



AutoNRT™: An automated system that measures ECAP thresholds with the Nucleus® Freedom™ cochlear implant via machine intelligence

Andrew Botros^{a,*}, Bas van Dijk^b, Matthijs Killian^b

^a Cochlear Ltd., 14 Mars Road, Lane Cove, NSW 2066, Australia

^b Cochlear Technology Centre Europe, Schaliënhoevedreef 20 I, 2800 Mechelen, Belgium

Received 24 January 2006; received in revised form 11 May 2006; accepted 30 June 2006

KEYWORDS

Cochlear implants;
Electrically evoked
compound action
potential;
Neural response
telemetry;
Threshold estimation;
Automated systems;
Machine learning;
Pattern recognition;
Decision trees

Summary

Objective: AutoNRT™ is an automated system that measures electrically evoked compound action potential (ECAP) thresholds from the auditory nerve with the Nucleus® Freedom™ cochlear implant. ECAP thresholds along the electrode array are useful in objectively fitting cochlear implant systems for individual use. This paper provides the first detailed description of the AutoNRT algorithm and its expert systems, and reports the clinical success of AutoNRT to date.

Methods: AutoNRT determines thresholds by visual detection, using two decision tree expert systems that automatically recognise ECAPs. The expert systems are guided by a dataset of 5393 neural response measurements. The algorithm approaches threshold from lower stimulus levels, ensuring recipient safety during postoperative measurements. Intraoperative measurements use the same algorithm but proceed faster by beginning at stimulus levels much closer to threshold. When searching for ECAPs, AutoNRT uses a highly specific expert system (specificity of 99% during training, 96% during testing; sensitivity of 91% during training, 89% during testing). Once ECAPs are established, AutoNRT uses an unbiased expert system to determine an accurate threshold. Throughout the execution of the algorithm, recording parameters (such as implant amplifier gain) are automatically optimised when needed.

Results: In a study that included 29 intraoperative and 29 postoperative subjects (a total of 418 electrodes), AutoNRT determined a threshold in 93% of cases where a human expert also determined a threshold. When compared to the median threshold of multiple human observers on 77 randomly selected electrodes, AutoNRT performed as accurately as the 'average' clinician.

* Corresponding author. Tel.: +61 2 9428 6555; fax: +61 2 9428 6353.
E-mail address: abotros@cochlear.com.au (A. Botros).

Conclusions: AutoNRT has demonstrated a high success rate and a level of performance that is comparable with human experts. It has been used in many clinics worldwide throughout the clinical trial and commercial launch of Nucleus Custom Sound™ Suite, significantly streamlining the clinical procedures associated with cochlear implant use.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Cochlear implants and Neural Response Telemetry (NRT™)

The cochlear implant is a device that electrically stimulates the auditory nerve, bypassing the non-functional inner ear of children and adults with moderate-to-profound hearing loss. Current cochlear implant systems consist of (i) a multichannel electrode array that is surgically implanted and (ii) an external sound processing unit (usually worn behind the ear) that controls the implant over a transcutaneous RF link. The system is configured and analysed via device-specific PC software. (For an in-depth coverage of cochlear implants, see Clark [1].)

The Nucleus® cochlear implant has the ability to measure electrically evoked compound action potentials (ECAPs) from the auditory nerve. The system applies an electrical pulse on a given intracochlear electrode and the evoked neural response is recorded at a neighbouring electrode. The measured potentials are telemetered back to the system's programming interface for clinical analysis.

This feature – ‘Neural Response Telemetry’ – was first available for commercial use in the Nucleus CI24M implant [2,3]. The technique is essentially that of Brown et al. [4]. In 2005 the Nucleus Freedom™ implant was released, offering NRT with additional functionality (such as the third phase

artefact reduction pulse) and a much-improved signal to noise ratio [5,6].

A sequence of NRT measurements that displays clear ECAPs is shown in Fig. 1 (left panel). Each measurement displays a clear negative and positive peak (N1 and P1, respectively). N1 occurs within a fraction of a millisecond. ECAP clarity varies widely: measurements may display a partial N1 peak, no P1 peak or a double positive peak (Lai and Dillier [7] provide an overview of ECAP morphologies). A sequence of NRT measurements that displays the absence of a neural response is shown in Fig. 1 (middle and right panels). Stimulus artefact and/or noise is observed—the stimulus may be too weak or the stimulus artefact may obscure the ECAP. Distinguishing between measurements that display ECAPs and those that do not is an important task when performing NRT. This can be difficult when the combination of stimulus artefact and noise gives the impression of an obscure ECAP.

NRT provides a number of clinical benefits. Intraoperatively, NRT can be used to verify implant and auditory nerve integrity during surgery; postoperatively, NRT can be used to monitor recipient progress and, perhaps most importantly, to objectively fit the sound processing system. ECAP features of interest include the threshold level, peak-to-peak amplitude growth functions, neural recovery functions, and measurements of the spatial spread of excitation (Brown [8] and Cafarelli Dees et al. [9] provide recent

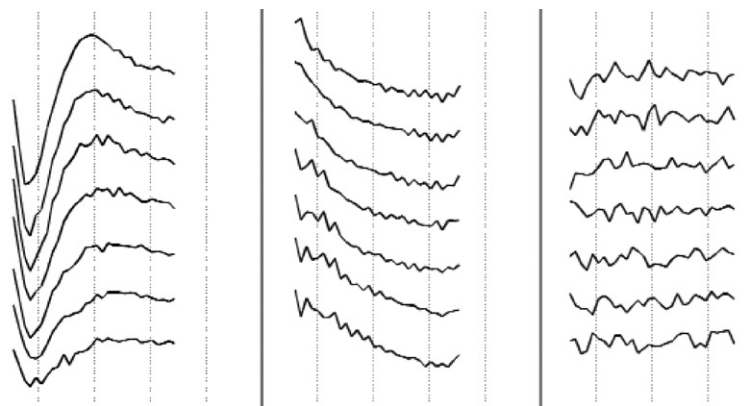


Figure 1 NRT measurements (horizontal axis: time; vertical axis: voltage). Left: Measurements displaying clear ECAPs. Middle: Measurements dominated by stimulus artefact, with no ECAPs evident. Right: Measurements containing noise only.

