beforehand based on good eCAP responses. The current study does not suffer from selection bias, there was no prior selection based on eCAP responses.

The success rates are important to keep in mind when analysing the results of the precision and accuracy of the thresholds. This study initially determined eCAP thresholds on 105 (104 for FG) electrode contacts, but TCIs could only be calculated when a valid eCAP threshold could be determined

and correlations could only be determined when at least three valid eCAP thresholds could be determined for a participant.

4.4. Precision of thresholds

The two-way ANOVA of the normalised TCIs (in dB) revealed that both the artefact reduction methods and the threshold detection methods had a significant effect on the mean TCIs. The combination of FG FM resulted in the highest precision of the threshold, and the sigmoid AGF method was the best automatic eCAP threshold detection regarding the precision. When comparing the 20 different combinations that have been examined in this study, the combination of FG AP with sigmoid AGF had the highest precision. This precision was significantly better than the precision of all other combinations, except from the combinations: FG AP with sigmoid AGF, FG AP with cross-covariance with maximum level and FG FM wit cross-covariance with maximum level.

TCIs were calculated per electrode contact for every participant, but the analysis was performed on all TCIs together, as a threshold detection method should be precise on all electrode contacts of all CI users. Nevertheless, the different participants and different electrode contacts can have an effect on the TCIs, and this effect was not incorporated into the two-way ANOVA. To correct for these possible effects, a linear mixed model can be used. In this model, the TCI would be the response variable, the artefact reduction methods, the threshold detection methods and the possible interaction between them would be the predictor variables or fixed effects (as the goal is to study the effects of these methods on the TCI), and the participants and electrode contacts would be the grouping variables or random effects (to correct for the differences in variances induced by these variables). The formula of this linear mixed model would be as follows:

$$\begin{aligned}\text{TCI} \sim \text{threshold detection method} * \text{artefact reduction method} + (1|\text{participant}) \\ + (1|\text{electrode contact}) + (1|\text{electrode contact: participant})\end{aligned}$$

An initial linear mixed model was made, without the grouping variable electrode contact, but this model was not completed, as the amount of data that was collected was relatively small in comparison to the complexity of the model. In addition, as all the fixed effects are categorical variables, some form of dummy coding should be used as well. The initial model did reveal a significant contribution from the random effect participant, meaning that the different participants partially explain the differences in TCI. Thus, more data should be collected to construct a complete linear mixed model and to examine whether this changes the analysis of the TCIs.

From the studies included in the previous review, only one study determined the precision of eCAP thresholds. Gärtner et al. (13), who used FG and AP in combination with the SNR threshold detection method, determined the precision of eCAP thresholds by means of the test-retest variability. They found a mean TCI of 0.59 nC, but they did not normalise their TCIs and therefore this mean cannot be directly compared to the TCIs of this study. Biesheuvel et al. (11) also studied TCIs as a measure of precision of eCAP thresholds. They used the combination of SA FM and a semi-automatic linear AGF threshold detection. The calculation of their TCIs was also based on the 95% bounds of the linear fit, but instead of just extrapolating these confidence bounds towards the x-axis (as has been done in this study), they calculated the steepest and least steep linear fits within these confidence bounds and determined the intersection of these fits with the x-axis. Because of this different method and the fact that their TCIs were not normalised, their TCIs cannot be directly compared to the TCIs of this study. Nevertheless, it would be interesting to study the differences between the methods to determine the TCIs, and whether these different methods result in different TCIs.

4.5. Accuracy of thresholds

As mentioned before, there are substantial measuring differences between behavioural measurements and eCAP measurements, as behavioural measurements make use of tone bursts at higher stimulation rates. The lower stimulation rates used for eCAP measurements are necessary to obtain measurable eCAPs. Increasing the stimulus rate will make the nerve fibres enter their refractory periods and thus the neurons will fire less synchronously (27). Therefore, in this study, it was chosen to calculate the correlation between T-levels measured as in normal clinical practice in SW and eCAP thresholds, and the correlation between T-levels measured with eCAP stimuli and eCAP thresholds.

The one-way ANOVA to compare the Fisher transformed correlation coefficients between these two types of correlation revealed a significant difference, the correlation coefficients of the correlation between T-levels measured with eCAP stimuli and eCAP thresholds were significantly higher than the correlation coefficients of the correlation between T-levels measured in SW and eCAP thresholds. Besides, the (non-transformed) correlation coefficients of the correlation with T-levels measured in SW seemed to cover the whole range between -1 and 1, as can be seen from Figures 3.12 to 3.16. The two-way ANOVA of the correlation with T-levels measured in SW showed that nor the artefact reduction method, nor the threshold detection method, nor the interaction of both had a significant effect on the Fisher transformed correlation coefficients. Therefore, the correlation between T-levels measured in SW and eCAP thresholds seemed rather random and is probably not a good measure of accuracy. The correlation coefficients of the automatic threshold detection methods included in the previous review cannot be compared directly to the correlation coefficients of this study, as most of the studies included in the review suffered from analytical bias by performing grouped correlations.

The correlation between T-levels measured with eCAP stimuli and eCAP thresholds may be a better measure of the accuracy of eCAP thresholds. The two-way ANOVA of the Fisher transformed correlation coefficients of this correlation did reveal a significant effect of artefact reduction method, threshold detection method and the interaction of both on the mean correlation coefficients. The FG FM combination resulted in the highest accuracy, but only significantly higher compared to the FG AP combination. Likewise, the sigmoid AGF method was the best automatic eCAP threshold detection regarding the accuracy, but only significantly better than the two cross-covariance methods. When comparing all 20 different combinations of methods, the FG AP with sigmoid AGF combination produced the highest accuracy, but only significantly higher than the combinations: FG FM with linear AGF, SA AP with SNR, FG AP with SNR, SA AP with cross-covariance between adjacent levels, FG AP with cross-covariance between adjacent levels and FG AP with cross-covariance with maxim level.

The correlation between T-levels measured with eCAP stimuli and eCAP thresholds has not been examined in previously published studies, and thus no comparisons can be made with other automatic threshold detection methods.

As previously stated, a lot of other studies suffered from analytical bias by only performing a grouped correlation. This bias was not present in the current study, correlation coefficients were calculated separately for each participant. Nonetheless, the difference between participants can have an effect on the correlation coefficients. Similar as for the TCIs, linear mixed models can be used here as well, but with the extra fixed effect of T-level measurement method and without the random effect electrode contact, as correlations are calculated over the electrode array of every participant. The initial linear mixed model did not reveal a significant contribution of the random effect participant to account for the differences observed between correlation coefficients. Nonetheless, when more data is collected, it would be interesting to complete the linear mixed model and to study whether the contribution of the different participants changes.

The question remains whether the correlation between T-levels and eCAP threshold is the best measure of accuracy. Clinical fitting of CIs nowadays is based on these T-levels, and the goal of the fitting is to reach optimal speech comprehension. If the subject of interest is CI fitting based on eCAP thresholds, the correlation between speech comprehension and eCAP thresholds could be a measure of accuracy as well. Smoorenburg et al. (28) studied the effect on speech comprehension when fitting a CI based on eCAP thresholds, but the exact effect remained unclear.

5. Conclusions

This study examined the success rates, precision and accuracy of five different automatic eCAP threshold detection methods: sigmoid AGF, linear AGF, SNR, cross-covariance between adjacent levels and cross-covariance with maximum level. These threshold detection methods have been studied using a combination of two different averaging methods: SA and FG, and two different artefact reduction methods: AP and FM. In total, 20 different combinations of methods have been examined.

Based on the success rates, the FG FM method is the best in processing the eCAP data and the SNR method is the best automatic eCAP threshold detection method.

From the TCI analysis, it can be concluded that the FG FM combination resulted in the highest precision. The sigmoid AGF was the automatic eCAP threshold detection method with the highest precision.

From the analysis of the correlation coefficients, it can be concluded that the correlation between eCAP thresholds and T-levels measured with eCAP stimuli was higher than the correlation between eCAP thresholds and T-levels measured in SW. From the analysis of the correlation coefficients with T-levels measured with eCAP stimuli, it can be concluded that the FG FM resulted in the highest accuracy. The sigmoid AGF was the automatic eCAP threshold detection method with the highest accuracy.

