Data from Past Patients used to Streamline Adjustment of Levels for Cochlear Implant for New Patients

Jim Torresen and Andreas Høyer Iversen
Department of Informatics
University of Oslo
Oslo, Norway
e-mail: jimtoer@ifi.uio.no

Ralf Greisiger
Department of Otorhinolaryngology
Oslo University Hospital
Oslo, Norway
e-mail: rgreisig@ous-hf.no

***Abstract*—Cochlear implant technology gives deaf people the ability to sense sound and speech. It consists of an electrode array inserted (implanted) in the cochlea of the ear and an external device that wirelessly connects to the electrodes. Each electrode stimulates different areas of the auditory nerve based of what frequency they represent. The sensitivity of auditory nerve fibers varies from patient to patient so upper and lower limits for each electrode must be set for each patient. Currently, this is undertaken by an operator adjusting the levels while the patient gives oral feedback. This process is both time consuming and challenging in that many patients are young children who are only partly able to provide feedback. At Oslo University Hospital, data has been collected on adjustment levels and response measurements from a number of former patients (158 have been used in this project). In this paper, we consider to what extent it is possible to predict values for *new* patients by using various machine learning techniques on data from *previous* patients. Although it is not possible to achieve fully automatic adjustments, the experiments show that a good starting point can be provided for manual adjustment. Further, the work has also shown which electrodes are most important to measure to automatically predict levels of other electrodes.**

Keywords—medical application; learning and adaptive control; machine learning; cochlear implant

I. Introduction

According to the WHO, there are more than 360 million people worldwide with loss of hearing [1]. Depending on the cause of deafness, many of these can benefit from hearing aids that amplify sound waves. There are, however, some cases that are too severe for normal hearing aids to be of help. In these cases it is often possible to bypass the damaged parts of the inner ear in order to stimulate the auditory nerve fibres directly. The auditory nerve fibres are responsible for sending signals to the brain and by stimulating them directly with a *cochlear implant* (CI) device, a hearing sensation can be achieved. The CI consists of an electrode array inserted (implanted) in the cochlea of the ear and an external device with a microphone that wirelessly connects to the electrodes, see Fig. 1. The electrode array provides direct electrical stimulation of the auditory nerve fibers for a number of frequencies corresponding to the number of electrodes (typically 20-22, where the lowest frequencies correspond to nerve fibres furthest inside the cochlear). The hearing sensation does not appear immediately after the CI has been implanted, and the patient needs some time to learn and adapt to the new way of perceiving sounds. Depending on various factors, such as the duration of deafness and reason for deafness, the performance of a CI can vary greatly [2, 3], but many people are able to hear well enough to follow a normal conversation and in the best cases, the hearing and language skills are close to those with normal hearing, at least in quiet listening situations.

Fig. 1. A Cochlear Implant. (Image provided courtesy of Cochlear)

Today, no method exists for using objective measures, i.e., that can be recorded from medical patient responses during and/or after a CI surgery, to program the CI. As a result, CI programming is mainly based on oral feedback from the patient, combined with the experience of the audiologist. This makes it especially difficult for audiologists who are new or inexperienced with programming a CI. They may need additional time to find an appropriate configuration, which is challenging because patients have varying durations of concentration when the CI programming takes place [4]. The longer the programming takes, the more likely it also is that the patient starts giving unreliable feedback. Since the feedback is important for how the parameters are set, an increase in the time it takes to find the correct parameters may have a negative impact on the final result or require more future programming sessions.

There are many other areas (e.g. cancer detection and classification) that have benefited from using a machine learning model that can make accurate predictions based on patient data [5-8]. Using a prediction model to predict the CI parameters could potentially automate the programming session, or at least reduce the programming time. If programming time is reduced, it would not only benefit the patient, but also reduce the total cost of a CI as the audiologist needs less time for each patient. Reduced programming time may also improve speech recognition scores, especially for young children where the period of giving reliable feedback can be very short. Further, infants almost never give a reliable feedback so an automatic adjustment for these would be of great help. Even for adults with post-lingual deafness[1], it can be difficult to give an accurate estimate of loudness. This makes the programming especially challenging for clinics and audiologists with little experience in programming cochlear implants who might have difficulty programming an accurate configuration and providing adequate sound for patients. This paper addresses this challenge by training models for predicting adjustment value settings using data from former patients. We are not aware of any earlier work applying such an approach.