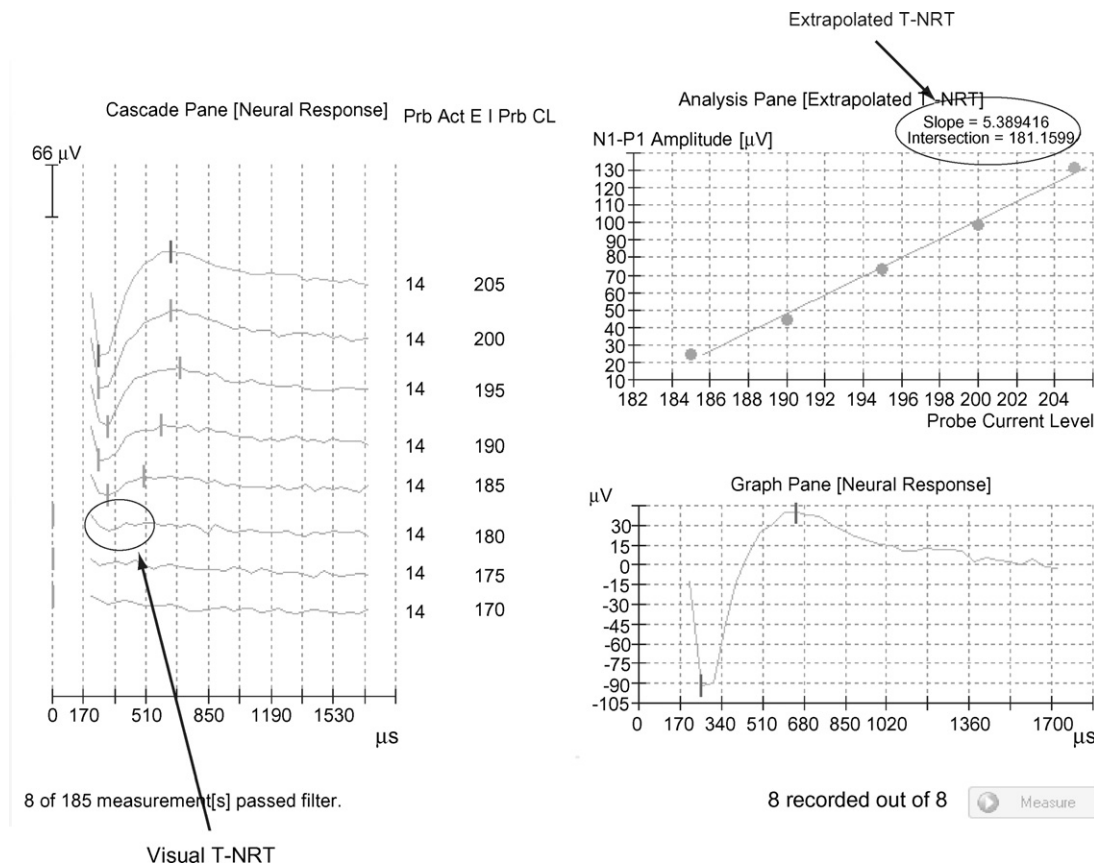


Figure 2 ECAP threshold measurements using the Nucleus Custom Sound EP software. A sequence of NRT measurements is performed on electrode 14 with a stimulus range of 170–205CL. To determine visual threshold, a clinician searches for the first instance of an ECAP (180CL). To determine extrapolated threshold, the AGF is extrapolated to the current level of zero N1–P1 amplitude (181CL).

overviews). The first of these – the threshold current level¹ at which an ECAP is obtained (T-NRT) – is the clinical parameter of most interest.

To fit a cochlear implant for a given recipient's requirements, a clinician must subjectively determine the individual's hearing dynamic range on each electrode (softest and loudest current levels). This task is difficult and time consuming, particularly with young children, and thus objective fitting methods can assist clinicians. Several researchers have presented methods for predicting these psychophysical levels from T-NRT levels (e.g. [10–14]). Measuring T-NRT levels can be difficult also: recording parameters (such as amplifier gain) may need to be optimised for a given recipient, and an appreciable level of expertise is required to interpret NRT recordings effectively.

AutoNRT™, a new feature of the Nucleus Freedom cochlear implant system, measures T-NRT levels automatically. It is available in Nucleus Custom Sound™ Suite, comprising Custom Sound and Custom Sound EP. AutoNRT is available in both software applications; in addition to AutoNRT, Custom Sound EP offers a wide range of advanced NRT functionality.

tom Sound™ Suite, comprising Custom Sound and Custom Sound EP. AutoNRT is available in both software applications; in addition to AutoNRT, Custom Sound EP offers a wide range of advanced NRT functionality.

1.2. T-NRT measurement methods

T-NRT levels are typically measured in one of two ways: by visual detection or by extrapolation of the amplitude growth function (AGF). These two methods are illustrated in Fig. 2 (see also [8]).

Visual threshold is determined by manually observing the minimum current level at which ECAP peaks are visible and can be replicated. A variation on the visual threshold method is the correlation threshold technique: a clear, suprathreshold ECAP is used as a template, and threshold is defined at a lower level where the correlation coefficient degrades sufficiently when the given NRT measurement is matched with the template. The extrapolated threshold method is based on the assumption that the ECAP peak-to-peak amplitude grows linearly with increasing current level above threshold.

¹ The current level (CL) scale is logarithmic. For the Nucleus Freedom implant, $I (\mu\text{A}) = 17.5 \times 100^{\text{CL}/255}$. Each current level step (1CL) is a 0.16 dB change in current.

Threshold is defined as the zero-amplitude intercept of the AGF slope.

1.3. Automated T-NRT measurements

Systems that automatically measure T-NRT levels (or determine them offline with a given set of NRT measurements) have been built in the past [15,16] and continue to be built [17]. In all cases, the chosen method has been extrapolated threshold. An expert system analyses NRT measurements at a range of current levels; those that are deemed to represent ECAPs are used to construct an AGF, from which a T-NRT level is extrapolated. The expert systems have taken various forms: Charasse et al. [15] used an artificial neural network (ANN) where the output neurons corresponded to one of five ECAP morphologies (both N1 and P1 visible, N1 missing, no neural response, etc.); Charasse et al. [18] also compared the ANN to a cross-correlation (CC) technique, where a given NRT measurement is compared with an array of fixed neural responses, grouped according to the five ECAP morphologies; and Nicolai et al. [19] presented an expert system that combined the ANN and CC techniques with additional rule-based criteria.

The AGF linearity assumption is not valid at all current levels however. Typically, the AGF is linear at higher current levels and tails off near threshold (the AGF also flattens at very high current levels, giving an overall sigmoidal function, but these levels are not often reached). Fig. 3 illustrates this characteristic shape. The nonlinearity near threshold poses a difficulty for automated systems that are based on the extrapolated threshold method. If the linear portion of the AGF is desired, a clinician must first determine the maximum current level that the recipient can withstand. This provides the system with an upper bound on the AGF current levels it can

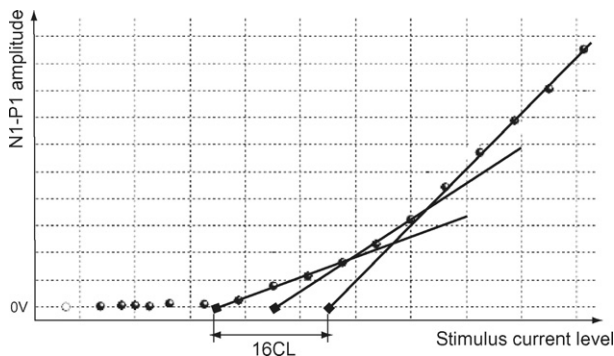


Figure 3 Characteristic AGF shape. Near threshold, the AGF is nonlinear: a number of regression lines are possible, leading to inaccurate extrapolated T-NRTs. Circles: individual NRT measurements. Diamonds: extrapolated T-NRTs.

examine. Without such a bound, the system must evaluate the AGF at lower current levels to ensure safety. As Fig. 3 shows, extrapolated threshold is poorly defined at these levels. Indeed, previous systems have required maximum current level measurements from clinicians, or required clinicians to perform the NRT measurements prior to analysis; thus, these systems are not strictly automated.