References

1. Dobie RA, Van Hemel S, Council NR. Basics of sound, the ear, and hearing. Hearing Loss: Determining Eligibility for Social Security Benefits: National Academies Press (US); 2004.
2. Vaerenberg B, Smits C, De Ceulaer G, Zir E, Harman S, Jaspers N, et al. Cochlear implant programming: a global survey on the state of the art. ScientificWorldJournal. 2014;2014:501738.
3. Mens LH. Advances in cochlear implant telemetry: evoked neural responses, electrical field imaging, and technical integrity. Trends Amplif. 2007;11(3):143-59.
4. Garrido LC, Schwarz K, Lauss K, Vernetta CD, Kontides A, Gomez MD, et al. Comparison of a Traditional and Novel Evoked Compound Action Potentials Recording Approach and Evoked Auditory Brainstem Responses in Pediatric Cochlear Implants Users. Journal of International Advanced Otology. 2018;14(3):353-8.
5. Gärtner L, Lenarz T, Büchner A. Fine-grain recordings of the electrically evoked compound action potential amplitude growth function in cochlear implant recipients. Biomed Eng Online. 2018;17(1):140.
6. Baudhuin JL, Hughes ML, Goehring JL. A Comparison of Alternating Polarity and Forward Masking Artifact-Reduction Methods to Resolve the Electrically Evoked Compound Action Potential. Ear and Hearing. 2016;37(4):E247-E55.
7. He S, Teagle HFB, Buchman CA. The Electrically Evoked Compound Action Potential: From Laboratory to Clinic. Front Neurosci. 2017;11:339.
8. Herrmann DP, Kretzer KVA, Pieper SH, Bahmer A. Effects of electrical pulse polarity shape on intra cochlear neural responses in humans: Triphasic pulses with anodic and cathodic second phase. Hear Res. 2021;412:108375.
9. Hey M, Muller-Deile J. Accuracy of measurement in electrically evoked compound action potentials. Journal of Neuroscience Methods. 2015;239:214-22.
10. Hughes ML, Goehring JL, Baudhuin JL. Effects of Stimulus Polarity and Artifact Reduction Method on the Electrically Evoked Compound Action Potential. Ear Hear. 2017;38(3):332-43.
11. Biesheuvel JD, Briaire JJ, Frijns JHM. The Precision of eCAP Thresholds Derived From Amplitude Growth Functions. Ear Hear. 2018;39(4):701-11.
12. de Vos JJ, Biesheuvel JD, Briaire JJ, Boot PS, van Gendt MJ, Dekkers OM, et al. Use of Electrically Evoked Compound Action Potentials for Cochlear Implant Fitting: A Systematic Review. Ear Hear. 2018;39(3):401-11.
13. Gärtner L, Spitzer P, Lauss K, Takanen M, Lenarz T, Hoth S. Optimized SNR-based ECAP threshold determination is comparable to the judgement of human evaluators. PLoS One. 2021;16(11):e0259347.
14. Estienne P, Scaglia A, Kontides A, Lauss K, Schwarz K, Arauz SL. Comparison of automated and traditional ECAP recording approaches in clinical practice. Int J Audiol. 2022;61(7):583-91.
15. Hoth S, Spitzer P, Praetorius M. A new approach for the determination of ECAP thresholds. Cochlear Implants Int. 2018;19(2):104-14.
16. Suthakar K, Liberman MC. A simple algorithm for objective threshold determination of auditory brainstem responses. Hear Res. 2019;381:107782.
17. van der Beek FB, Briaire JJ, Frijns JH. Population-based prediction of fitting levels for individual cochlear implant recipients. Audiol Neurootol. 2015;20(1):1-16.
18. Telmesani LM, Said NM. Electrically evoked compound action potential (ECAP) in cochlear implant children: Changes in auditory nerve response in first year of cochlear implant use. Int J Pediatr Otorhinolaryngol. 2016;82:28-33.
19. Spivak L, Auerbach C, Vambutas A, Geshkovich S, Wexler L, Popecki B. Electrical compound action potentials recorded with automated neural response telemetry: threshold changes as a function of time and electrode position. Ear Hear. 2011;32(1):104-13.
20. Brown CJ, Abbas PJ, Etlert CP, O'Brient S, Oleson JJ. Effects of long-term use of a cochlear implant on the electrically evoked compound action potential. J Am Acad Audiol. 2010;21(1):5-15.

21. Molisz A, Zarowski A, Vermeiren A, Theunen T, De Coninck L, Siebert J, et al. Postimplantation changes of electrophysiological parameters in patients with cochlear implants. Audiol Neurootol. 2015;20(4):222-8.
22. Bahmer A, Baumann U. Application of triphasic pulses with adjustable phase amplitude ratio (PAR) for cochlear ECAP recording: I. amplitude growth functions. J Neurosci Methods. 2012;205(1):202-11.
23. Charasse B, Thai-Van H, Chanal JM, Berger-Vachon C, Collet L. Automatic analysis of auditory nerve electrically evoked compound action potential with an artificial neural network. Artif Intell Med. 2004;31(3):221-9.
24. Estienne P, Scaglia A, Kontides A, Lauss K, Schwarz K, Arauz SL. Comparison of automated and traditional ECAP recording approaches in clinical practice. Int J Audiol. 2021:1-9.
25. van Dijk B, Botros AM, Battmer RD, Begall K, Dillier N, Hey M, et al. Clinical results of AutoNRT, a completely automatic ECAP recording system for cochlear implants. Ear Hear. 2007;28(4):558-70.
26. Skidmore J, Ramekers D, Colesa DJ, Schvartz-Leyzac KC, Pfingst BE, He S. A Broadly Applicable Method for Characterizing the Slope of the Electrically Evoked Compound Action Potential Amplitude Growth Function. Ear Hear. 2022;43(1):150-64.
27. Koch DB, Overstreet EH. Neural Response Imaging: Measuring Auditory-Nerve Responses from the Cochlea with the HiResolution™ Bionic Ear System. Advanced Bionics Corporation, Bengaluru. 2003.
28. Smoorenburg GF, Willeboer C, van Dijk JE. Speech perception in nucleus CI24M cochlear implant users with processor settings based on electrically evoked compound action potential thresholds. Audiol Neurootol. 2002;7(6):335-47.

Appendix

A. Literature review on automatic eCAP threshold detection methods

This review was written for the course TM30003. This review has already been graded and should not be graded again. The review is inserted as background information.

Automatic detection of eCAP thresholds: A systematic review

Eleen Schupp

Objectives. Besides restoring speech perception, cochlear implants (CI) can also record the response of the auditory nerve to a stimulus via its telemetry function. This reaction is called the electrically evoked action potential (eCAP). The fitting of a CI is crucial in obtaining good speech perception for a patient, but it can be a time-consuming process which often requires active cooperation of a patient. eCAP measurements can aid in the fitting and monitoring of CIs via eCAP thresholds. The eCAP threshold is defined as the minimum amount of current that evokes a measurable eCAP response. There is currently no golden standard for the determination of these eCAP thresholds, and often the judgment of an experienced clinical specialist is required. Automatic detection of eCAP thresholds could further elaborate the use of eCAP measurements in fitting and monitoring of a CI. Different methods for automatic detection have already been designed and tested in clinical practice. The goals of this review was to create an overview of different automatic detection methods of eCAP thresholds.
Methods. A literature search was performed in the electronic databases PubMed, Embase and Web of Science. The search strings were constructed with the terms CI, eCAP, automatic and their derivatives. The studies were assessed for eligibility and relevant data concerning eCAP threshold detection methods was extracted.
Results. 14 studies were included in this review. All threshold detection methods (standard and automatic) can be subdivided into mainly two categories: amplitude growth function (AGF) and last visible. Most used outcome measures to evaluate automatic threshold detection methods are success rates, correlations of automatic methods with standard methods and behavioural measurements and differences between standard and automatic methods. Outcome measures can suffer from different types of bias; selection bias or analytical bias (no individual correlations or no indication of significance).
Conclusions. Based on the limitations of the outcome measures that have been researched in the literature, no conclusions are drawn regarding which automatic threshold detection methods perform better or worse. This review solely gives an overview of the different automatic threshold detection methods that have been described in literature thus far. Additionally, it demonstrates that further research should focus on different outcome measures, namely noninferiority of the automatic methods, individual correlations and the precision of the different automatic threshold detection methods.

Introduction

Cochlear implants

A cochlear implant (CI) is an auditory prosthesis used to restore hearing in people with severe to profound hearing loss. A CI consists of outer parts including the microphone and speech processor to capture and code the sounds, and a coil. This external coil transmits the coded signals to the internal coil, which in turns transmits these electrical signals to the cochlea via its multiple-channel electrode array. In the cochlea, the electrical stimulus generates action potentials that propagate along the auditory nerve and can lead to auditory perception and speech comprehension.

The fitting of a CI is crucial in obtaining good speech perception for a patient. There is currently no uniform fitting, which could lead to variances if fitting is depending on the clinical specialists. Besides, fitting is often a time-consuming process which requires the active cooperation of a patient (1). Fitting of a CI is now often based on behavioural measurements. The behavioural threshold (lowest stimulus level for which a patient indicates it is audible) is called the T-level, the maximum level that is still comfortable for a patient is called the M-level or C-level. It would be beneficial if a CI could be fitted based on objective measurements, this could lead to more uniform fittings across clinical specialists and it can be performed without a patient's active cooperation (which would be especially useful in for example pediatric patients).

Electrically evoked compound action potentials

Besides transmitting electrical stimuli, a CI can also record the response of the auditory nerve to a stimulus via its telemetry function. This reaction is called the electrically evoked compound action potential (eCAP), which represents the neural response from multiple auditory nerve fibers to an electrical stimulus.