In the next section an introduction to the CI operation and configuration will be followed by a description of the former patient data. Relevant machine learning techniques are outlined in section III. Results of the experiments and discussion are given in section IV and conclusion in section V.

II. CI BACKGROUND

Receiving a CI requires the implant to be surgically implanted behind the ear. The surgery is, however, only one part of the process. After the surgery a sequence of sessions with an audiologist will follow. During these sessions the audiologist tries to as best as possible to adjust a number of stimulation level parameters (minimum and maximum stimulation levels for each electrode, respectively) of the speech processor according to the sensitivity of the patient. For each electrode, the two following values are to be set:

- T-levels (Threshold): Minimum level of current
- C-levels (Comfort): Maximum level of current

The aim is to make the patient hear as well as possible but limit the maximum stimulation levels to avoid discomfort. How well these parameters are set can have a large impact on hearing quality [9]. In order to set these parameters, the audiologists use data measured during and after the surgery. This is commonly referred to as *objective data,* and are outlined in section A below. Since the objective data is not sufficient to make an accurate setting, oral feedback from the patient is also used; this is commonly referred to as *subjective data*. As mentioned in the introduction, sometimes it is difficult to get any useful feedback from the patient for various reasons. Young children are especially challenging, since they usually do not give accurate feedback on how well they perceive a sound. Thus, there is a risk of giving too high stimulation levels which can be uncomfortable for the child and may cause the child to refuse to use the sound processor. It is also important not to set the parameters too low, because then the stimulation of the nerve fibres may be too weak. This is especially important for young children since hearing is important for how well they learn speech and language at an early age. Studies have shown that the earlier children get their implant, the closer they can come to the auditory skill of children with normal hearing [10].

Fig. 2. Measurement of objective data during CI surgery.

A. Objective Data

Since C-level and T-level are based on subjective behavioural experiences, this can cause some inaccuracy. To complement the subjective data, it is desirable to use objective measures as well during the programming process. Therefore, a set of objective measurements are performed during and after surgery, see Fig. 2. The idea is that these measurements can say something about what levels to use without having to rely on feedback from the patient. The objective data includes the following:

- ESRT – Electrically evoked stapedius reflex threshold (contraction of a tiny muscle in the middle ear in response to loud sounds)
- ECAP – Electrically evoked compound action potential (auditory nerve activity in response to an electrical stimulus, that can be recorded in situ on cochlear implant (CI) using one electrode to stimulate the nerve and the neighboring electrode to measure the response)
- Impedance (normally used to check whether all electrodes work correctly)
- Age

There exist some studies of the correlation between these measured values and appropriate C-/T-values. The relationship between hearing quality and the ECAP for CI users with short electrode array has been investigated [11, 12]. However, the results have generally shown poor correlation. The ECAP level has also been used as an estimator for C-levels, but with limited success [13]. However, others have found a positive correlation between ECAP and C-/T-levels for infants [14]. There seems to be some variations for how useful various clinics find these parameters which may come from variations between the CIs or clinical practises.

[1] Post-lingual deafness is deafness that occurs after the development of speech and language.

B. Former Patient Data Set

Data from 300 patients from Oslo University Hospital was made available for this work. However, some of them were excluded, e.g., due to missing objective measurements or unreliable oral feedback from patients at a young age, and the data set applied consisted of 158 patients.

The data for each patient consists of the objective data presented in the previous section – the measured ESRT, ECAP and impedance together with age and the corresponding C- and T-levels found by the audiologist. Each of the measurements can be collected for each electrode, in practice, however, the measures are only collected for every second electrode to save time. There are two sets of measures of impedance, C-level and T-level, one from the initial programming and one from the stable programming found after multiple sessions (being averaged for each patient in the experiments).

There are a number of different cochlear implants available (four main manufacturers, each with different cochlear implants). To make sure that the data available is comparable, only data from patients with comparable devices have been used. The device type used in our data set is the Cochlear® CI24 containing 22 electrodes. There are a few variations of this CI, the electrode design is either the standard Contour® electrode array or the Contour Advanced Soft Tip®. The housing type is either 24R, Freedom, or CI512. These variations should not have any effect on the data or stimulation strategy used.

A set of C-levels and T-levels for all electrodes is called a MAP (named from the word MAPping). A patient can have multiple MAPs depending on preferences. Having multiple MAPs lets the patient change C- and T-levels depending on what works best in the current situation. Since there is no information about which MAP the patient has used, the C- and T-levels used in this paper is the *average* of all the MAPs for each patient.