AutoNRT differs from previous systems. AutoNRT measures T-NRT levels by visual detection, approaching threshold from low current levels and halting as soon as an ECAP is obtained. With this approach, AutoNRT provides a completely automated method for measuring ECAP thresholds in both intraoperative and postoperative settings. This paper describes the AutoNRT algorithm and its pattern recognition component. A discussion of the design and clinical results to date is also provided.

2. The AutoNRT algorithm

2.1. Summary flow

The AutoNRT algorithm consists of two logical phases: an 'ascending series' and a 'descending series'. The ascending series performs NRT measurements at increasing current levels until an ECAP is detected by the expert system. Thereafter, the descending series performs NRT measurements at decreasing current levels with finer step sizes to establish threshold more accurately.

To ensure safety postoperatively, AutoNRT begins at a low current level (default 100CL²). Intraoperatively, when the recipient is under general anaesthesia, AutoNRT begins at a level that is closest to the expected T-NRT: this is either the population mean (170CL) or the interpolated value from neighbouring electrodes that have already been measured. The ascending series increases the current level in 6CL² steps. The descending series decreases the current level in 3CL steps. Postoperatively, if the rising current level is perceived by the recipient to be too loud, the clinician simply cancels the measurement on the current electrode and AutoNRT continues on the remaining selected electrodes.

Two separate expert systems are used. The ascending series uses an expert system (ES1) that has a low false positive rate: the goal of the ascending series is to establish the presence of ECAPs with high confidence. To reduce the error rate further, two consecutive ECAP positive predictions are required before the ascending series is complete. The descending series uses an expert system (ES2)

² This value can be adjusted by the clinician.

that has a low error rate overall: the goal of the descending series is to establish an accurate threshold once ECAPs are obtained.

If the implant amplifier saturates at any stage during the measurement, AutoNRT attempts to optimise a number of NRT recording parameters. If this is unsuccessful, the measurement is cancelled and AutoNRT continues on the remaining electrodes. Similarly, if voltage compliance cannot be achieved at high levels of stimulation (i.e. the implant cannot deliver the required current), or if the maximum current level is reached (255CL), the measurement is cancelled.

The descending series completes when two consecutive ECAP negative predictions are given by ES2. Threshold is (roughly) defined as the mean current level of ES2's lowest ECAP positive measurement and highest ECAP negative measurement. Fig. 4 gives a more precise specification of the AutoNRT algorithm flow.

2.2. NRT recording parameter optimisation

AutoNRT uses default NRT recording parameters, with the exception of: (i) a stimulation rate of 250 Hz is used intraoperatively, to minimise the time taken during surgery (default is 80 Hz) and (ii) 35 averages are used per measurement (default is 50). For default NRT measurements: (i) the implant amplifier gain is set to 50 dB; (ii) a measurement delay of 120 μ s is used (the latency between stimulation and recording); (iii) the forward masking paradigm is used to reduce artefact [4]; and (iv) the third phase artefact reduction pulse³ is not used. Each NRT measurement contains 32 samples, sampled at 20 kHz.

ECAPs are much smaller than (artefactual) stimulus potentials; in some measurements, the stimulus artefact saturates the implant amplifier. When this occurs, AutoNRT attempts to use a third phase artefact reduction pulse and/or reduce the amplifier gain, as such:

1. Use the third phase artefact reduction pulse, automatically optimising its current level such that stimulus artefact is minimised.
2. If the amplifier still saturates, (i) reduce the gain to 40 dB; (ii) increase the number of averages by a factor of 1.5 (to maintain the signal to noise

ratio with the lower gain setting); and (iii) turn off the third phase artefact reduction pulse.

3. If amplifier still saturates, use the third phase artefact reduction pulse with the 40 dB gain setting. If this is also unsuccessful, cancel the AutoNRT measurement.

2.3. Supporting measurements

Nucleus Custom Sound Suite enforces impedance measurements prior to performing AutoNRT. This is particularly important during surgery where the extracochlear electrodes can become dry, effectively open circuiting the implant system. If high impedances are found, the clinician is advised to check electrode placement.

Additionally, intraoperative AutoNRT is preceded by an electrode conditioning phase. High current stimulation is applied to the selected electrode until its impedance stabilises. The interface between electrode and fluid changes over time: impedances decrease as the electrode surfaces settle into contact with the underlying perilymph. Electrical stimulation facilitates this process. The decrease in impedance leads to less stimulus artefact, improving AutoNRT's efficacy.

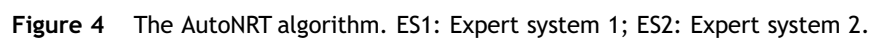
3. The AutoNRT expert system

3.1. Specification

The ascending series and descending series expert systems are shown in Fig. 5. They take the form of decision trees. The decision node parameters are the following features of a given NRT measurement:

- **N1P1:** N1–P1 amplitude (μ V) = $ECAP_{P1} - ECAP_{N1}$. Peaks are selected according to the following rules (see Fig. 6): N1 is the minimum of the first 8 samples; P1 is the maximum of the samples after N1, up to and including sample 16; if any one of the following conditions is true however, N1–P1 = 0 μ V:
 - N1–P1 < 0 μ V;
 - latency between N1 and P1 < 2 samples;
 - latency between N1 and P1 > 12 samples; or
 - latency between N1 and the maximum sample after N1 > 15 samples and ratio of N1–P1 to the range from N1 onwards < 0.85 (explained in the next section).
- **Noise:** The noise level (μ V) is defined as the range (maximum – minimum) of samples 21–32 after subtracting the least-squares regression line through these 12 samples.

³ Implant stimulation consists of a train of alternate polarity biphasic pulses (25 μ s pulse width per phase; 7 μ s inter-phase gap); the Nucleus Freedom implant allows a small-amplitude, 10 μ s pulse width third phase per pulse to reduce the stimulus artefact of the second phase.



