AuDiET: Auditory Diagnostics and Error-based Treatment - Towards performance-based fitting

Enrico Migliorini[1,2], Nikki Philpott[2,1], Jan-Willem Wasmann[2], Bas van Dijk[1], Birgit Philips[1], Emmanuel Mylanus[2], Wendy Huinck[2]

1) Cochlear Technology Centre Belgium, Mechelen, Belgium
2) Department of Otorhinolaryngology, Donders Institute for Brain, Cognition and Behaviour, Radboud university medical center Nijmegen, the Netherlands

Unexplained poor performance in Cochlear Implant (CI) users is a problem which is as hard to predict as it is to address. People who could be expected to perform well with a CI may turn out to have suboptimal speech recognition with their implant.

Reducing unexpected poor performance has been difficult so far. **Numerous factors** contribute to speech perception, and several of them, such as aetiology, cannot be intervened upon.

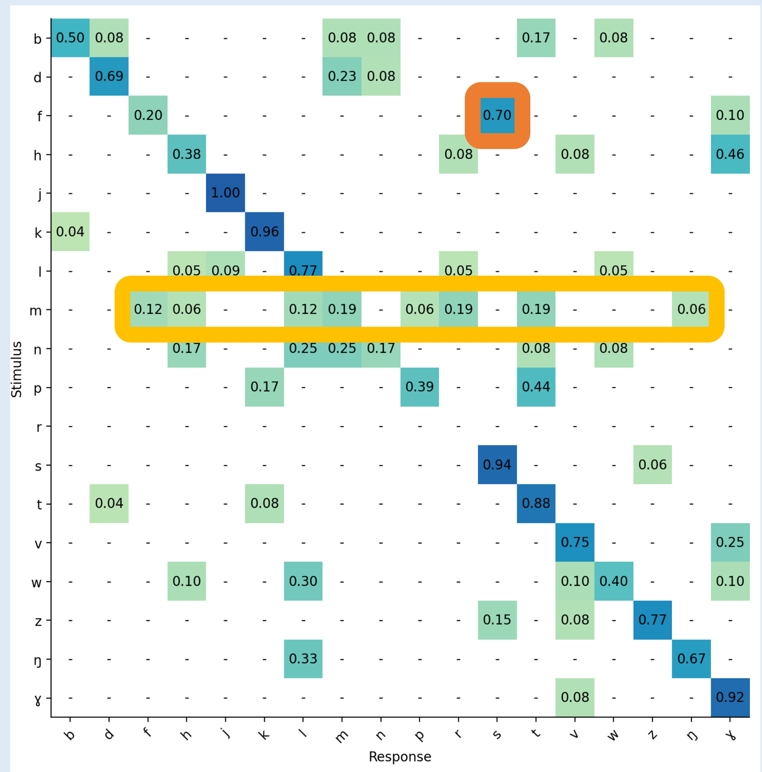
What can, however, be done, is to move towards **individualized care**. Implant fitting and post-intervention rehabilitation are left to the discretion of hospitals and clinics. The lack of standardised, evidence-based clinical practices can then result in **specific individual needs** not being effectively addressed.

Because of this, unexpected poor performance is difficult to address effectively, leaving users dissatisfied with their implants. The Auditory Diagnostics and Error-based Treatment (AuDiET) study aims to change this situation by taking steps towards evidence-based clinical practices.

How to spot error patterns?

The **Confusion Matrix** is valuable Data Science tool, useful for analysing in detail which phonemes a CI user is struggling with.

It is a target-response plot in which each row and column corresponds to a phoneme in a recognition test. Each row corresponds to a presented stimulus, and each column to a response.



The value of cell (X, Y) corresponds to the percentage of times that the subject, when presented with the phoneme X, reported hearing Y. Values on the diagonal correspond to correct answers.

A Confusion Matrix can help us separate Systematic errors from Random ones.
Systematic errors are the ones where a phoneme is predominantly identified as a different one. For example, in the matrix above /f/ is misclassified as /s/ in 70% of occurrences (highlighted in orange).

Random errors are instead those where instead multiple different errors are made for a single phoneme. For example, in the matrix above /m/ is misclassified as 8 different phonemes (highlighted in yellow).

The assumption driving the AuDiET study is that **different error patterns should be addressed differently.**

The AuDiET study aims at providing a proof of concept for individualized interventions. The study population comprises 25 adult, post-lingually deafened, Dutch-speaking experienced CI users.