An eCAP occurs within roughly one millisecond after the stimulation pulse and typically contains a negative peak N1 and a positive peak P1. The difference between those two peaks is the eCAP amplitude, which varies between 50 and 1500 μV on average. Besides the eCAP, a recording also contains signal noise, recording artefacts and stimulus artefacts (2). Signal noise is reduced by averaging and the recording artefact can be eliminated by template subtraction (recording with zero amplitude). To reduce the stimulus artefact, different artefact reduction methods exist, such as alternating polarity, forward masking and triphasic pulses (3, 4). Artefact reduction methods exceed the scope of this review and will therefore not be explained in more detail, however, other studies have demonstrated significant differences regarding eCAP thresholds between different artefact reduction methods. There is no clear consensus on which artefact reduction method results in the most precise and accurate eCAP threshold (2-7).

eCAP measurements can be used intra-operatively to ensure a good placement of the electrode, and could post-operatively be used for, among other things, monitoring and fitting (4, 8).

An important advantage of eCAP measurements which make them beneficial for use in clinical practice, is that no extra equipment is involved (besides from the components of the CI that are already accessible); no external electrode is needed, and software is provided by the manufacturer. eCAP measurements can be performed for all types of CIs, but each manufacturer has its own terminology: it is called Neural Response Imaging (NRI) by Advanced Bionics (AB) (Valencia, CA, USA), Neural Response Telemetry (NRT) by Cochlear (Sydney, Australia) and Auditory Response Telemetry (ART) by MED-EL (Innsbruck, Austria).

eCAP threshold

From eCAP measurements, the eCAP threshold can be estimated, which is defined as the minimum amount of current that evokes a measurable eCAP response. This eCAP threshold could potentially be used in fitting of CIs, as an addition to behavioural measurements. The correlation between behavioural measurements and eCAP thresholds has been widely studied, but to date no strong correlations are found (4, 8, 9).

There are different eCAP threshold detection methods. Thresholds can be determined by clinical specialists or automatically via algorithms. Determining eCAP thresholds by clinical specialists is more time-consuming and can lead to a wider distribution of the results as each clinical specialists develops his own method and uses different criteria for threshold detection. Therefore, automatic threshold detection would be advantageous and could potentially lead to higher correlations with behavioural measurements.

Goals and objectives

The goal of this review is to provide an overview of different automatic eCAP threshold detection methods.

Methods

Search strategy

The electronic databases PubMed, Embase and Web of Science were searched on the 23rd of June 2022. Specific search strings were made for the different databases with the cooperation of an experienced librarian, all search strings contained the following components (with their derivatives and synonyms): cochlear implants, eCAP and automatic. Only publications in English and Dutch and published in the past 20 years were included. The complete search strings can be found in appendix A.

Eligibility criteria and study selection

The author of this review identified the eligibility of the obtained articles. Only original articles, including reviews and letters to editor containing original data were included. Articles were excluded if they did not involve CIs, if the performed measurements were not eCAP measurements and did not involve threshold detection, if the threshold detection method was not (fully) automatic or when there was uncertainty whether the method was automatic. Since automatic threshold detection is the topic of this review, this was the most important inclusion criterion. For an article to be included in this review, the methods section should clearly describe which automatic threshold detection method was used. The first screening was made based on title and abstract and next full-text articles were assessed for eligibility. If full texts of publications were not available, authors were contacted inquiring the possibility of receiving the article via email, articles were excluded if the author could not be reached.

Data extraction/collection

The reviewer screened the data from the included studies. The following data was extracted from the included articles: type of CIs, measurement details, details/description of the automatic threshold detection method and the results of these methods.

Evaluation

The goal of this review is to create an overview of different automated threshold detection methods, to explore which methods are already tested and demonstrated promising results, or which methods can be refined further or need further testing. Therefore, there will be no evaluation to define one best method.

Results

Included articles

The results of the electronic database searches are summarized in the flow diagram shown in figure 1. The searches in PubMed, Embase and Web of Science produced a total of 333 publications (156, 60 and 117 respectively). After removing of duplicates, there were 184 publications remaining, which were all screened based on title and abstract. After this first screening, 39 full-text articles were assessed for eligibility, of which 14 were included in this systematic review. One of those articles (Botros et al., 2007) (10) does not contain own data/results but does contain an extensive description of an automated detection method and is therefore included in this review.

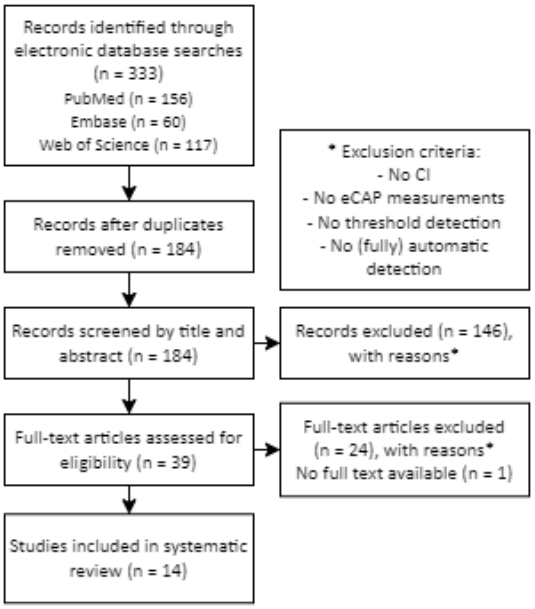


Figure 25. Flow diagram of literature search and study selection.
CI = cochlear implant, eCAP = electrically evoked compound action potential

Data extraction

Data was collected from the remaining 13 articles (the article without original data is not included in the overview). In appendix B table B.1, summarised data can be found about the CI type, the timing of the measurements, the tested electrodes, the artefact reduction method and the threshold detection methods. The different (automatic) threshold detection methods are described in detail below and summarised in table 1. The results of these different automatic threshold detection methods are depicted in table 2.

Threshold detection methods

For both non-automatic and automatic threshold detection methods, there are two major techniques: last visible and linear extrapolation or amplitude growth function (AGF). In the last visible technique, the eCAP traces are ordered on stimulus intensity, and the clinical specialist or the algorithm picks the trace with the lowest stimulus intensity in which a clear eCAP can still be distinguished. For the AGF, the eCAP amplitudes are calculated and plotted in function of the stimulus levels. This plot is then linearly extrapolated, and the intersection with the x-axis is labeled as the threshold. Both techniques have their advantages and disadvantages. A disadvantage of the last visible technique is that the threshold will always be above the noise floor of the measurement system, as responses within this noise floor will not be visible. The noise floor of the NRI system is approximately 20 µV peak to peak (11). An advantage of the last visible technique over the AGF technique is that fewer (suprathreshold) measurements are required. In order to be able to make a linear extrapolation for the AGF technique, at least three suprathreshold measurements are needed (12). This can lead to a longer duration of the measurements, and in some patients suprathreshold measurements are close to the patient's maximum acceptable loudness level, which can make it challenging to record enough suprathreshold measurements. Another disadvantage of the AGF technique, is that the AGF is often non-linear (especially near the threshold), which can make a linear extrapolation less reliable. An advantage of the AGF technique over the last visible technique, is that the AGF technique relies less on the stimulus levels and step sizes, while thresholds defined by the last visible technique can only be equal to one of the stimulus levels.

When a clinical specialist handpicked the N1 and P1 peaks or when an algorithm automatically picks the peaks but manual adjustments by clinical specialists are applied, this technique will be called standard AGF for the continuation of this review.
When eCAP thresholds were determined by a clinical specialist via the last visible technique, these were simply referred to as 'last visible'.

FineGrain and AutoART
The threshold determination in the FineGrain (FG) method, developed originally for MED-EL CIs, is based on the AGF technique, but what distinguishes the FG method is the linear extrapolation. As mentioned above, the AGF is not completely linear. In FG, a sigmoidal function is fitted to the AGF, and a linear extrapolation is made of the inflection point (steepest slope of sigmoid function). The threshold is determined as the intersection of this extrapolation with the x-axis (13). Another important characteristic of the FG method is the averaging method. In normal averaging, multiple measurements are made at the same stimulus level and the resulting eCAP trace is the average over all these measurements. As multiple measurements prolong the measuring time, measurements are usually restricted to a limited number of stimulus levels with relatively bigger step sizes. The FG method makes use of averaging as well, but instead of averaging over multiple measurements at the same stimulus levels, it averages the measurements within a small range of stimulus levels. As a result, during the same measuring time, eCAP measurements can be made at more stimulus levels with a smaller step size. Consequently, this leads to a higher resolution of the AGF, which in turn can improve the accuracy of the threshold determination (14). The FG method can determine the threshold fully automatically by means of peak-picking algorithms.

AutoART (developed by MED-EL), is based on the FG method. The only difference between FG and AutoART, is that FG requires a manual stop of the measurements, while AutoART automatically stops once an ECAP threshold is reached (15). As the topic of this review is automatic threshold detection methods, but not necessarily automatic measurement methods, FG and AutoART are considered as the same technique in the remainder of this review.

NRT-ANN
The NRT-ANN method involves the NRT technique of Cochlear combined with an artificial neural network (ANN). The constructed ANN consist of 30 input neurons and 5 output neurons, as every NRT trace is made of 30 points and there are five different possible patterns of traces. These five patterns are the following: clear response, both N1 and P1 peaks are identifiable; clear response, but N1 peak is missing; clear response, both N1 and P1 peaks are identifiable, the first point of the response being upside maximum of P1 peak; clear response with prominent P1 in the absence of N1 peak; non-recordable response. NRT-traces of the first and third type of pattern were extracted, an automatic peak picker algorithm detected N1 and P1 peaks. Next, the AGF technique is applied to determine the threshold (16).