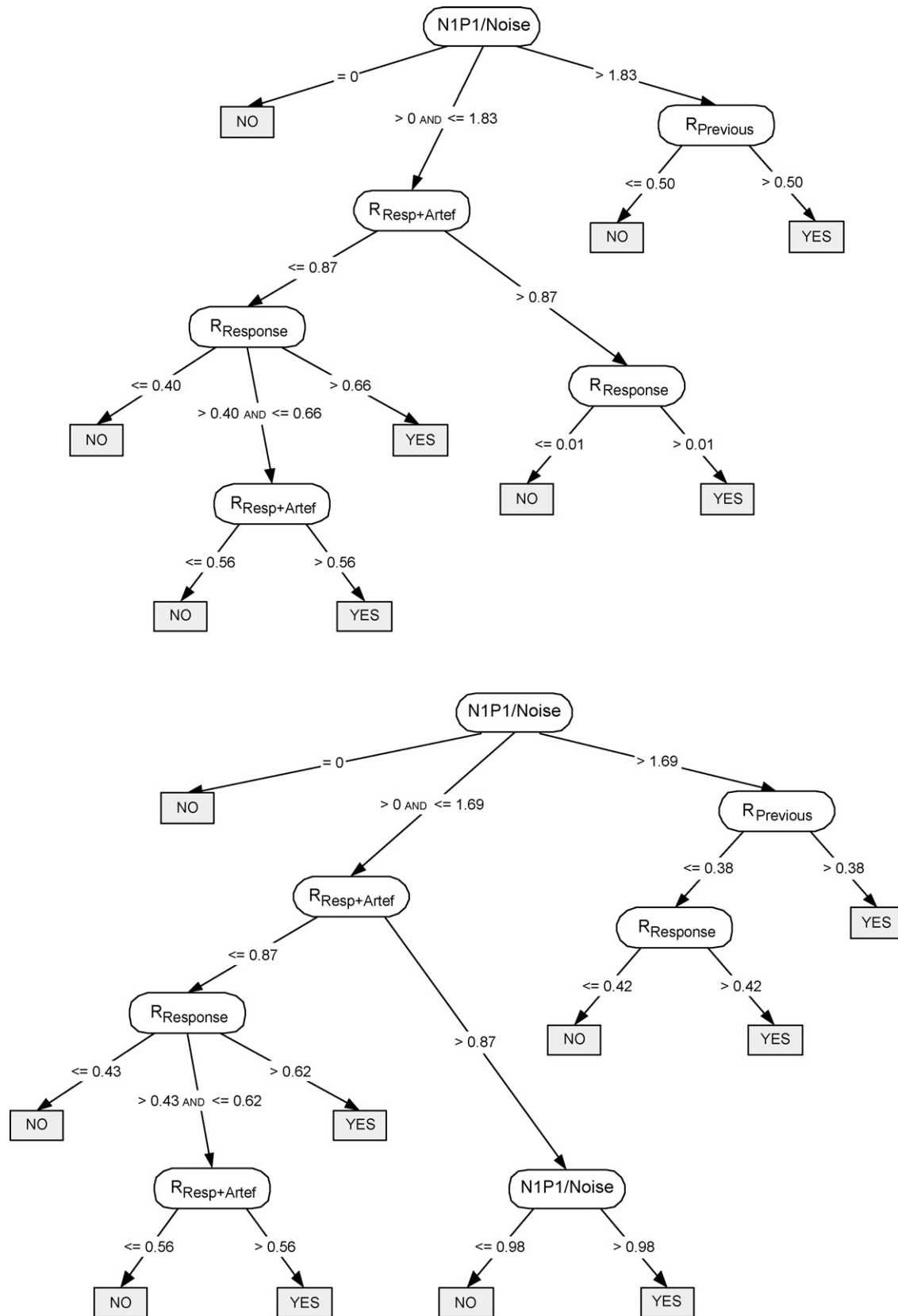


Figure 5 The AutoNRT expert systems. Each decision tree determines whether a given NRT measurement represents an ECAP or not. Top: Expert system 1 (high specificity for ascending series). Bottom: Expert system 2 (specificity and sensitivity equal for descending series).

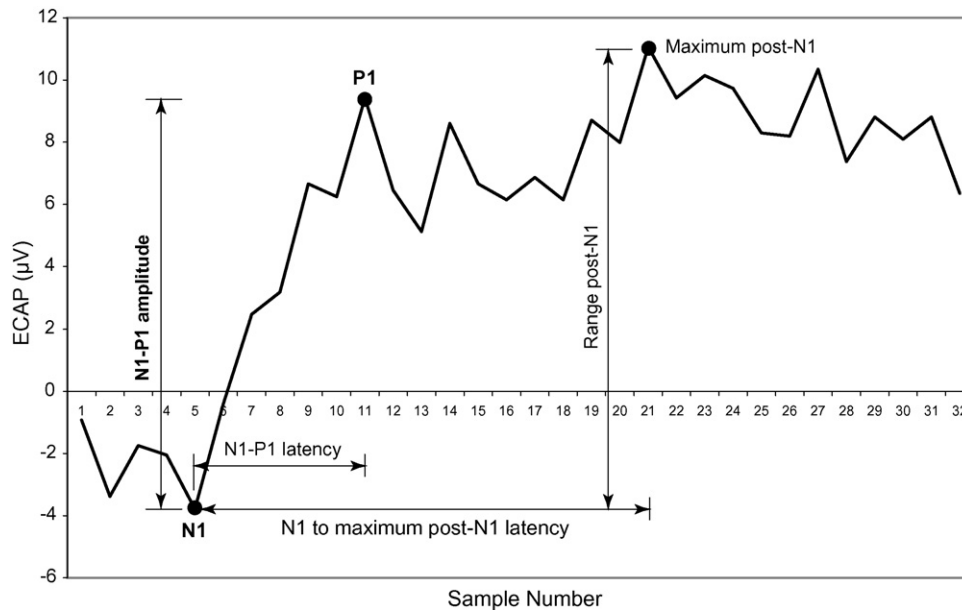


Figure 6 Peak picker feature extraction. The NRT measurement, which contains only stimulus artefact and noise, looks remarkably like a valid ECAP. The peak picker does not reject this measurement, but the ascending series expert system makes a correct classification ('NO') by virtue of its R_{previous} decision node.

- $N1P1/\text{Noise}$: The ratio of the N1–P1 amplitude to the noise level. Since the ECAP morphology is of more interest than the absolute ECAP amplitude, a normalised measure of signal amplitude is preferred.
- R_{Response} : The correlation between the given NRT measurement and a fixed clear neural response (Fig. 7, left), calculated over samples 1–24. (The template is the average of all ECAPs in the experimental dataset.)
- $R_{\text{Response+Artefact}}$: The correlation between the given NRT measurement and a fixed measurement containing both neural response and stimulus artefact (Fig. 7, middle), calculated over samples 1–24. (The template is the average of 200 manually selected ECAPs in the experimental dataset that are contaminated with stimulus artefact.)
- R_{previous} : The correlation between the given NRT measurement and the NRT measurement of immediately lower stimulus current level during AutoNRT's execution (regardless of step size), calculated over samples 1–24.