C. Restoring Missing Values

Since the dataset has many missing values, these need to either be handled by the learning algorithm or corrected for beforehand. Since the machine learning methods used are not all able to handle missing data well during training, the missing data was handled *before* training. By looking at the correlation matrices in Fig. 3 and Fig. 4, it is clear that close by electrodes have fairly similar values. This means that if we have the adjacent value of a missing value, we should be able to find a reasonable accurate estimate for the missing one. Thus, to restore the missing values, interpolation was used. The interpolation was performed with Mathworks ® MATLAB® using the function *interp1* with Piecewise Cubic Hermite Interpolating Polynomial (PCHIP).

The number of missing values varies from patient to patient, some have a measure on almost every electrode and some have on almost none. Because of this, we may end up interpolating with very few values, if there is no requirement for how many missing values we allow.

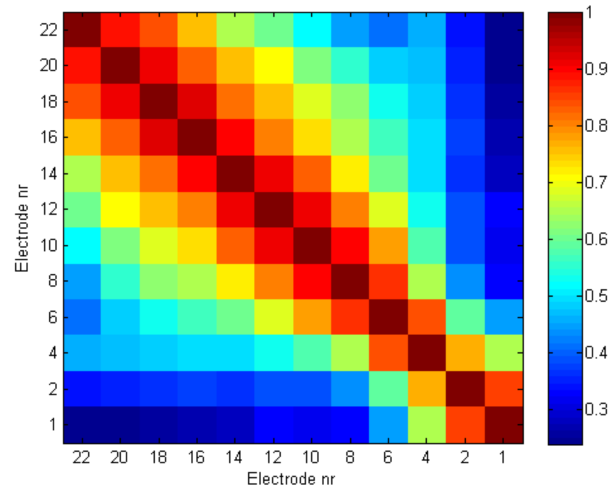

Fig. 3. Correlation plot for ESRT. The plot only shows the correlation between pairs of electrodes where the objective measures are not missing.

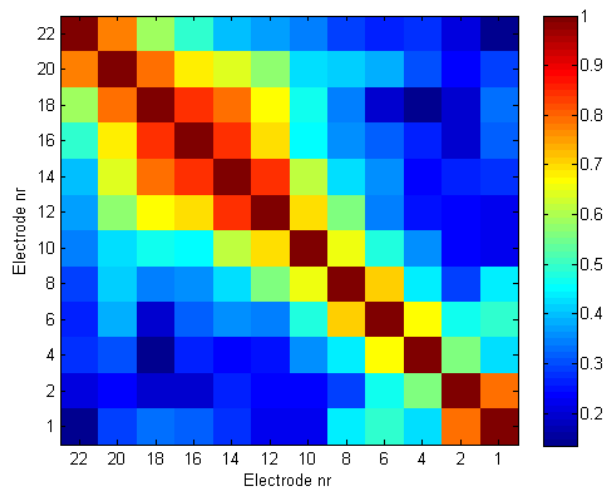

Fig. 4. Correlation plot for ECAP. The plot only shows the correlation between pairs of electrodes where the objective measures are not missing.

Interpolating with a very small number of values is undesirable, since it is likely to yield values that are far off from their actual value. Many missing values may also be an indication of possible measurement outliers. Even though we do not know the reason why values are missing, it is reasonable to assume that many missing values could have been caused by some complication or anomaly. To avoid interpolating these samples, we need to limit how many missing values that should be allowed. Having at least 8 samples of each type of measure seem reasonable (given experience from manual adjustment); this means that the allowed number of missing values for each measure is 14 for each objective measure. With this requirement there were data from 20 patients (out of the 300) that couldn't be interpolated.

III. MACHINE LEARNING METHODS

The key focus of this paper is to investigate which sort of algorithms and features may be best suited for making models that predict the C- and T-levels and to determine the achievable accuracy. Three different and widely used machine learning models have been applied in this work to make C/T-level predictions:

- *Linear Regression*
- *Support Vector Machine (SVM)*
- *Artificial Neural Networks (ANN)*

Linear regression is a method that tries to find a linear function to fit some data, by minimizing the sum of the squared error. To further increase the performance, *ensemble learning* is applied using multiple prediction models that, when combined, can provide better performance than a single model. We have used the variant called *boosting* where the training data are given different weights. The combination of models is typically by averaging or voting with the former used in this work.