AutoNRT
The AutoNRT method, developed by Cochlear, is based on the last visible technique for threshold determination and can measure eCAP thresholds automatically. AutoNRT consists of a peak-picking algorithm to define the N1 and P1 peaks combined with a decision tree to decide whether an eCAP is visible in a given trace. Both components are machine-learned. The eCAP measurements in AutoNRT consist of two series: an ascending and a descending series. The measurements start with the ascending series, stimulus level is increased until an eCAP is detected by the decision tree in two consecutive traces. Thereafter, the descending series is initiated, stimulus level is decreased (in smaller steps than in the ascending series) until an eCAP is no longer detected. The eCAP threshold is defined as the mean of the lowest stimulus level (of the descending series) for which an eCAP was detected

and the highest stimulus level (of the descending series) for which no eCAP could be detected. If these lowest and highest stimulus levels in which an eCAP is detected are more than 10 current levels apart, the algorithm does not determine a threshold because in this case the estimate would be unreliable (10, 17, 18).

(FG-)SNR

As the name suggests, the FG signal-to-noise (FG-SNR) method is partially based on the FG method. The eCAP measurements are conducted according to the FG method and a AGF is made. However, to construct this AGF, the SNR ratio (instead of the eCAP amplitude) is plotted in function of the stimulus level. Because the SNR ratio reflects on whether a trace contains a neural response or not, the theory behind the FG-SNR method is rather based on the last visible technique than on the AGF technique. To determine the SNR ratio, the eCAP measurements are split into two time windows: one time window with the actual eCAP and residual noise, the other window only contains noise. Next, the variance is calculated in these two time windows, and the SNR ratio is defined as the variance of the time window with the eCAP and noise divided by the variance of the time window with noise only. As explained above, this SNR ratio is plotted in function of the stimulus level, a linear fit is applied, and the threshold is defined as the crossing of this function with a horizontal line at a certain SNR ratio. Different levels of this horizontal line were researched, a SNR threshold of 6 dB led to the best results (highest correlation with eCAP thresholds determined by experts) (19). Hoth et al. (20) researched the SNR method as well, but they did not make use of the FG method to record the eCAP measurements.

SmartNRI

SmartNRI is an automated method of the NRI technique of AB. The statistical method used in SmartNRI assumes that all measurements consist of neural responses, noise and residual artefact. Firstly, the noise component is reduced using a principle-component analysis. Afterwards, the NRI traces are reconstructed using a linear sum of seven basic functions. In addition, a best-fit residual artefact model was constructed (based on an analysis of over 1000 NRI traces). The reconstructed NRI traces are compared to this artefact model, any trace that differs significantly from the model possibly contains a real neural response. Next, a strength-of-response (SOR) metric is computed to quantify how much a trace deviates from the artefact. A trace is assumed to contain a neural response if its SOR exceeds a certain criterion value (27 appears to be a good value based on previous research) (21, 22). Since this method is based on whether a trace contains a neural response or not, it can be considered as a variant of the last visible technique.

Peak-picking algorithms

Some automatic threshold detection methods do not make use of adaptive measurements (adapt the measurements when an eCAP is detected) as several methods described above do. Instead, measurements are performed in a fixed manner and each trace is analysed separately with a peak-picking algorithm, these peak-picking algorithms are often the same as the ones used in the other methods.

Advanced NRT, developed by Cochlear, uses the same peak-picking algorithm as in AutoNRT. In this algorithm, N1 is defined as the minimum of the first 8 samples and P1 is the maximum of the samples after N1, up to and including sample 16 (10, 12).

Similar for AB devices, the NRI peak-picking algorithm defines N1 as the first most negative peak in the range of 270 to 670 µsec and P1 as the following most positive peak in the range of 370 µsec or N1 time + 100 µsec (whichever is greater) to 1000 µsec. In the NRI peak-picking algorithm, only traces with a peak-to-peak amplitude higher than the noise floor, 20 µV for AB devices, are included in further analyses (12).

Once the peaks are defined, the eCAP threshold can be defined by means of the AGF technique or the last visible technique. The AGF threshold is determined by simple linear regression of the traces that were included in the analysis. The last visible threshold is defined as the lowest level where the algorithm picked peaks of the traces that were included in the analysis.

The automated threshold detection method that makes us of a peak-picking algorithm in combination with AGF will be called peak-picking AGF in this review, the combination of a peak-picking algorithm and the last visible technique will be called peak-picking last visible.

Outcome measures and possible bias

The outcome measures that were calculated in the included studies and summarised in table 2 are success rate, correlation (with standard threshold detection methods and with behavioural measurements) and difference (between standard threshold detection methods and automatic threshold detection methods).

Success rate

Success rate is defined as the percentage of the total measured electrode contacts in which an eCAP threshold could be defined. Bias can be present in this outcome measure when patients, ears and/or electrode contacts are selected on beforehand based on good eCAP responses. One study that determined the success rates, Chauhan et al. (23), suffered from this type of selection bias. The other studies that determined the success rates had no indications of selection bias.

Correlation

Two correlations can be calculated: the correlation between eCAP thresholds determined via standard threshold detection methods and eCAP thresholds determined via automatic threshold detection methods and the correlation between eCAP thresholds determined via automatic threshold detection methods and behavioural measurements (T- and C-levels). As with the success rates, selection bias is possible for correlations as well. Four studies; Glassman et al. (12), Jeon et al. (11), Chauhan et al. (23) and Akin et al. (24) are affected by the selection bias. Another type of bias that can occur when determining correlations is an analytical bias. This bias can be induced when group correlations are performed. Group correlations can suppress the possible variations of individual correlations, while these individual correlations are important to investigate eCAP-based fitting. This form of analytical bias is known as the Simpson's Paradox and has also been described in the review by de Vos et al. (8). Only two studies of the included articles in this review determined individual correlations, namely Gärtner et al. (19) and Hoth et al. (20). Another type of bias is caused by the fact that only measurements for which both measurement methods (standard method and automatic method) agreed on the presence of an eCAP threshold are used for the calculation of the correlation. This bias in inevitable and therefore present in all studies calculating correlations, as you cannot determine a correlation between an existing and non-existing threshold. However, it is still important to bare this bias in mind when evaluating the correlations. When the percentage of agreement on which measurements contain an eCAP threshold is small, correlations are only calculated for this small percentage. As all studies suffered from this bias, it is not indicated in table 2.

Difference

Difference simply indicates the (absolute) difference between eCAP thresholds determined with standard and automatic methods. Same as for the other outcome measures, the outcome measure difference suffers from selection bias as well. Differences can suffer from a type of analytical bias as well, when there is no indication of significance (with for example p-values, standard deviations or confidence intervals) of this difference. Four studies; Garrido et al., 2018 (14), van Dijk et al., 2007 (18), Gärtner et al., 2021 (19) and Gärtner et al., 2021 (19) suffered from this type of analytical bias, as indicated in table 2.

Table 1. Overview of threshold detection methods.
AGF = amplitude growth function, NRT = Neural Response Telemetry, FG-SNR = FineGrain Signal-to-noise ratio, NRI = Neural Response Imaging, ART = Auditory Response Telemetry, ANN = artificial neural network

	LAST VISIBLE	**AGF**
ADVANTAGES (+) & DISADVANTAGES (-)	+ Fewer (suprathreshold) measurements required - Threshold can only be equal to one of the stimulus levels - Threshold above noise floor of system	- At least 3 suprathreshold measurements are required + Relies less on stimulus levels and step sizes - AGF is often non-linear
METHODS BASED ON THIS TECHNIQUE	AutoNRT, FG-SNR, SmartNRI	FineGrain and AutoART, NRT-ANN

Table 2. Results of the different automatic threshold detection methods.
FG = FineGrain, AGF = amplitude growth function, ART = Auditory Response Telemetry, NRT = Neural Response Telemetry, ANN = artificial neural network, SNR = Signal-to-noise ratio, NRI = Neural Response Imaging, qu or CU = charge units (1 qu ≈ 1nC), CL = current level ($I[\mu A] = 17.5*100^{(CL/255)}$)
* = Selection bias, ° = analytical bias (no individual correlations), † = analytical bias (no indication of significance)