3.2. Construction methods

The AutoNRT expert systems are two-tiered—they each consist of a peak picker and a decision tree classifier, combined in the one tree structure (the peak picker is common to both the ascending and descending series expert systems). Both components are machine-learned using the C5.0 decision tree algorithm [20,21]; decision trees are the most popular choice in data mining applications today, providing quick and informative data analysis with potentially large sets of features.

Learning was guided by a large dataset of 5393 NRT measurements. Most of the measurements were performed postoperatively with a group of 18 recipients, using random intracochlear electrodes. 268 intraoperative NRT measurements that are dominated by stimulus artefact are also included in the dataset. Each measurement was classified as 'YES' (ECAP positive, 60% of the dataset) or 'NO' (ECAP negative, 34% of the dataset) by two experts. No distinction is made between different ECAP



Figure 7 NRT measurement templates. The AutoNRT expert systems correlate a given NRT measurement with these templates to assist classification. Left: Clear ECAP. Middle: ECAP plus stimulus artefact. Right: Stimulus artefact only.

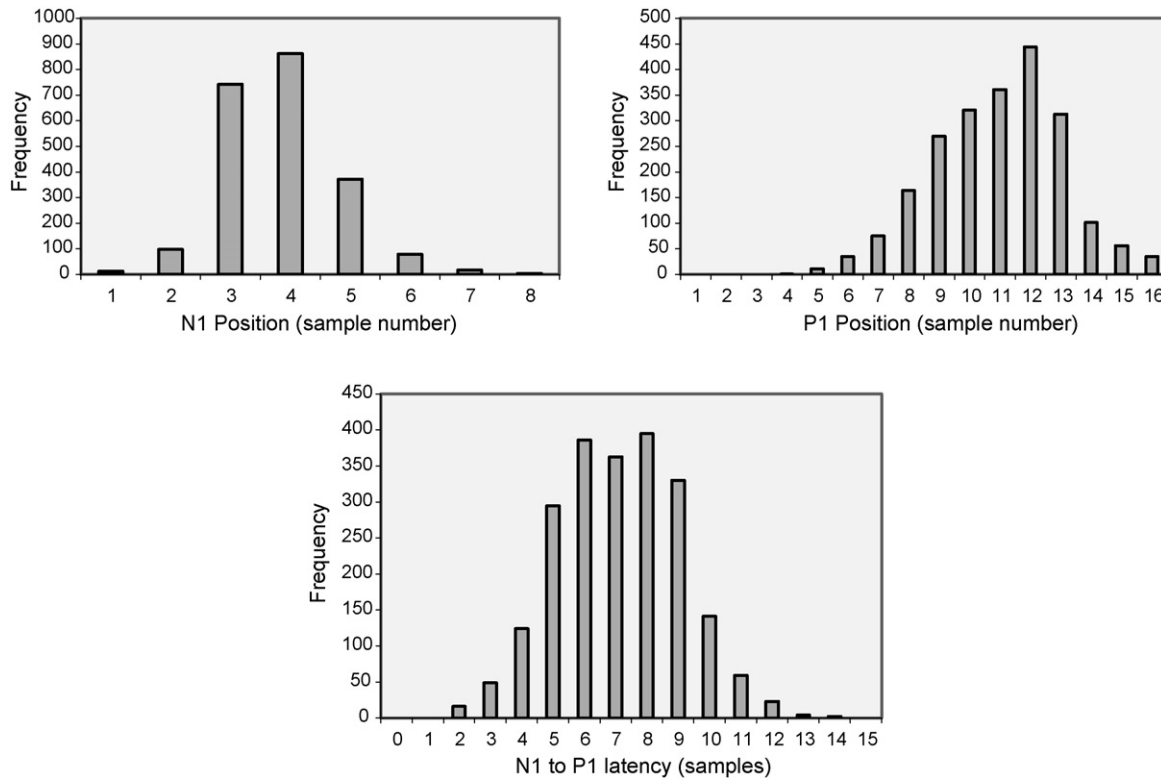


Figure 8 Top: The distribution of N1 and P1 position amongst 2187 training instances. Bottom: The distribution of N1–P1 latency. For AutoNRT measurements, sample 1 is taken 120 μ s after the stimulus completes, and each sample is separated by 50 μ s.

morphologies. Measurements with different classifications by the two experts were discarded (6%); of the remaining measurements, 3638 were used for training (63% ‘YES’; 37% ‘NO’) and 1443 were used for testing (65% ‘YES’; 35% ‘NO’).

3.2.1. Peak picker construction

The task of the peak picker is to identify potential N1 and P1 peaks and discard NRT measurements with false peaks. The peak picker pre-processes data for the classification stage: whereas the peak picker selects the measurement samples that are potentially the peaks of an ECAP, it is the decision tree classifier that determines whether the *entire trace* represents a valid ECAP or not.

Peak picking is a non-trivial task: N1 and P1 peaks are not always prominent, and traces that are dominated by stimulus artefact can display peak-like characteristics. Furthermore, a P1 peak may not always be present—the peak picker must select a suitable maximum in its place. Thus, to correctly select peaks in such a domain, a simple search for global extrema is insufficient.

2187 ECAP positive measurements were selected from the training dataset for peak analysis. Fig. 8 (top) shows the distribution of N1 and P1 position for these measurements. We base the N1 and P1

windows on these results: N1 is the minimum of the first eight samples; P1 is the maximum of the samples after N1, up to and including sample 16.

To determine whether the selected peaks are due to stimulus artefact, appropriate rules were machine-learned from the dataset. 24 troublesome artefact measurements were added to the 2187 ECAP positive measurements. These 24 measurements, such as the one in Fig. 6, display a characteristic upward slope that strongly suggests the shape of an ECAP. Only 24 such measurements exist in the experimental dataset (they are relatively rare). Eight features that we considered to be potentially useful in distinguishing artefact traces were identified, such as: the latency between N1 and P1; the latency between N1 and the global maximum after N1; the latency between P1 and the global maximum after N1; the ratio of N1–P1 amplitude to the global range from N1 onwards (intuitively, N1–P1 amplitude should be a significant proportion of the global range); etc. From these features, C5.0 learned the following rules:

- if N1–P1 latency > 12 samples, reject peaks;
- if the latency between N1 and the global maximum after N1 > 23 samples and the ratio of

N1–P1 amplitude to the global range from N1 onwards < 0.69 , reject peaks;

- otherwise, accept peaks.