The SVM used soft margins and a radial basis kernel function. The Levenberg-Marquardt backpropagation training algorithm was used for the ANN [15]. The network has one input neuron for each feature, three neurons in the hidden layer and one output neuron for C- or T-level. A search for optimal parameters in ANN and SVM was undertaken using cross-validation on the training set.

The input to each model consists of three objective data measurements (see section II-A) for each of the 22 electrodes. In addition, age is also an input, so a total number of 67 inputs were used. Rather than making a single model for all 2 × 22 levels, we train one model for each level to be predicted (and end up with 44 models in total). Linear regression is implemented using the MATLAB Curve Fitting Toolbox™, the ANN is from the MATLAB Neural Network Toolbox™ and the SVM is from LIBSVM[16]. Further details on these implementations can be found in [17].

The error of a *constant* function will be listed together with the performance of the trained models for comparison. The predicted value of the constant function is the mean of all the C- or T-levels for that electrode. Thus, for the constant function, we will always predict the same value for a given electrode.

A. Accuracy Measure

The C-level and T-level are set to a given *current level* (CL) out of which 256 levels are available where the C-level always would be larger than the T-level. The mean C-level in the data set is approximately CL equal to 180, while the mean of the T-level is approximately CL equal to 100. The variance is rather large with max and min values being typically more than 50 CL above and below the mean values.

To give an indication of the model accuracy, an error measure with a certain error tolerance will be used. The two error measures of the model will be the percentage of errors below 5 and 10 CL, respectively, see Fig. 5. That is, the percentage of test vectors having an absolute value below 5 and 10 CL, respectively. The reason for this is that a discrepancy of 5 CL from the correct value is close enough that the patient would often not notice the difference. Case studies at our clinic have shown that programming levels vary during programming sessions on the same day. Thus, asking the patient twice can often lead to different values. Therefore, if the error is less than 5 CL, we say it is a correct prediction. Mean absolute error will also be used to see how the overall error of a model changes. The error measure is defined as follows:

Mean absolute error (MAE) =

$$\frac{1}{n}\sum_{i=1}^{n} |(h(x_i) - Y_i)|$$

Threshold measure:

$$\frac{1}{n}\sum_{i=1}^{n} (|(h(x_i) - Y_i)| < threshold)$$

where $h(x_i)$ is the output we get from testing the model on data x_i, for electrode i. Y_i is the correct value (C- or T-level) for electrode i. If nothing else is specified, the error is measured using *leave-one-out* cross-validation where 20% of the data set is used as a test set after training. Thus, experiments are repeated five times with different test sets.

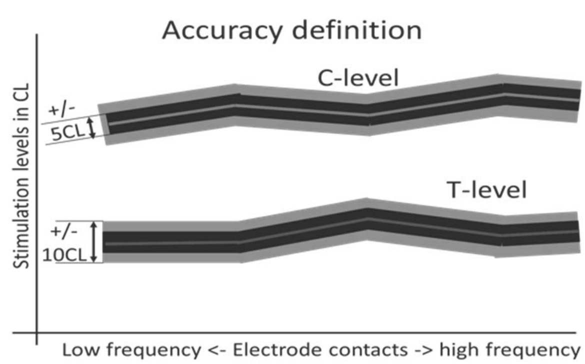

Fig. 5. Illustration of what area of errors that is covered by an error measure with threshold 5 and 10 CL. The percentage of error related to the area marked with 5 CL in the figure will be the key criteria in the following experiments.

B. Feature Selection

There is a large amount of uncertainty in the relationship between the objective data measurement and C-/T-levels. Although some studies have found relationships between the objective data and the CI parameters, the relationship has not been found to be strong enough for accurate parameter prediction. To find the most optimal features, a variety were tested in the experiments (see section IV-A). These consisted of one or a combination of several of the various objective data being available in the data set. For the later experiments, the best features found were used.

IV. RESULTS

In this section, the experimental results will be presented. First, the results of prediction using only objective measurements will be described, followed by results when one or two electrode measurements are known in advance and, finally, when objective measurements are combined with known electrode measurements.