During Visit 1 each subject first undergoes a battery of tests so that their unique error profile can be collected. These tests include:

- Tone Audiometry
- Spectrotemporal Assessment
- Phonemes in Quiet tests
- Consonant-Vowel-Consonant Words in Quiet
- Digits in Noise

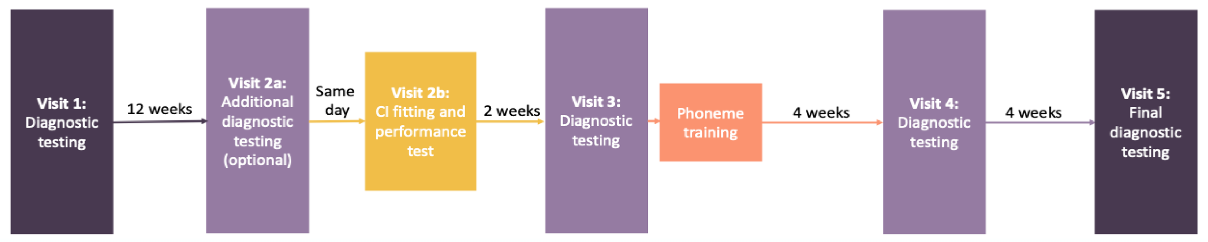
Based on the results of these tests and the current fitting profile, a Fitting Intervention is drafted by two experienced audiologists. This intervention modifies the fitting parameters in order to address their most common errors. During Visit 2 this intervention is loaded onto each subject's processor.

During Visit 3 the subjects are tested again in order to evaluate the effects of the Fitting Intervention.
Additionally, during the visit the subjects are given a mobile application which will provide them with personalized training exercises focusing on their most common errors. This is the Training Intervention.

During Visit 4 the tests are repeated in order to evaluate the effectiveness of the Training Intervention.

During Visit 5 the subjects are tested again, in order to evaluate whether the subjects retain any change in speech recognition when not actively training.

Should fitting prove impactful, the relationship between changes to fitting and speech recognition will be investigated in detail, with the aim to move towards **evidence-driven, targeted fitting strategies** to form the basis of **standardised clinical practices**.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 860718

Interim Data Analysis – Quantification of Error Patterns

- **Accuracy (percentage correct): 70% for both Stephen and Randy**

Accuracy is insufficient to show the differences between the two, therefore we introduce two additional measures.

Weighted Accuracy

Let $f_{i,j}$ be the frequency of the confusion between phonemes i and j. Let $d(i, j)$ be the distance between those phonemes, using a weighted measure similar to the one used by Preston et al[1]. The weighted accuracy can then be defined as $\sum_i (1 - \sum_j f_{i,j} d(i,j))$. In other words, the larger the errors made, the lower the weighted accuracy.

- **Weighted Accuracy: 90% for Stephen, 88% for Randy**

Note that Accuracy forms a lower bound for Weighted Accuracy.

Error Dispersion

Derived from the paper by van Son[2], this measure is based on Shannon's entropy, i.e., the mean amount of information conveyed by each symbol in a sequence. It "*can be interpreted as the effective number of error classes per stimulus token*".

- **Error Dispersion: 1.0 for Stephen, 3.0 for Randy**

Error Dispersion correlates negatively with accuracy ($r < -0.9$, $p < 0.001$) in both PRQ and CVC tests. This is intuitive as a smaller number of errors will likely be spread over fewer options.

Example 1: "Stephen"

p	0.7	0.3	-	-
b	0.3	0.7	-	-
t	-	-	0.7	0.3
d	-	-	0.3	0.7
	p	b	t	d

Example 2: "Randy"

p	0.7	0.1	0.1	0.1
b	0.1	0.7	0.1	0.1
t	0.1	0.1	0.7	0.1
d	0.1	0.1	0.1	0.7
	p	b	t	d

Weighted confusion matrix

p	-	0.1	-	-
b	0.1	-	-	-
t	-	-	-	0.1
d	-	-	0.1	-
	p	b	t	d

p	-	0.03	0.03	0.06
b	0.03	-	0.06	0.03
t	0.03	0.06	-	0.03
d	0.06	0.03	0.03	-
	p	b	t	d

Subject	Accuracy	Weighted Accuracy	Error Dispersion	Systematic Errors	Random Errors
S04	65.4%	81.1%	0.94	/f/ - /v/ /p/ - /k/	/d/ /l/
S05	65.4%	80.4%	1.06	/n/ - /m/ /v/ - /z/	/f/ /l/
S14	65.4%	82.6%	0.88	/k/ - /t/ /v/ - /z/	/f/ /n/