METHOD	STUDY	SUCCES RATE	CORRELATION		DIFFERENCE BETWEEN STANDARD AND AUTOMATIC METHOD
			WITH STANDARD METHOD	WITH BEHAVIOURAL MEASUREMENT	
FINEGRAIN AND AUTOART	**ESTIENNE ET AL., 2021 (15)**	Success rates: FG: 92%, standard AGF: 89%	Correlation between FG and standard AGF: r=0.94, p<2.2e-16°	Correlation between FG and C-levels: r=0.65, p=0.005 and r=0.53, p=0.042 at electrodes 2 and 6, no statistically significant correlation at electrode 10°	Standard AGF significantly lower than FG (t(47)=11.8, p=1.22e–15) with mean of differences being 1.9 qu
	GARRIDO ET AL., 2018 (14)	Success rates: FG: 95.2%, standard AGF: 90.3%	Correlation between FG and standard AGF: r=0.754, p=2.4e-14°		Standard AGF generally lower than FG†
NRT-ANN	**CHARASSE ET AL., 2004 (16)**		Correlation between NRT-ANN and standard AGF: 0.9<r<0.96°		- No significant difference between NRT-ANN and standard AGF (p=0.03) - Mean difference 3.6 current level units
PEAK-PICKING AGF	**GLASSMAN ET AL., 2013 (12)**		Correlation between peak-picking AGF and standard AGF: r>0.99, p<0.001 for Cochlear; r=0.99, p<0.001 for AB*°		- On average, peak-picking AGF was lower than standard AGF for both devices* - Mean absolute-value differences were 1.10 CL (0.99 standard deviation) for Cochlear, and 3.61 CU (5.31 standard deviation) for AB*
	JEON ET AL., 2010 (11)		Correlation between peak-picking AGF and last visible: r=0.905, p<0.0001*°		- Significant difference between peak-picking AGF and last visible (t=6.48, p<0.0001)* - Last visible on average 23 CU higher than peak-picking AGF*
AUTONRT	**CHAUHAN ET AL., 2021 (23)**	Success rates: AutoNRT: 97.4%, T-levels: 88.64%, C-levels: 89.82%*		- Correlation between AutoNRT and T-levels: r=0.391, p=0.005*° - Correlation between AutoNRT and C-levels: r=0.390, p=0.005*°	
	GÄRTNER ET AL., 2010 (17)	Success rates: AutoNRT: 96.8%, last visible: 89.7–95.2%			

	SCORPECCI ET AL., 2016 (25)	Success rate AutoNRT: 93.3%		- Correlation between AutoNRT and C-levels: r=0.569, p<0.001° - Correlation between AutoNRT and T-levels: r=0.723, p<0.001°	
	VAN DIJK ET AL., 2007 (18)	Success rate AutoNRT: 93%	Correlation between last visible and AutoNRT: r=0.97, p<0.0001°		Median absolute difference between last visible and AutoNRT was 3 CL†
	WESARG ET AL., 2010 (26)		Correlation between AutoNRT and standard AGF: r=0.93, p<0.001°		- AutoNRT about 2 CU higher than standard AGF, not significant (p=0.1)
(FG-)SNR	**GÄRTNER ET AL., 2021 (19)**		Correlation FG-SNR and standard AGF: 0.78<r<0.93		- Agreement FG-SNR and standard AGF on eCAP presence/absence: 64% - FG-SNR on average approximately 2.7 nC higher than standard AGF†
	HOTH ET AL., 2018 (20)	Success rates: SNR: 95%, Standard AGF: 77%		- Correlation between SNR and T-level profiles: -0.01<r<0.92 (mean 0.52) - Correlation between SNR and C-level profiles: 0.26<r<0.84 (mean 0.53)	SNR predominantly higher than standard AGF and not below 5 qu†
SMARTNRI	**AKIN ET AL., 2012 (24)**		Correlation between standard AGF and SmartNRI: r=0.933, p<0.000; r=0.907, p<0.005; r=0.905, p<0.000; and r=0.911, p<0.001, for electrode pairs 3/1, 7/5, 11/9, and 15/13*°		
PEAK-PICKING LAST VISIBLE	**GLASSMAN ET AL., 2013 (12)**		Correlation between peak-picking last visible and last visible: r=0.98, p<0.001 for Cochlear; r=0.89, p<0.001 for AB*°		- On average, last visible was slightly higher than peak-picking last visible for both devices* - Mean absolute-value differences: 3.25 CL (3.32 standard deviation) for Cochlear and 10.64 CU (17.76 standard deviation) for AB*

Discussion

As can be seen from the results in table 1, for the studies who examined the success rates, the success rates of automated threshold detection methods were higher than those of standard methods. Only one study that calculated success rates suffered from selection bias, but the other studies evaluating the same automatic threshold detection method did not. Based on these results, the automatic threshold detection methods of which success rates were calculated, namely FG, AutoNRT and SNR, did not perform worse than standard threshold detection methods.

From the differences that were researched between automatic and standard threshold detection methods, no conclusions can be drawn as some studies found a significant difference, others did not and some studies did not give an indication of significance. Besides, there is no clear golden standard for eCAP threshold detection and therefore questions can be raised regarding between which techniques this difference should be calculated. For this reason, further research can focus on testing noninferiority. Noninferiority studies can assess whether the automatic threshold detection method does not perform worse than a standard threshold detection method when the difference between the two methods does exceed a prespecified margin (27).

Correlation coefficients between the automated methods and the behavioural measurements (T- and C-levels) were rather low and not always significant. This can partially be attributed to the measuring differences; behavioural thresholds are usually measured with pulse trains with a higher stimulation rate. In AB CI's, so-called speech bursts are often used to determine subjective thresholds. Speech bursts typically are delivered to groups of four adjacent electrodes, whereas eCAP stimuli are normally delivered to only one electrode. Akin et al. (28) studied eCAPs recorded at groups of electrodes similarly to speech bursts, but this is outside the scope of this review as it does not concern a fully automatic method. Because of these differences in measuring eCAP thresholds and behavioural thresholds and the fact that all but one study suffered from analytical bias by only showing group correlations, no conclusions are drawn based on these results.

Besides, the correlations with standard threshold detection methods should be handled with some alertness as well. This correlation suffers from analytical bias as well. As with the differences, there is currently no golden standard for eCAP threshold detection to calculate the correlation with and the principle of noninferiority based on differences cannot be used in for this outcome measure.

Hence, other outcome measurements should be examined to evaluate the automated threshold detection methods. For methods to be used in clinical practice, especially in aiding fitting, accuracy and precision are two important parameters. Accuracy is a measure for the nearness of the calculated threshold to the actual value. In this regard, the correlation between the eCAP threshold and the subjective measurements can be deemed as a measure for the accuracy. However, as explained above, there are substantial measuring differences between them and therefore subjective measurements may not be good representatives of the actual value. Residual error approximation may be a better alternative for determining the accuracy, Hey et al. (6) have researched different error approximation methods for eCAP measurements. Precision is the nearness of measurements as obtained with repeated measures, the test-retest variability is a good representative of the precision as demonstrated by Stronks et al. (29). One study included in this review, Gärtner et al. (19) who researched the FG-SNR automatic threshold detection method, determined the precision with this method, their 95% confidence intervals of the eCAP thresholds are maximally ± 1.24 nC wide (mean 0.59, standard deviation 0.21). Biesheuvel et al. (9) created threshold confidence intervals to estimate the precision of eCAP thresholds.

Furthermore, when references were made to a more extensive explanation of an automated method or when referring to a newly tested method, these references were often conference posters or speeches. As these did not always result in a published article, the referred info could not always be retrieved. This raises suspicion about the results of these methods, that they possibly failed in achieving results that could be published. This review thus gives an overview of different automatic threshold detection methods of which a sufficient description of the method and performance results could be found. Hence it is possible that not all methods that have been tried thus far are included, but the methods that are most researched and used in clinical practise are included.

Conclusions

eCAP measurements, and especially eCAP thresholds, can aid in the fitting and monitoring of CIs. eCAP thresholds can be detected by a clinical specialist or automatically, but all methods are roughly based on two techniques: AGF and last visible. This review provides an overview of the different automatic threshold detection methods that have been described in literature: FG/AutoART, NRT-ANN, peak-picking AGF, AutoNRT, (FG-)SNR, SmartNRI and peak-picking last visible. All methods have their specific advantages and disadvantages. The outcome measures that are mostly used to evaluate these methods are success rates, correlations of the automatic methods with standard threshold detection methods and behavioural measurements and differences between standard and automatic methods. However, many of these outcome measures suffer from some sort of bias (selection or analytical) and there is currently no golden standard for eCAP threshold detection to compare these new automatic methods to. Besides, there are substantial differences between measurements that could influence the results (e.g., the timing of the measurements, eCAP thresholds versus behavioural measurements). Based on these limitations, no conclusions are drawn on the ranking of these automatic threshold detection methods (which methods perform better or worse). This review only provides an overview and demonstrated that further research should focus on other outcome measures: noninferiority of automatic threshold detection methods based on differences between standard and automatic methods, individual correlations and the precision of the different automatic threshold detection methods.

References

1. Vaerenberg B, Smits C, De Ceulaer G, Zir E, Harman S, Jaspers N, et al. Cochlear implant programming: a global survey on the state of the art. ScientificWorldJournal. 2014;2014:501738.
2. Mens LH. Advances in cochlear implant telemetry: evoked neural responses, electrical field imaging, and technical integrity. Trends Amplif. 2007;11(3):143-59.
3. Baudhuin JL, Hughes ML, Goehring JL. A Comparison of Alternating Polarity and Forward Masking Artifact-Reduction Methods to Resolve the Electrically Evoked Compound Action Potential. Ear Hear. 2016;37(4):e247-55.
4. He S, Teagle HFB, Buchman CA. The Electrically Evoked Compound Action Potential: From Laboratory to Clinic. Front Neurosci. 2017;11:339.
5. Herrmann DP, Kretzer KVA, Pieper SH, Bahmer A. Effects of electrical pulse polarity shape on intra cochlear neural responses in humans: Triphasic pulses with anodic and cathodic second phase. Hear Res. 2021;412:108375.
6. Hey M, Müller-Deile J. Accuracy of measurement in electrically evoked compound action potentials. J Neurosci Methods. 2015;239:214-22.
7. Hughes ML, Goehring JL, Baudhuin JL. Effects of Stimulus Polarity and Artifact Reduction Method on the Electrically Evoked Compound Action Potential. Ear Hear. 2017;38(3):332-43.
8. de Vos JJ, Biesheuvel JD, Briaire JJ, Boot PS, van Gendt MJ, Dekkers OM, et al. Use of Electrically Evoked Compound Action Potentials for Cochlear Implant Fitting: A Systematic Review. Ear Hear. 2018;39(3):401-11.