A. Best Model with only Objective Measurements

The first experiments used linear regression and only the objective measures to predict C- and T-levels. The results show that some objective measures are more useful than others, see Fig. 6 and Fig. 7, respectively. The best accuracy is from using ESRT and Impedance together for C-levels and ESRT for T-levels. Using ESRT and Impedance for C-levels give 17.2% accuracy (5 CL) which is an improvement of 6% compared to a constant function with a performance of 11.2%. The best accuracy for T-levels estimation is 18.4%, which is a 1.9% improvement compared to a constant function (16.5%). This tells us that T-levels are easier to predict, and the objective measures are most useful when predicting C-levels. It is not beneficial to apply as many objective parameters as possible. The reason for this may be that the model easily becomes overfitted; i.e., the model specializes too much on the training set so the performance on untrained data is reduced.

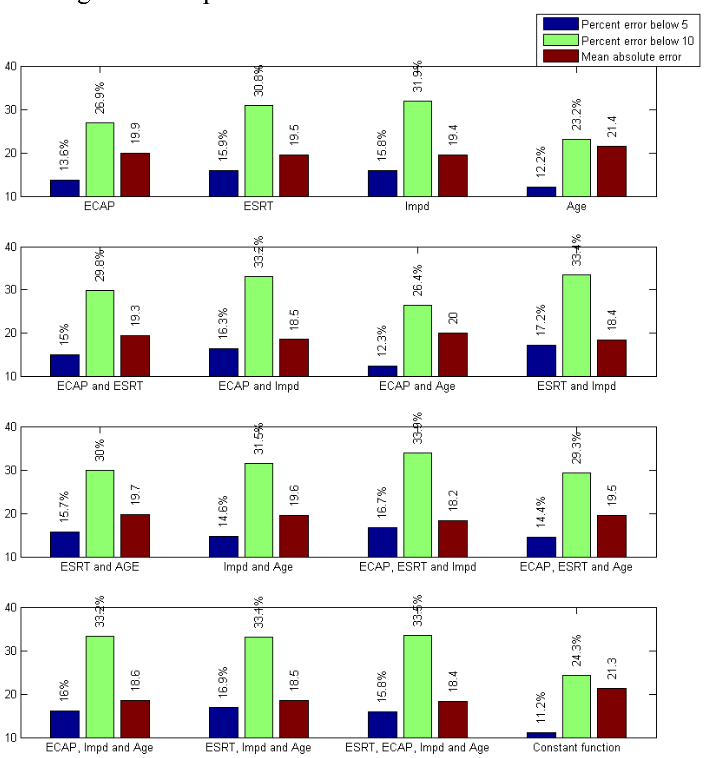

Fig. 6. Cross-validation error when predicting C-levels for all combinations of features. What features that were used are listed under the bars. All models were created using linear regression.

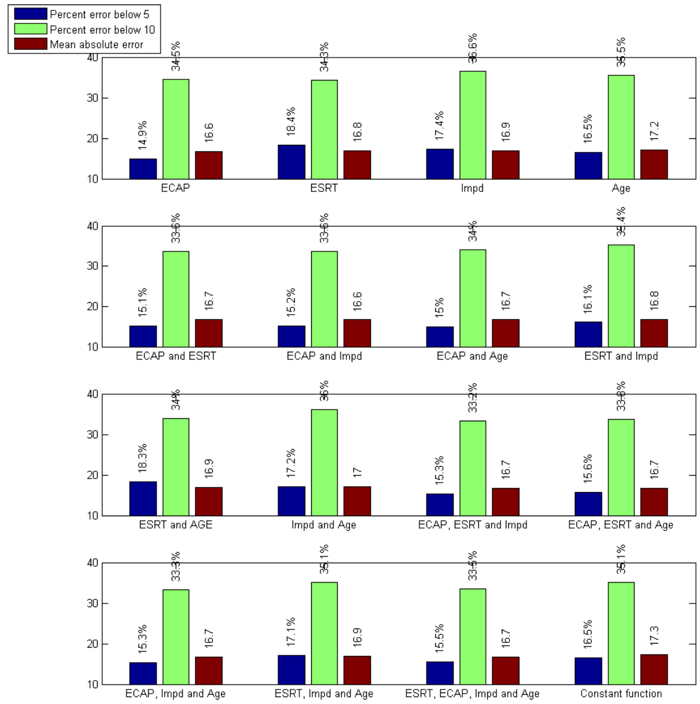

Fig. 7. Cross-validation error for combination of features, using linear regression for predicting T-levels.