The data above is limited to Phoneme tests in quiet for consonants

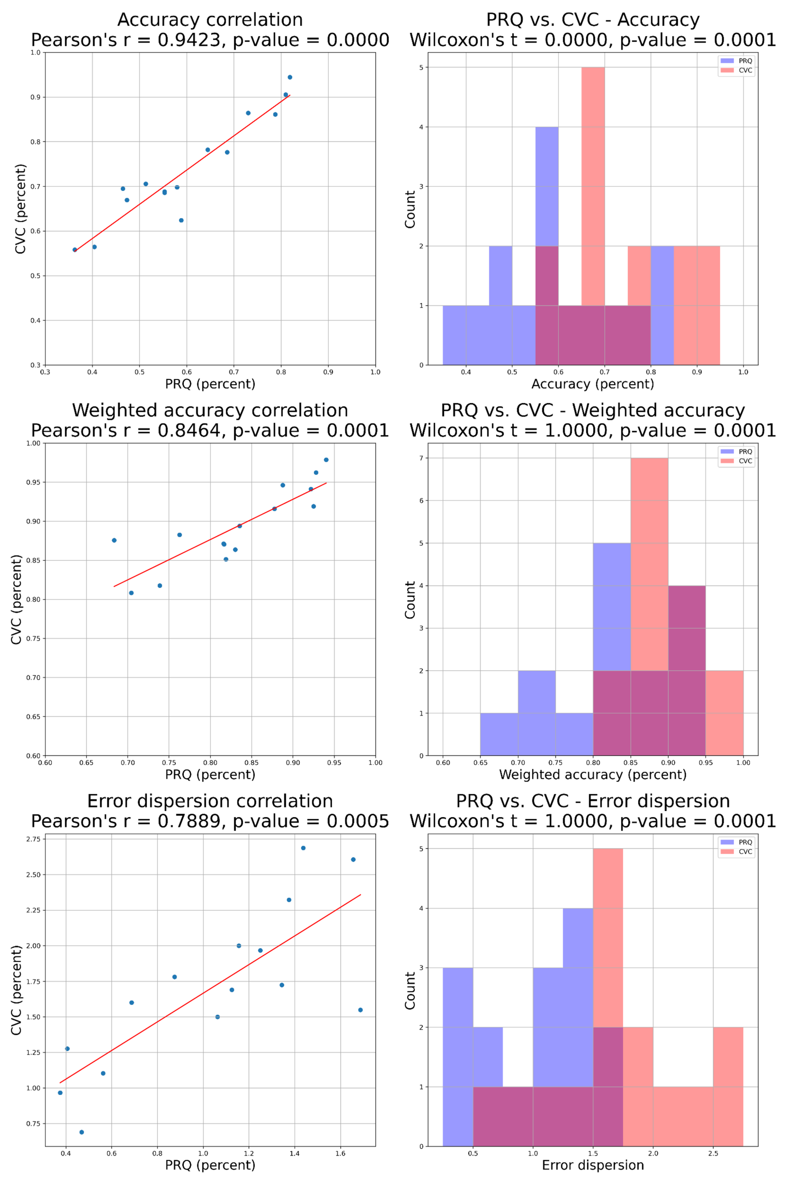
These measures do not describe qualitatively error patterns.

- Strong correlation with accuracy (*|r|>0.9, p<0.00001*)
- High dimensionality of data points
- No indication of features of confused phonemes

We need to **represent differences in error patterns (whether between or within subjects) as a low-dimensional variable for statistical analysis.**
Under consideration: grouping by (IPA) features; limiting analysis to errors addressed in interventions; expert case-by-case interpretation.

We would love to hear your thoughts on the topic!

Interim Data Analysis – Phonemes vs. Word tests



Phoneme Recognition in Quiet (PRQ) tests: *closed* set recognition of *nonsense* triphones of the form /aCa/ or /hVt/, with C and V being Consonant and Vowel respectively.
Word Recognition in Quiet (WRQ) tests: open set recognition of meaningful triphonemic Consonant-Vowel-Consonant (CVC) Dutch words.

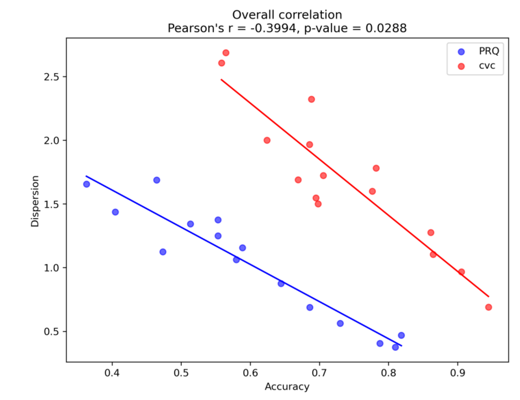
How do PRQ and CVC scores relate to each other?

- PRQ and CVC scores all strongly correlate (r and p values in the graph).
- PRQ and CVC scores come from meaningfully distinct distributions (Wilcoxon's signed rank test).

The average score for word tests is significantly higher than that for phoneme tests. This is likely due to the addition of meaning, leading the subject to correct errors that would result in nonsensical words.

On the right the accuracy and error dispersions for the PRQ and CVC tests are plotted separately, with their individual regressions ($r < -0.9$). Despite having higher accuracy, CVC tests also have higher error dispersion. Potential reasons include:

- Context mitigating systematic errors
- Context leading to cognition-related errors
- Set openness
- Potential bias for closed set options



A better representation of error patterns will lead to further insight into these results.

References

1) Preston, J. L., Ramsdell, H. L., Oller, D. K., Edwards, M. L., & Tobin, S. J. (2011). Developing a weighted measure of speech sound accuracy.
2) Van Son, R. J. J. H. (1995). A method to quantify the error distribution in confusion matrices. In EUROSPEECH.

emigliorini@cochlear.com

scan me
www.mosaics-eid.eu

Radboudumc
Cochlear®
MOSAICS