9. Biesheuvel JD, Briaire JJ, Frijns JHM. The Precision of eCAP Thresholds Derived From Amplitude Growth Functions. Ear Hear. 2018;39(4):701-11.
10. Botros A, van Dijk B, Killian M. AutoNR: an automated system that measures ECAP thresholds with the Nucleus Freedom cochlear implant via machine intelligence. Artif Intell Med. 2007;40(1):15-28.
11. Jeon EK, Brown CJ, Etler CP, O'Brien S, Chiou LK, Abbas PJ. Comparison of electrically evoked compound action potential thresholds and loudness estimates for the stimuli used to program the Advanced Bionics cochlear implant. J Am Acad Audiol. 2010;21(1):16-27.
12. Glassman EK, Hughes ML. Determining electrically evoked compound action potential thresholds: a comparison of computer versus human analysis methods. Ear Hear. 2013;34(1):96-109.
13. Strahl S, Dierker A, Spitzer P, Schwarz K. AutoART–A system for automatic determination of eCAP thresholds. 21 Jahrestagung Der Deutschen Gesellschaft Für Audiologie. 2018.
14. Garrido LC, Schwarz K, Lauss K, Vernetta CP, Kontides A, Gomez MD, et al. Comparison of a Traditional and Novel Evoked Compound Action Potentials Recording Approach and Evoked Auditory Brainstem Responses in Pediatric Cochlear Implants Users. J Int Adv Otol. 2018;14(3):353-8.
15. Estienne P, Scaglia A, Kontides A, Lauss K, Schwarz K, Arauz SL. Comparison of automated and traditional ECAP recording approaches in clinical practice. Int J Audiol. 2021:1-9.
16. Charasse B, Thai-Van H, Chanal JM, Berger-Vachon C, Collet L. Automatic analysis of auditory nerve electrically evoked compound action potential with an artificial neural network. Artif Intell Med. 2004;31(3):221-9.
17. Gärtner L, Lenarz T, Joseph G, Büchner A. Clinical use of a system for the automated recording and analysis of electrically evoked compound action potentials (ECAPs) in cochlear implant patients. Acta Otolaryngol. 2010;130(6):724-32.
18. van Dijk B, Botros AM, Battmer RD, Begall K, Dillier N, Hey M, et al. Clinical results of AutoNRT, a completely automatic ECAP recording system for cochlear implants. Ear Hear. 2007;28(4):558-70.
19. Gärtner L, Spitzer P, Lauss K, Takanen M, Lenarz T, Hoth S. Optimized SNR-based ECAP threshold determination is comparable to the judgement of human evaluators. PLoS One. 2021;16(11):e0259347.
20. Hoth S, Spitzer P, Praetorius M. A new approach for the determination of ECAP thresholds. Cochlear Implants Int. 2018;19(2):104-14.
21. Arnold L, Boyle P. SmartNRI: algorithm and mathematical basis. In: Proceedings of the 8th EFAS Conference, June 2007; Heidelberg, Germany.
22. Frachet B, Ormezzano Y, Arnold L, Boyle P. Smart NRI Multicenter Study in Europe: Preliminary Results. In: Advanced Bionics Auditory Research Bulletin Biennial Edition 2007.
23. Akin I, Götze R, Basta D, Böhnke B, Müller-Deile J, Ormezzano Y, et al. Results of a multicenter clinical study evaluating a new Smart algorithm to measure neural response imaging. Otol Neurotol. 2012;33(5):736-9.
24. Chauhan I, Swami H, Natraj R. A study on relationship between Neural Response Telemetry and behavioural Threshold/Comfort levels in children with cochlear implant. Med J Armed Forces India. 2021;77(2):224-9.
25. Scorpecci A, D'Elia A, Malerba P, Cantore I, Consolino P, Trabalzini F, et al. Maps created using a new objective procedure (C-NRT) correlate with behavioral, loudness-balanced maps: a study in adult cochlear implant users. Eur Arch Otorhinolaryngol. 2016;273(12):4167-73.
26. Wesarg T, Battmer RD, Garrido LC, Dillier N, Garcia-Ibáñez L, Hey M, et al. Effect of changing pulse rate on profile parameters of perceptual thresholds and loudness comfort levels and relation to ECAP thresholds in recipients of the Nucleus CI24RE device. Int J Audiol. 2010;49(10):775-87.
27. Akin I, Mutlu M, Kuran G, Dincer H, Arnold L, Boyle P. One-year results of the banded neural response imaging study. Otol Neurotol. 2008;29(5):635-8.
28. Stronks HC, Biesheuvel JD, de Vos JJ, Boot MS, Briaire JJ, Frijns JHM. Test/Retest Variability of the eCAP Threshold in Advanced Bionics Cochlear Implant Users. Ear Hear. 2019;40(6):1457-66.

Appendix

A. Search strings

Search string PubMed

(("Cochlear Implants"[Mesh] OR "cochlear implant"[tw] OR "cochlear implants"[tw] OR "cochlea implant"[tw] OR "cochlea implants"[tw] OR "cochlear prosthesis"[tw] OR "auditory prosthesis"[tw] OR "auditory implant"[tw] OR "auditory implants"[tw] OR "Cochlear Implantation"[Mesh] OR "Cochlear Implantation"[tw] OR "cochlear implant*"[tw] OR "Cochlea Implantation"[tw] OR "Cochlea Implant*"[tw]) AND ("eCAP"[tw] OR "eCAPs"[tw] OR "electrically evoked compound action potential"[tw] OR "electrically evoked compound action potentials"[tw] OR "electrically evoked action potential"[tw] OR "electrically evoked action potentials"[tw] OR "electrically evoked"[tw] OR "EAP"[tw] OR "EAPs"[tw] OR "auditory nerve response telemetry"[tw] OR "auditory response telemetry"[tw] OR "auditory nerve response"[tw] OR "auditory response"[tw] OR ("ART"[tw] AND "auditory nerve"[tw]) OR "neural response imaging"[tw] OR "NRI"[tw] OR "neural response telemetry"[tw] OR "NRT"[tw]) AND ("auto*" [All Fields] OR "software*" [All Fields] OR "algorithm*" [All Fields] OR "computer*" [All Fields])) AND (english[la] OR dutch[la]) AND ("2002/01/01"[PDAT] : "3000/12/31"[PDAT])

Search string Embase

((exp *"Cochlea Prosthesis"/ OR "cochlear implant".ti OR "cochlear implants".ti OR "cochlea implant".ti OR "cochlea implants".ti OR "cochlear prosthesis".ti OR "auditory prosthesis".ti OR "auditory implant".ti OR "auditory implants".ti) AND ("eCAP".ti OR "eCAPs".ti OR "electrically evoked compound action potential".ti OR "electrically evoked compound action potentials".ti OR "electrically evoked action potential".ti OR "electrically evoked action potentials".ti OR "electrically evoked".ti OR "EAP".ti OR "EAPs".ti OR "auditory nerve response telemetry".ti OR "auditory response telemetry".ti OR "auditory nerve response".ti OR "auditory response".ti OR ("ART".ti AND "auditory nerve".ti) OR "neural response imaging".ti OR "NRI".ti OR "neural response telemetry".ti OR "NRT".ti) AND ("auto*".af OR "algorithm*".af OR "computer*".af OR "software*".af)) AND (english.la OR dutch.la) AND 2002:2023.(sa_year) NOT (conference review or conference abstract).pt