The performance achieved using the SVM and ANN was similar to the linear classifier and is, therefore, not included here. A more robust model was created using *ensemble learning* based on multiple linear regression models and boosting. The boosted model is also likely to be more robust than the other models and less likely to lead to overfitting, which is of importance in clinical use. The ten feature combinations giving the best performance in the initial experiments were combined using ensemble learning. Normally each would be trained on parts of the training set but since our training set is rather small, we train each on the whole training set. As the error dropped slightly more using this model, it was applied in the last experiments (see section B-c).

B. Prediction with Known Parameters

Since using objective data alone does not seem to give a high enough accuracy for automatic parameter prediction, an alternative method was tested where the levels for one or more electrode of the CI were assumed known and could therefore be used as input to the model. If we assume that an audiologist is able to measure one or two C- and T-levels of our choosing, we can then make a model that can predict the remaining C- and T-levels. To find which electrodes are optimal to use for the model, each was tested using cross-validation on the training set. The score of a model using linear interpolation is used as a reference to better understand how our model performs compared to the simplest possible technique. In this context, interpolation refers to a simple straight line through the known points (i.e., known electrode measurements). In cases where we only have one known electrode, a constant function is used.

a) Prediction using Knowledge about Levels for One Electrode

The results for predicting C-levels when the C-level is known for one electrode show that model performance greatly depends on which electrode is known. Fig. 8 shows the errors when using one electrode at a time as input to predict C-levels of the other electrodes with a linear regression model using ensemble learning. The x-axis is the input electrode used to create the model, while the y-axis is the error when using it as input. The electrodes in the middle give the best result. The best performance of 61% is obtained (5 CL) when applying electrode 15 as a part of the prediction model.

Similar performance is seen for the T-level in Fig. 9 with the best performance of 66% when using electrode 9. Although a different electrode number than for the C-level, they are both among the electrodes in the middle of the electrode array. The results from using the SVM and ANN were only marginally better and so are not included here.

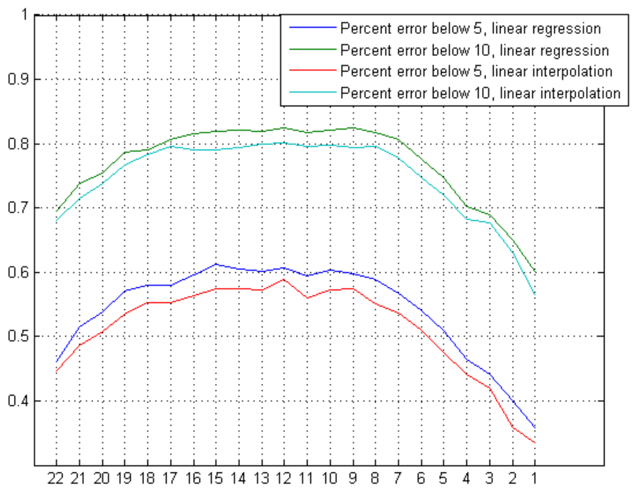

Fig. 8. Errors when using one electrode at a time as input to predict C-levels with a linear model. The x-axis is the input electrode used to create the model, while the y-axis is the error when using it as input.

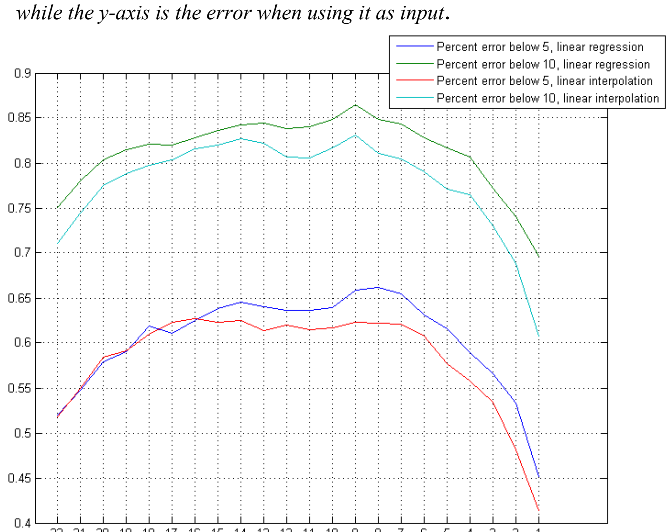

Fig. 9. Errors when using one electrode at a time as input to predict T-levels. The x-axis is the input electrode used to create the model, while the y-axis is the error when using it as input.