Of the 2187 ECAP positive measurements, 7 were rejected based on these rules; of the 24 artefact traces, 2 were falsely accepted, giving an overall 0.4% error rate. To increase the specificity of the peak picker, we chose to strengthen the second rule manually. This raised the peak picker error rate to 1.5% over the training data. Admittedly, the 24 artefact traces form a small-sized training set; however, we note that the peak picker is only the first stage of the expert system and that the performance impact is reasonably small.

A final guard is the rejection of peaks that are too close to each other. The distribution of N1–P1 latency amongst ECAP positive measurements is shown in Fig. 8 (bottom). No peaks occur at consecutive samples, so a simple added rule is: if N1–P1 latency < 2 samples, reject peaks. This rule pairs with the upper bound of 12 samples set by C5.0.

3.2.2. Decision tree classifier construction

NRT measurements that are rejected by the peak picker were discarded from the dataset, since these measurements are classified as 'NO' before the decision tree stage. A training set of 3020 measurements and a test set of 1223 measurements remain.

Six features were extracted from each measurement: the four features given in the decision tree nodes and, additionally: (i) the correlation between a given NRT measurement and a fixed measurement containing stimulus artefact only (Fig. 7, right) and

(ii) the gradient of the least-squares regression line through the noise portion of the measurement (samples 21–32). (The latter two features are not used in the expert systems—C5.0 deemed them insignificant.)

To construct the ascending series expert system, we set the cost of a false ECAP positive prediction to be five times worse than the converse error (higher weightings raised the overall error rate without significantly improving specificity). This allows the ascending series to give ECAP positive predictions with a higher level of confidence. To construct the descending series expert system, all errors received the same weighting, allowing C5.0 to generate an unbiased classifier.

An important consideration in machine learning is ensuring that training data are not overfitted. If an algorithm attempts to fit training instances as closely as possible, the performance of the resulting system with unseen data is likely to be reduced. A decision tree is quite capable of fitting training data perfectly since there is no limit to the degree of branching that may occur. To avoid such overfitting, C5.0 provides a mechanism for pruning decision trees: the data analyst may specify a minimum number of training instances that must follow at least two of the branches at each node. Insignificant branches are replaced by leaf node classifications.

We evaluated the cross-validation error, test set error and, for the ascending series expert system, the specificity at different levels of C5.0 pruning, selecting the decision tree that performed well over all three measures. Cross-validation randomly divides the training instances into a number of

Table 1 Ascending series decision tree performance at different levels of pruning

C5.0 PRUNING LEVEL	CROSS-VALIDATION ERROR (%)	TEST SET ERROR (%)	CROSS-VALIDATION SPECIFICITY (%)	TEST SET SPECIFICITY (%)
2	5.4	6.4	95.9	90.3
5	6.1	6.7	96.0	90.6
10	5.8	7.0	96.3	90.3
15	6.0	6.3	96.5	91.5
20	6.1	6.6	96.3	92.7
30	6.7	6.6	96.1	91.8
40	6.7	6.6	96.5	93.3
50	7.0	6.3	96.4	93.3
60	7.0	7.5	96.4	93.6
70	6.3	7.5	96.0	93.6
80	7.1	7.3	95.9	94.5
90	7.8	7.3	95.3	93.3
100	9.3	7.4	96.0	90.3
150	12.0	9.9	97.5	95.1
200	12.5	12.7	98.3	98.2

Selected tree is highlighted. Pruning level is the minimum number of instances that at least two branches must carry at a decision node.

Table 2 Descending series decision tree performance at different levels of pruning

C5.0 PRUNING LEVEL	CROSS-VALIDATION ERROR (%)	TEST SET ERROR (%)
2	4.7	6.7
5	5.0	7.0
10	5.2	6.7
15	5.1	6.8
20	5.4	6.7
30	6.0	6.1
40	6.4	7.4
50	6.2	7.9
60	6.3	8.0
70	6.5	8.0
80	6.4	8.0
90	6.8	8.0
100	6.6	8.0
150	7.9	9.2
200	9.2	10.5

Selected tree is highlighted

Table 3 Training set confusion matrix for the ascending series expert system

	Predicted YES (%)	Predicted NO (%)
YES	2105 (91.4)	199 (8.6)
NO	13 (1.0)	1321 (99.0)

Table 4 Test set confusion matrix for the ascending series expert system

	Predicted YES (%)	Predicted NO (%)
YES	834 (88.6)	107 (11.4)
NO	22 (4.4)	480 (95.6)

Table 5 Training set confusion matrix for the descending series expert system

	Predicted YES (%)	Predicted NO (%)
YES	2177 (94.5)	127 (5.5)
NO	44 (3.3)	1290 (96.7)

Table 6 Test set confusion matrix for the descending series expert system

	Predicted YES (%)	Predicted NO (%)
YES	857 (91.1)	84 (8.9)
NO	39 (7.8)	463 (92.2)

blocks with approximately equal class distribution. For each block in turn, a decision tree is constructed from data in the remaining blocks and tested on the instances in the hold-out block. In this way, each

Table 7 Test set performance comparison of AutoNRT's ascending series expert system (ES1) with artificial neural network (ANN) and cross-correlation (CC) techniques

	Specificity (%)	Sensitivity (%)
AutoNRT (ES1)	96	89
ANN (Charasse et al. [15,18])	95	68
CC (Charasse et al. [18])	95	78
ANN + CC + rules (Nicolai et al. [19])	93	80

The AutoNRT expert system is based on measurements from an implant with improved signal to noise ratio.

instance can be used exactly once as a test case. Tables 1 and 2 show the results of this evaluation for the ascending and descending series expert systems. The selected trees are highlighted.

3.2.3. Expert system evaluation

Tables 3–6 show the training set and test set confusion matrices for the ascending and descending series expert systems (including the peak picker stage).

The descriptive quality of decision trees allows an easy insight into the expert system. At a glance, the structure of the expert system is intuitive: N1P1/Noise is placed at the top of the decision trees, as expected, and the remaining branches form plausible rules.

Table 7 compares the test set specificity and sensitivity of the ascending series expert system (the more critical of the two) with those of previous researchers. It is important to note, however, that the results are not directly comparable, since (i) previous systems have been based on NRT measurements with Nucleus CI24M/R implants, which are noisier and (ii) previous systems only consider measurements with clear N1 and P1 peaks to be ECAP positive—AutoNRT places no such restriction on the ECAP definition.

4. Results

AutoNRT has been used extensively throughout the clinical trial and commercial launch of the Nucleus Freedom cochlear implant system. A sizeable body of clinical data exists; van Dijk et al. provide the results of the first large study [22], and these are summarised briefly here. It is important to note, however, that the results of van Dijk et al. span both the validation and commercial iterations of AutoNRT (this paper describes the current commercial release).

van Dijk et al. performed AutoNRT with 29 intraoperative and 29 postoperative subjects, a total of 418 electrodes. On 21 electrodes, no ECAP threshold could be determined by either AutoNRT or a human observer. Of the remaining 397 electrodes, thresholds were determined by both AutoNRT and an expert clinician in 370 cases (93%). Of the 27 discrepancies, half were due to algorithm error and half were due to AutoNRT giving no threshold due to low confidence (an element of earlier designs).