Search string Web of Science

((TI=("Cochlear Implants" OR "cochlear implant" OR "cochlear implants" OR "cochlea implant" OR "cochlea implants" OR "cochlear prosthesis" OR "auditory prosthesis" OR "auditory implant" OR "auditory implants" OR "Cochlear Implantation" OR "Cochlear Implantation" OR "cochlear implant*" OR "Cochlea Implantation" OR "Cochlea Implant*") OR AK=("Cochlear Implants" OR "cochlear implant" OR "cochlear implants" OR "cochlea implant" OR "cochlea implants" OR "cochlear prosthesis" OR "auditory prosthesis" OR "auditory implant" OR "auditory implants" OR "Cochlear Implantation" OR "Cochlear Implantation" OR "cochlear implant*" OR "Cochlea Implantation" OR "Cochlea Implant*") OR AB=("Cochlear Implants" OR "cochlear implant" OR "cochlear implants" OR "cochlea implant" OR "cochlea implants" OR "cochlear prosthesis" OR "auditory prosthesis" OR "auditory implant" OR "auditory implants" OR "Cochlear Implantation" OR "Cochlear Implantation" OR "cochlear implant*" OR "Cochlea Implantation" OR "Cochlea Implant*")) AND (TI=("eCAP" OR "eCAPs" OR "electrically evoked compound action potential" OR "electrically evoked compound action potentials" OR "electrically evoked action potential" OR "electrically evoked action potentials" OR "electrically evoked" OR "EAP" OR "EAPs" OR "auditory nerve response telemetry" OR "auditory response telemetry" OR "auditory nerve response" OR "auditory response" OR ("ART" AND "auditory nerve") OR "neural response imaging" OR "NRI" OR "neural response telemetry" OR "NRT") OR AK=("eCAP" OR "eCAPs" OR "electrically evoked compound action potential" OR "electrically evoked compound action potentials" OR "electrically evoked action potential" OR "electrically evoked action

potentials" OR "electrically evoked" OR "EAP" OR "EAPs" OR "auditory nerve response telemetry" OR "auditory response telemetry" OR "auditory nerve response" OR "auditory response" OR ("ART" AND "auditory nerve") OR "neural response imaging" OR "NRI" OR "neural response telemetry" OR "NRT") OR AB=("eCAP" OR "eCAPs" OR "electrically evoked compound action potential" OR "electrically evoked compound action potentials" OR "electrically evoked action potential" OR "electrically evoked action potentials" OR "electrically evoked" OR "EAP" OR "EAPs" OR "auditory nerve response telemetry" OR "auditory response telemetry" OR "auditory nerve response" OR "auditory response" OR ("ART" AND "auditory nerve") OR "neural response imaging" OR "NRI" OR "neural response telemetry" OR "NRT")) AND (ALL=("auto*" OR "algorithm*" OR "computer*" OR "software*"))) AND la=(english OR dutch) AND py=(2002 OR 2003 OR 2004 OR 2005 OR 2006 OR 2007 OR 2008 OR 2009 OR 2010 OR 2011 OR 2012 OR 2013 OR 2014 OR 2015 OR 2016 OR 2017 OR 2018 OR 2019 OR 2020 OR 2021 OR 2022)

B. Study details

Table B.1. Measurement details of the included articles.
CI = cochlear implant, AB = Advanced Bionics, eCAP = electrically evoked compound action potential, NRI = Neural Response Imaging, AGF = amplitude growth function, NRT = Neural Response Telemetry, ANN = artificial neural network, FG = FineGrain, SNR = Signal-to-noise ratio

ARTICLE	TYPE CI	TIMING OF MEASUREMENTS	ELECTRODES	ARTEFACT REDUCTION	METHODS
AKIN ET AL., 2012 (23)	CII Bionic Ear or HiRes 90K system (AB)	Intra-operative, at first fitting and after 3 months of CI use	4 electrode pairs typically 3/1, 7/5, 11/9, and 15/13	Alternating polarity	SmartNRI and standard AGF
CHARASSE ET AL., 2004 (16)	Nucleus (Cochlear)	Not mentioned	Electrodes 5, 10, 15 and 20	Forward masking	NRT-ANN and standard AGF
CHAUHAN ET AL., 2021 (23)	Cochlear	Intra-operative	All electrodes		AutoNRT and behavioural measurements (T- and C-levels)
ESTIENNE ET AL., 2021 (15)	MED-EL	Post-operative	Electrodes 2, 6 and 10	Alternating polarity and signature subtraction	FG, standard AGF and behavioural measurements (C-levels)
GARRIDO ET AL., 2018 (14)	MED-EL	Intra-operative and post-operative	Not mentioned	Alternating polarity and signature subtraction	FG and standard AGF
GÄRTNER ET AL., 2010 (17)	Cochlear	Post-operative	All 22 electrodes	Forward masking	AutoNRT and last visible
GÄRTNER ET AL., 2021 (19)	CI with an i100 platform (MED-EL)	Post-operative	All 12 electrodes	Alternating polarity and signature subtraction	FG-SNR and standard AGF
GLASSMAN ET AL., 2013 (12)	- 10 Cochlear (9 Nucleus® 24RE[CA] and 1 Nucleus® CI512)	At least 3 months of CI use	- Cochlear: electrodes 3, 11, and 20 - AB: electrodes 1, 8, and 14	- Cochlear: forward masking	Peak-picking AGF, peak-picking last visible, standard AGF and last visible

	- 10 AB (5 CII and 5 HiRes 90K)			- AB: alternating polarity	
HOTH ET AL., 2018 (20)	MED-EL	eCAP: intra-operative Subjective measurements: 4 weeks post-operative	All 12 electrodes	Alternating polarity and signature extraction	SNR, standard AGF and behavioural measurements (T- and C-level profiles)
JEON ET AL., 2010 (11)	CII or HiRes 90K (AB)	At least 1 year of CI use	Electrodes 3, 7, 11 and 15	Alternating polarity	Peak-picking AGF and last visible
SCORPECCI ET AL., 2016 (25)	Nucleus Freedom or CI512 CI (Cochlear)	At least 6 moths of CI use	Electrodes 1, 3, 6, 9, 11, 13, 16, 19, 22		AutoNRT, behavioural measurements (T- and C-levels)
VAN DIJK ET AL., 2007 (18)	Freedom system (Cochlear)	Intra-operative, regular intervals post-operative	Not mentioned	Forward masking	AutoNRT and last visible
WESARG ET AL., 2010 (26)	CI24RE (Cochlear)	Intra-operative and up to 22 weeks post-operative	Electrodes 3, 5, 10, 15, 20 or electrodes 1, 6, 11, 16, and 22	Forward masking	AutoNRT and standard AGF

B. 8-points loudness scale

8. HOOGST ACCEPTABELE LUIDHEID

7. LUID

6. AANGENAAM MAAR LUID

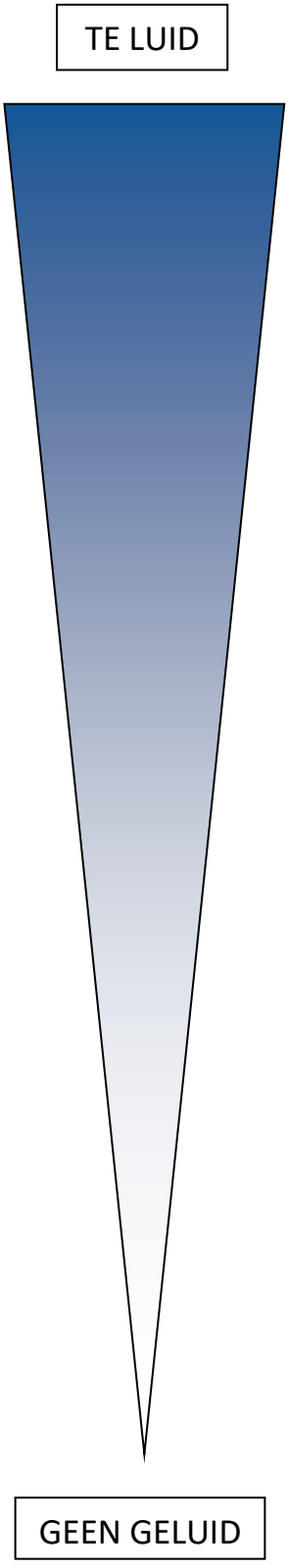
5. AANGENAAMSTE LUIDHEID

4. AANGENAAM MAAR ZACHT

3. ZACHT

2. ZEER ZACHT

1. DREMPELWAARDE



C. Measuring protocol

A - Protocol Soundwave testen

Processor = Naida Q70

Stimulatie = 3x ‘tone burst’.

0 – Dagelijkse ‘map’

1. Maak een print van de dagelijkse map waarop voor alle elektrode contacten de T- en M-levels zichtbaar zijn.

I - Meet de impedanties

1. Bepaal de impedanties voor alle oneven elektrode contacten. Meet het naastgelegen basale contact als een contact uitgeschakeld is in de dagelijkse ‘map’ (variabele = **Inactive electrodes**), of als de impedantie hoger is dan 10 kOhm.
2. Bewaar de impedanties in Castor (variabele = **Impedance electrode #**).

II - Fitting data opslaan

Haal de volgende gegevens op uit SoundWave en sla deze op in Castor.

1. Fitting strategie (variabele = **Fitting strategy**).
2. Pulse width (variabele = **Pulse width**).

III - T-levels bepalen

Meet de T-levels op alle elektrode contacten waarvoor je ook de impedanties hebt gemeten (bij A-I).

1. Neem map-parameters over van huidige fitting (strategy, pulse width, etc.).
2. Laat de stimulus tweemaal duidelijk horen op een luidheid tussen T-level en M-level van de dagelijkse ‘map’ (zie A-0), zodat de proefpersoon het geluid duidelijk kan horen en weet wat hij/zij moet horen. Dit is het initiële stimulusniveau, bewaar dit in Castor (variabele = **Initial stimulus level electrode #**).
3. Zak dan terug naar 5 CU onder het T-level van de dagelijkse ‘map’ (zie A-0) en laat de stimulus horen.
4. Indien het gehoord wordt, zak dan 5 CU en laat het opnieuw horen.
5. Indien het niet gehoord wordt, stijg dan 5 CU en laat het opnieuw horen.
6. Het T-level is het laagste niveau dat nog gehoord wordt.
7. Indien je twee keer hetzelfde (zo laag mogelijke) T-level bepaalt, is dit het uiteindelijke T-Level. Bewaar dit in Castor (variabele = **T-Level electrode #**).