In conclusion, using one electrode measurement did improve the results significantly in all the experiments. The best results differed for the C- and T-levels, but in both cases were provided by an electrode in the middle of the range of electrodes.

b) Prediction using knowledge about levels for two electrodes

We continued the investigation to see if the performance further increased when using two known electrode measurements. A selection of the electrode combinations were tested and resulted in substantial improvement. The best known parameter pair seemed to be electrodes 5 and 15 when two C-levels were available. This gave a performance of 78.3% which is close to 17% improvement compared to using only one electrode. For T-level, the best known electrode combination was of electrodes 5 and 17 providing a performance of 78.2%, presenting a 12% improvement from using only one electrode.

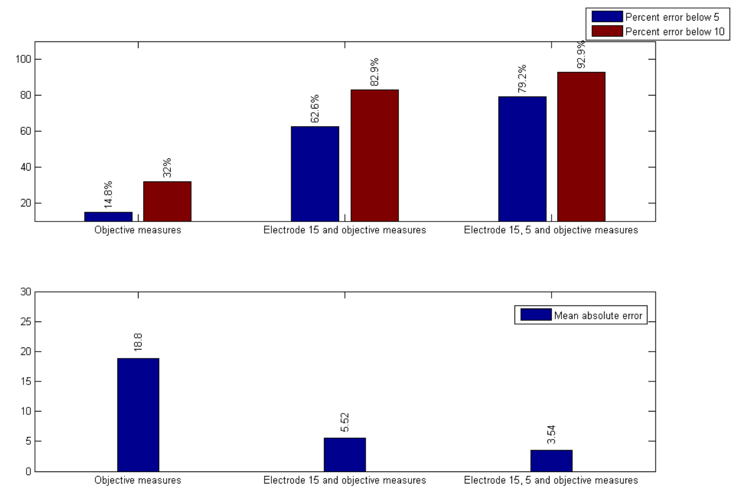

Fig. 10. Predicting C-levels using objective measurement with a varying number of known, optimal electrodes.

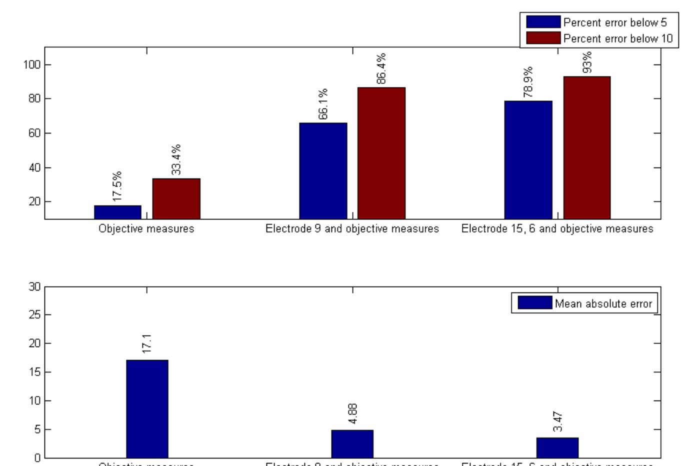

Fig. 11. Predicting T-levels using objective measurement with a varying number of known, optimal electrodes.

Although various models were tested in this investigation, they did not result in much difference in performance. The rather simple linear regression model seemed to be competitive compared to the more advanced models for the CI data set used in this work.

c) Prediction Using Best Known Electrode Parameters and Objective Data

We have now seen the possible performance using objective data (section IV-A) or using one or two electrode values (previous sections). This section will combine the two approaches to see how that impacts the performance. The best features found in the previous sections will be combined using ensemble learning as outlined in section IV-A. The results for using ensemble learning of multiple linear regression models and boosting are shown in Fig. 10 and 11 for the C- and T-level, respectively. For both levels, we see that the mean absolute error is decreased as electrodes are introduced to the model. The best performances of 79.2% and 78.9% for cross-validation on the training set for C- and T-level are regarded as very satisfactory. In other words, by having the audiologist make two measurements for each level, the rest of the electrode levels can, to a large extent, be accurately predicted. This would be of major importance especially for young patients. The main contribution to the satisfactory performance is knowledge about one or two electrode settings since the addition of objective measurements only increases the performance by around 1%.

In this study, more advanced machine learning techniques such as SVM and ANN only provided marginally better performance. The lack of need for these may be explained by the data set being limited in the number of patients.