For the 370 electrodes where both AutoNRT and the expert clinician determined an ECAP threshold, the absolute difference between the two was less than 9CL in 90% of cases, with a median of 3CL and a maximum of 37CL. However, when AutoNRT was compared to multiple human observers, AutoNRT performed just as well as the ‘average’ clinician. Five human observers (four experts and one novice) determined T-NRT levels on 77 randomly selected electrodes. The observers did not perform any recording parameter optimisation (this was already performed by AutoNRT), and the AutoNRT T-NRT levels were hidden from them. For each electrode, the median T-NRT of the four expert observers was nominally set as the ‘true’ T-NRT level. Each observer – AutoNRT included – was compared with this median. Fig. 9 shows the result of the comparison, demonstrating the ability of AutoNRT to perform just as well as an experienced clinician. Interestingly, the novice clinician also performs just as well as two of the experts (discussed below). Further, two of the experts differed by as much as 30CL: returning to the single-human comparison, AutoNRT’s maximum error of 37CL should be

considered with this inter-observer variability in mind.

Intraoperatively, AutoNRT had a mean execution time of 23 s per electrode (S.D. 5 s). All intraoperative measurements were performed with a fixed starting current level; thus, when multiple electrodes are measured in a session, the mean execution time is less than 23 s because the starting current level is based on T-NRTs from neighbouring electrodes. Postoperatively, where AutoNRT must begin at a low current level and uses a lower stimulation rate, the mean execution time was 46 s (S.D. 11 s). A manual procedure typically takes a few minutes.

Thus, AutoNRT is successful and accurate in the vast majority of cases. Compared to a manual procedure, AutoNRT saves time and gives objective results that are more consistent across clinics worldwide. Furthermore, as with previous releases, we endeavour to improve the accuracy of AutoNRT in future releases of Nucleus Custom Sound Suite as more training data becomes available.

5. Discussion

True automation requires a high level of performance in a number of aspects: (i) the automated system must function at the single press of a button; (ii) the system must produce results in almost all cases; and (iii) the system must be sufficiently accurate. Although these requirements are difficult to satisfy simultaneously, AutoNRT provides a successful balance.

To achieve ECAP thresholds at the press of a button, AutoNRT takes an infrathreshold approach so that, postoperatively, safety is assured from the start of the measurement. If the stimulation becomes too loud, a clinician must intervene and cancel the measurement; in the absence of this event however, AutoNRT operates at a single button press. Whilst this approach enhances automation, it places a heavy burden on the expert system. This is for two reasons: (i) the expert system must detect ECAPs near threshold, where the signal is less clearly defined and (ii) the expert system does not have the benefit of seeing large ECAPs at high stimulation levels—ECAPs that can be used as a correlation template at lower stimulation levels. A comparison with measurements of the auditory brainstem response (ABR) highlights the latter factor further.

Visual detection of threshold is common with the ABR. This is widely performed to detect neonatal deafness. Typically, a high level acoustic stimulation is used to establish a template response, and this is correlated with responses at lower volumes to find threshold. This is easy to do acoustically because the

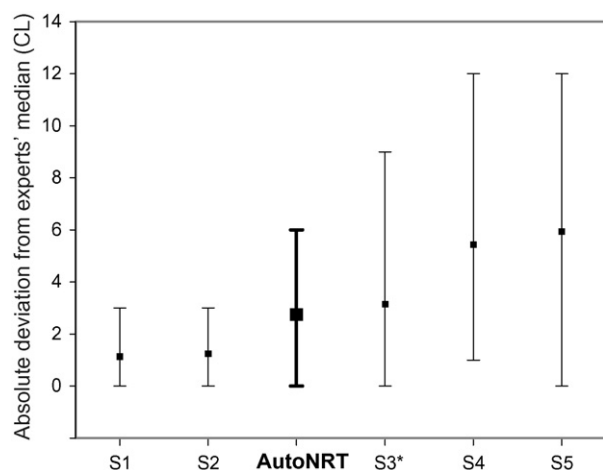


Figure 9 Performance of AutoNRT compared to five human observers (S1–S5). The median T-NRT of the four experts is defined as the ‘true’ T-NRT for 77 T-NRT measurements. Data points are the mean absolute deviations from this median; error bars are the 10th and 90th percentiles. Novice observer denoted by asterisk (*).

dynamic range of acoustic hearing is extremely large and sound level scales are perceived consistently across the population (for example, 70 dB SPL speech is similarly loud to different listeners). Thus, it is simple to define a starting level that is both safe and likely to evoke a large response. Accordingly, automated systems exist that detect ABR thresholds by visual detection (e.g. [23]). In contrast, ECAP thresholds can be close to the maximum acceptable level, and stimulation levels differ largely across recipients and even across electrodes.

To achieve T-NRT levels with a high success rate, AutoNRT is sufficiently sensitive with all possible ECAP morphologies. Whereas the systems of Charasse et al. [15] and van Dijk et al. [16] only use responses with clear N1 and P1 peaks (a prudent precaution with Nucleus CI24M/R waveforms), AutoNRT makes no distinction in ECAP morphology. This provides AutoNRT with a greater chance of success on any given electrode. Similarly, the AutoNRT expert system is trained with NRT measurements containing many obscure morphologies near threshold. By comparison, Litvak and Emadi [17] reject 40% of their dataset, only including traces that are classified unanimously by five clinicians. Charasse et al. [15] and van Dijk et al. [16], with small subject pools, do not provide a firm indication of their systems' success rates; a reduced rate is suggested by the sensitivities of their expert systems (68% [18] and 80% [19], respectively, with clear peaks required) and the requirements of obtaining an AGF (Charasse et al. [15] require five valid ECAPs). Thus, previous systems are designed to be highly specific, and this reduces the success rate and hence the level of automation.

The pursuit of sensitivity, however, directly reduces accuracy. Notwithstanding this trade-off, AutoNRT has demonstrated a level of accuracy that is comparable with a human expert. The use of two separate expert systems for the ascending and descending phases provides the required balance between sensitivity and accuracy: the ascending series expert system is highly specific (specificity of 99% during training and 96% during testing), and the descending series expert system treats all misclassifications equally. Since the visual detection method requires ECAP recognition at low signal levels, where noise and artefact are significant, the expert system features are designed to be morphology-sensitive rather than amplitude-sensitive: evoked potentials are normalised to the noise level (N1P1/Noise), downward sloping artefact is tracked by template matching, and upward sloping artefact is tracked by the peak picker rules.