B - Protocol eCAP testen

Formule[1] voor omrekenen CU naar uA: $Amplitude = \frac{CU}{pulse\ width \times 0.013}$ met PW = 32 us

I – Levels (T/M) bepalen

Meet de T-levels en MAL op alle elektrode contacten waarvoor je ook de T-levels hebt gemeten (bij A).

1. Vul bij de 'Impedances (kOhm)' uit Castor (variabele = **Impedance electrode #**) in.
2. Vul bij 'MAL (uA)' het **Initial stimulus level in uA electrode #** uit Castor in, en laat dit horen via de knop 'Play MAL'. Indien dit gehoord wordt, ga dan naar stap 3. Indien dit niet gehoord wordt, ga dan naar 4.
3. Bepaal het T-level zoals beschreven bij A-III 4 t/m 7 met behulp van de 'Up' en 'Down' knoppen in 'Set MAL' (stimulus wordt automatisch afgespeeld wanneer je klikt op 'Up' of 'Down').
    a. Laat eerst nog een keer het **Initial stimulus level** horen.
    b. Gebruik een stapgrootte van 40 uA (vul dit in bij 'Stepsize') en laat het stimulus level zakken (via de knop 'Down') tot het niet meer gehoord wordt.
    c. Wijzig de stapgrootte naar 10 uA (wat ongeveer overeenkomt met 5 CU).
    d. Ga verder met het bepalen van het T-level zoals beschreven bij A-III 4 t/m 7.
    e. Ga verder naar stap 5.
4. Stappenplan voor wanneer **Initial stimulus level in uA electrode #** niet gehoord wordt.
    a. Vul een stapgrootte van 40 uA in bij 'Stepsize'. Verhoog het stimulus level door te klikken op 'Up'. Blijf het stimulus level ophogen tot de patiënt het hoort.
    b. Laat de stimulus nog een keer horen op deze luidheid (die de patiënt hoort) via de knop 'Play MAL'.
    c. Zak terug 40 uA via de knop 'Down'.
    d. Verander de 'Stepsize' naar 10 uA en bepaal het T-level zoals beschreven bij A-III 4 t/m 7.
    e. Ga verder naar stap 5.
5. Vul de waarde die je bepaald hebt als T-level in bij 'T-Level eCAP (uA)'.
6. Ga nu verder met het bepalen van de MAL met behulp van de 8-punts luidheidsschaal. Vul weer bij 'MAL (uA)' het **Initial stimulus level in uA electrode #** uit Castor in indien dit gehoord werd of laat het zonet bepaalde T-Level staan en laat dit horen via 'Play MAL'. Laat de proefpersoon aanduiden met welke score van de 8-punts luidheidsschaal dit overeenkomt. Wanneer de proefpersoon een score van 4 of lager aangeeft, verhoog dan het stimulus level met een

[1] https://www.advancedbionics.com/content/dam/advancedbionics/Documents/Regional/US/libraries/Audiologists-&-Surgeons-Library/Device-Fitting-Manuals/soundwave-fitting-manual.pdf

'Stepsize' van 50 uA en druk op 'Up'. Laat de proefpersoon dit opnieuw scoren en blijf het stimulus level ophogen tot de patiënt een score van 5 aangeeft. Verander dan de 'Stepsize' naar 20 uA en ga verder met het ophogen van het stimulus level tot de patiënt een score van 7 aangeeft. Verander opnieuw de 'Stepsize' naar 10 uA, laat de proefpersoon dit opnieuw scoren en blijf het stimulus level ophogen tot de patiënt een score van 8 aangeeft. Het stimulus level waarbij de patiënt een score van 8 aangeeft komt overeen met de MAL (dit hoeft niet meer apart worden ingevoerd, bij het klikken op 'Up' wordt de 'MAL (uA)' automatisch aangepast).

7. Klik tot slot nog een keer in het vakje van 'T-Level eCAP (uA)' en druk op enter zodat alles juist wordt opgeslagen.
8. Doorloop deze stappen voor alle benodigde elektrodes.

<u>II - eCAP metingen</u>

Meet de eCAPs op alle elektrode contacten waarvoor je ook de T-levels hebt gemeten (bij A).

1. Selecteer 1 elektrode door te klikken op 'E#'.
2. Begin met de standaard biphasic meting door te klikken op 'Measure' onder 'Biphasic FM and AP'. De parameters voor standaard biphasic zijn als volgt:
    a. 32 averages
    b. Imasker = Iprobe, start met 0 uA, in 10 gelijke stappen tot MAL (gemeten met ADEPT-software)
    c. IPI = 400
    d. PW = 32
    e. Sampling rate = 56 kHz
    f. Start recording before stimulus
    g. Gain = 300
    h. 5 frames: masker, masker+probe, probe, signature, alternating polarity
3. Ga nadien door met het meten van de FineGrain door te klikken op 'Measure' onder 'Biphasic FM and AP FineGrain'. De parameters voor FineGrain zijn als volgt:
    a. Zelfde als standaard biphasic, maar met 2 averages, stimulus level start met 0 uA, in stappen van 7.5 uA tot MAL (gemeten met ADEPT-software). Daarna wordt er een moving average genomen van 7 opeenvolgende stimuluslevels.
4. Zet de elektrode weer uit door terug te klikken op 'E#'.
5. Selecteer de volgende elektrode en voer de metingen weer uit.

Druk op 'Exit' nadat je alle benodigde elektrodes hebt gemeten.

D. Results of multiple comparison tests for threshold confidence intervals

Table D.1: Population marginal means and their corresponding standard deviation for the normalised threshold confidence intervals (in dB). SA = standard averaging, AP = alternating polarity, FM = forward masking, FG = FineGrain averaging, AGF = amplitude growth function, SNR = signal-to-noise ratio.

Method	Mean	Standard deviation
SA AP	-5.70	0.94
SA FM	-8.81	0.75
FG AP	-14.95	0.76
FG FM	-18.98	0.66
AGF sigmoid	-19.59	0.84
AGF linear	-7.17	1.07
SNR	-8.07	0.77
Cross-covariance between adjacent levels	-9.72	1.01
Cross-covariance with maximum level	-16.00	0.94
SA AP with AGF sigmoid	-15.81	2.25
SA FM with AGF sigmoid	-13.86	2.18
FG AP with AGF sigmoid	-25.29	2.21
FG FM with AGF sigmoid	-23.41	2.11
SA AP with AGF linear	-4.54	3.37
SA FM with AGF linear	-0.91	3.26
FG AP with AGF linear	-12.17	2.27
FG FM with AGF linear	-11.04	2.06
SA AP with SNR	1.23	2.04
SA FM with SNR	-4.53	1.94
FG AP with SNR	-10.18	1.96
FG FM with SNR	-18.80	1.96
SA AP with cross-covariance between adjacent levels	-0.49	3.78
SA FM with cross-covariance between adjacent levels	-15.07	2.02
FG AP with cross-covariance between adjacent levels	-5.76	2.50
FG FM with cross-covariance between adjacent levels	-17.56	1.97
SA AP with cross-covariance with maximum level	-8.90	2.83
SA FM with cross-covariance with maximum level	-9.66	2.09
FG AP with cross-covariance with maximum level	-21.35	2.77
FG FM with cross-covariance with maximum level	-24.07	2.06

E. Results of multiple comparison tests for correlation coefficients

Table E.2: Population marginal means and their corresponding standard deviation for the Fisher transformed correlation coefficients. SA = standard averaging, AP = alternating polarity, FM = forward masking, FG = FineGrain averaging, AGF = amplitude growth function, SNR = signal-to-noise ratio.

Method	Mean	Standard deviation
T-levels in SW	0.71	0.08
T-levels with eCAP stimuli	0.12	0.08
SA AP	0.53	0.22
SA FM	0.86	0.20
FG AP	0.49	0.20
FG FM	0.87	0.18
AGF sigmoid	1.09	0.22
AGF linear	0.84	0.26
SNR	0.66	0.21
Cross-covariance between adjacent levels	0.31	0.25
Cross-covariance with maximum level	0.54	0.24
SA AP with AGF sigmoid	0.79	0.60
SA FM with AGF sigmoid	1.21	0.60
FG AP with AGF sigmoid	1.49	0.58
FG FM with AGF sigmoid	0.88	0.55
SA AP with AGF linear	1.04	0.76
SA FM with AGF linear	1.07	0.83
FG AP with AGF linear	0.93	0.58
FG FM with AGF linear	0.31	0.55
SA AP with SNR	0.17	0.53
SA FM with SNR	0.98	0.53
FG AP with SNR	0.33	0.55
FG FM with SNR	1.17	0.54
SA AP with cross-covariance between adjacent levels	0.03	0.83
SA FM with cross-covariance between adjacent levels	0.61	0.55
FG AP with cross-covariance between adjacent levels	-0.25	0.63
FG FM with cross-covariance between adjacent levels	0.82	0.54
SA AP with cross-covariance with maximum level	0.64	0.71
SA FM with cross-covariance with maximum level	0.41	0.54
FG AP with cross-covariance with maximum level	-0.07	0.71
FG FM with cross-covariance with maximum level	1.18	0.54