V. Conclusion

This paper has been concerned with using data collected from earlier cochlear implant patients to predict the programming levels for new patients. The results show that the data measured on earlier patients – ESRT, ECAP, impedance and patient age – cannot be used to make predictions accurate enough for a completely automatic fitting process. However, the results show that it was possible to achieve a useful accuracy if combined with a few patient measurements. That is, if a few of the CI parameters for a new patient can be measured, they can be used to develop a model that can predict the rest of the parameters with a fairly high accuracy. In addition to SVM, ensemble learning of multiple linear regression models provided high performance.

During the experiments, the optimal electrodes for interpolation and prediction were also found. Which electrodes were chosen seemed to have a large impact on the model performance, and this was independent of the model used. This means that even if the prediction model proposed in this study is not used, it is still beneficial to use these optimal electrodes when using, e.g., linear interpolation. Further, using the model proposed in this paper gives lower prediction error than linear interpolation which is commonly used today. The results of this work suggest ways to reduce the time for audiologist parameter tuning as well as improve the quality of the tuning. This would be beneficial for both the patient and the clinic.

Acknowledgment

This work is partially supported by The Research Council of Norway as a part of the Engineering Predictability with Embodied Cognition (EPEC) project, under grant 240862.

References

[1] WHO. Deafness and hearing loss. [October 11, 2014]; Available from: http://www.who.int/mediacentre/factsheets/fs300/en/.

[2] Blamey, P., et al., Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: an update with 2251 patients. Audiol Neurootol, 2013. 18(1): p. 36-47.

[3] Lazard, D.S., et al., Pre-, Per- and Postoperative Factors Affecting Performance of Postlinguistically Deaf Adults Using Cochlear Implants: A New Conceptual Model over Time. PLoS One, 2012. 7(11): p. e48739.

[4] Baudhuin, J., et al, Optimization of Programming Parameters in Children with the Advanced Bionics Cochlear Implant. J Am Acad Audiol. 2012 May; 23(5): 302–312.

[5] Thekkek, N. and R. Richards-Kortum, Optical imaging for cervical cancer detection: solutions for a continuing global problem. Nature Reviews Cancer, 2008. 8(9): p. 725-731.

[6] Tan, A.C. and D. Gilbert, Ensemble machine learning on gene expression data for cancer classification. Appl Bioinformatics. 2003;2(3 Suppl):S75-83, 2003.

[7] Charasse, B., et al., Automatic analysis of auditory nerve electrically evoked compound action potential with an artificial neural network. Artificial Intelligence in Medicine, 2004. 31(3): p. 221-229.

[8] Garg, A.X., et al., Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. Jama, 2005. 293(10): p. 1223-1238.

[9] Wasowski, A., et al., Influence of non-optimal levels of electrical stimulation in cochlear implantees on hearing benefits. Cochlear Implants Int, 2010. 11 Suppl 1: p. 485-8.

[10] Tomblin, J.B., et al., The effect of age at cochlear implant initial stimulation on expressive language growth in infants and toddlers. J. of Speech, Language, and Hearing Research, 2005. 48(4): p. 853-867.

[11] Kim, J.R., et al., The relationship between electrically evoked compound action potential and speech perception: a study in cochlear implant users with short electrode array. Otol Neurotol, 2010. 31(7): p. 1041-8.

[12] Cosetti, M.K., et al., Intraoperative neural response telemetry as a predictor of performance. Otol Neurotol, 2010. 31(7): p. 1095-9.

[13] Potts, L.G., et al., Relation between neural response telemetry thresholds, T- and C-levels, and loudness judgments in 12 adult nucleus 24 cochlear implant recipients. Ear Hear, 2007. 28(4): p. 495-511.

[14] Muhaimeed, H.A., et al., Correlation between NRT measurement level and behavioral levels in pediatrics cochlear implant patients. Int J Pediatr Otorhinolaryngol, 2010. 74(4): p. 356-60.

[15] Suratgar, A.A., et al., Modified Levenberg-Marquardt method for neural networks training. World Acad Sci Eng Technol, 2005. **6**: p. 46-48.

[16] Chang, C.-C. and C.-J. Lin, LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2011. 2(3): p. 27.

[17] Iversen, A.H. Use of Artificial Intelligence and Machine Learning algorithms to predict programming levels in Cochlear Implant patients. Master's thesis, University of Oslo, 2014. Available from: https://www.duo.uio.no/handle/10852/42385