When compared to the median of multiple human experts, AutoNRT's absolute mean deviation was

2.8CL (Fig. 9). This is similar to the results of Charasse et al. [15] (3.6CL) and van Dijk et al. [16] (2.3CL). Furthermore, a novice clinician performed just as well as an experienced clinician in the AutoNRT observer pool. Thus, we conclude that threshold determination is ideally suited for automation: discrepancies between multiple observers are most likely due to differences in the subjective definitions of 'threshold', rather than any inherent difficulty of the task.

Despite the level of automation that AutoNRT achieves, clinical experts may prefer to supervise AutoNRT measurements if they feel that the results can be improved from time to time. Nucleus Custom Sound Suite displays the NRT measurements as they occur, and clinicians can adjust the T-NRT level as they wish. Nevertheless, with or without human supervision, AutoNRT saves significant clinical time through its automated measurement sequence, recording parameter optimisation and machine analysis. AutoNRT is a powerful tool for all clinicians, both expert and novice.

6. Conclusions

AutoNRT offers a completely automated means of obtaining ECAP thresholds with the Nucleus Freedom cochlear implant. Whereas previous systems require considerable manual effort and expertise to provide NRT data or ensure safety prior to the automated procedure, AutoNRT performs all functions at the press of a button. AutoNRT has demonstrated a high success rate (93% of electrodes) and a level of performance that is comparable with human experts. It has been successfully used in many clinics worldwide, significantly streamlining the clinical procedures associated with cochlear implant use.

Acknowledgements

We thank Pascal Winnen of Cochlear Technology Centre Europe for technical assistance. We thank the clinics that gathered data during the development and validation of AutoNRT—in particular: the Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation (Melbourne and Sydney); University Hospital Zurich; Medizinische Hochschule Hannover; Universitätsklinikum Freiburg; Universitätsklinikum Kiel; AMEOS Klinikum St Salvator Halberstadt; St Augustinus Hospital Wilrijk. We also thank all implant recipients who participated in the Nucleus Freedom clinical trials.

References

- [1] Clark G. Cochlear implants: fundamentals and applications. New York: Springer-Verlag; 2003.
- [2] Abbas PJ, Brown CJ, Shallop JK, Firszt JB, Hughes ML, Hong SH, Staller SJ. Summary of results using the Nucleus CI24M implant to record the electrically evoked compound action potential. *Ear Hear* 1999;20:45–59.
- [3] Dillier N, Lai WK, Almqvist B, Frohne C, Müller-Deile J, Stecker M, von Wallenberg E. Measurement of the electrically evoked compound action potential (ECAP) via a neural response telemetry (NRT) system. *Ann Otol Rhinol Laryngol* 2002;111:407–14.
- [4] Brown CJ, Abbas PJ, Gantz B. Electrically evoked whole-nerve action potentials: data from human cochlear implant users. *J Acoust Soc Am* 1990;88:1385–91.
- [5] Daly CN, Nygard TM, Eder H. Method and apparatus for measurement of evoked neural response. US Patent Application Publication No. 20050101878.
- [6] Eder HC, Hurley PJ, Money DK, Nygard TM. Method and apparatus for measurement of evoked neural response. International (PCT) Patent Application Publication No. WO/2004/021885.
- [7] Lai WK, Dillier N. A simple two-component model of the electrically evoked compound action potential in the human cochlea. *Audiol Neurotol* 2000;5:333–45.
- [8] Brown CJ. The electrically evoked whole nerve action potential. In: Cullington HE, editor. *Cochlear implants: objective measures*. London: Whurr Publishers; 2003. p. 96–129.
- [9] Cafarelli Dees D, Dillier N, Lai WK, von Wallenberg E, van Dijk B, et al. Normative findings of electrically evoked compound action potential measurements using the neural response telemetry of the Nucleus CI24M cochlear implant system. *Audiol Neurotol* 2005;10:105–16.
- [10] Brown CJ, Hughes ML, Luk B, Abbas PJ, Wolaver A, Gervais J. The relationship between EAP and EABR thresholds and levels used to program the Nucleus 24 speech processor: data from adults. *Ear Hear* 2000;21:151–63.
- [11] Hughes ML, Brown CJ, Abbas PJ, Wolaver AA, Gervais JP. Comparison of EAP thresholds to MAP levels in the Nucleus CI24M cochlear implant: data from children. *Ear Hear* 2000; 21:164–74.
- [12] Franck KH. A model of a Nucleus 24 cochlear implant fitting protocol based on the electrically evoked whole nerve action potential. *Ear Hear* 2002;23:675–715.
- [13] 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 Neurotol* 2002;7:335–47.
- [14] Thai-Van H, Truy E, Charasse B, Boutitie F, Chanal J-M, Cochard N, et al. Modeling the relationship between psychophysical perception and electrically evoked compound action potential threshold in young cochlear implant recipients: clinical implications for implant fitting. *Clin Neurophysiol* 2004;115:2811–24.
- [15] 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;31:221–9.
- [16] van Dijk B, Krey C, Verhulst L, Marichal C, Charasse B, Collet L. Development of a prototype fully-automated intra-operative ECAP recording tool, using NRT™ V3. In: Shepherd RK, Svirsky MA, editors. *Abstracts of the 2003 Conference on Implantable Auditory Prostheses*. 2003. p. 178.
- [17] Litvak L, Emadi G. Automatic estimate of threshold from neural response imaging (NRI). In: Zeng F-G, Snyder R, editors. *Abstracts of the 2005 conference on implantable auditory prostheses*. 2005. p. 211.
- [18] Charasse B, Killian M, Berger-Vachon C, Collet L. Comparison of two different methods to automatically classify auditory nerve responses recorded with NRT system. *Acta Acust United Acust* 2004;90:512–9.
- [19] Nicolai J, Charasse B, Collet L, van Dijk B. Performance of automatic recognition algorithms in Nucleus neural response telemetry. In: Shepherd RK, Svirsky MA, editors. *Abstracts of the 2003 Conference on Implantable Auditory Prostheses*. 2003. p. 179.
- [20] Quinlan JR. C4.5: programs for machine learning. San Mateo: Morgan Kaufmann; 1993.
- [21] Quinlan JR. C5.0: an informal tutorial. Rulequest Research; <http://www.rulequest.com/see5-unix.html> (accessed 1 May 2006).
- [22] van Dijk B, Ambrosch P, Battmer R-D, Begall K, Botros A, Dillier N, Hey M, Lenarz T, Müller-Deile J, Weber B, Wesarg T, Zarowsky A, Offeciers E. AutoNRT™: first clinical results of a completely automatic ECAP recording system. In: Zeng F-G, Snyder R, editors. *Abstracts of the 2005 conference on implantable auditory prostheses*. 2005. p. 229.
- [23] Vannier E, Adam O, Motsch J-F. Objective detection of brainstem auditory evoked potentials with a priori information from higher presentation levels. *Artif Intell Med* 2002; 25:283–301.