Model-based validation framework for coding strategies in cochlear implants

Michele Nicoletti, Michael Isik and Werner Hemmert

Technische Universität München, IMETUM, Bioanaloge Informationsverarbeitung

michele.nicoletti@tum.de

IMETUM Zentralinstitut für Medizintechnik

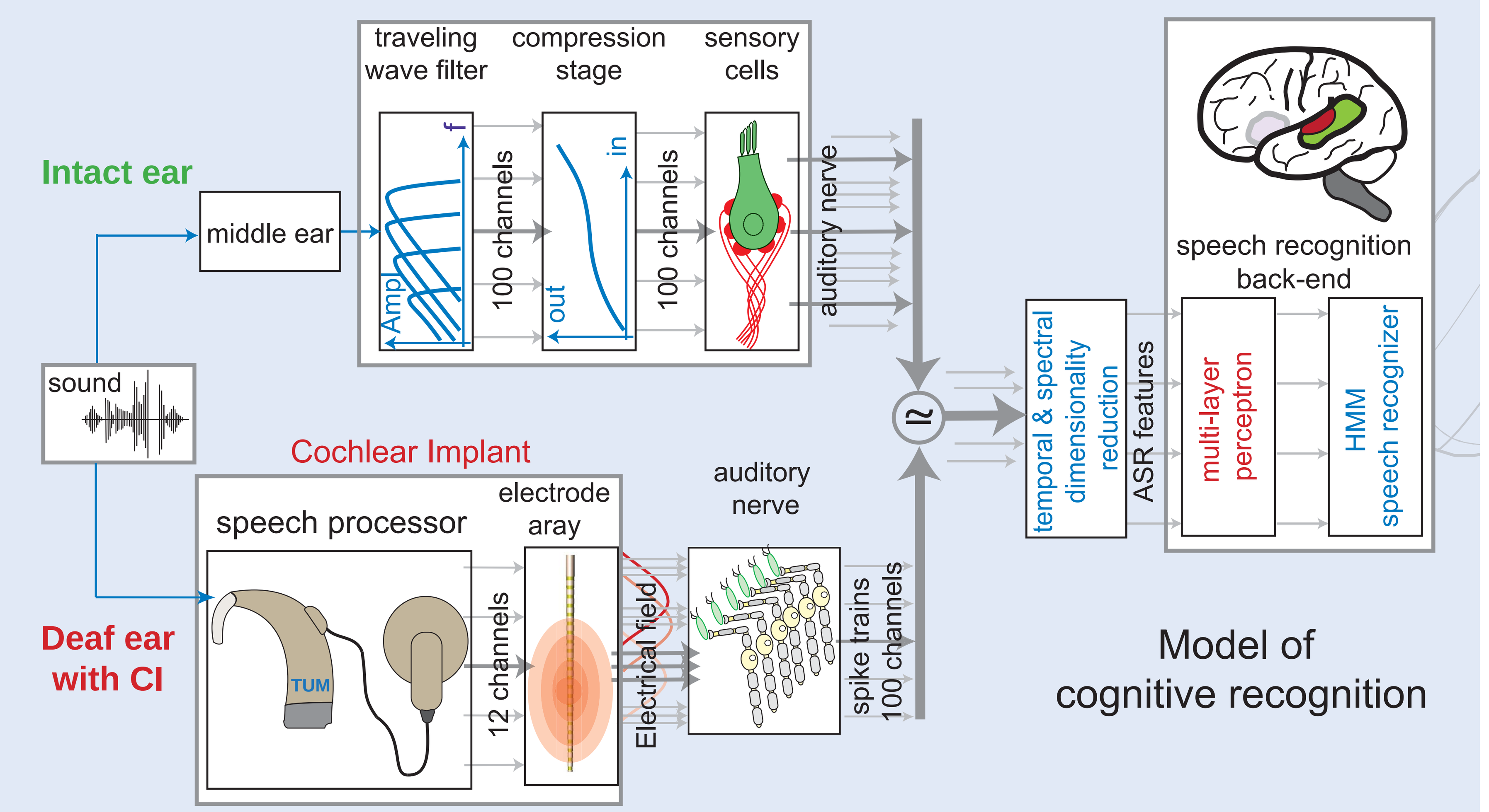
Bernstein Center for Computational Neuroscience Munich

Bio-Inspired Information Processing

Abstract

Although modern cochlear implants (CI) are able to restore speech perception to a high degree, there is still a large potential for improvements e.g. in music perception and speech discrimination in noise. To evaluate and optimize novel coding strategies, we have developed a toolbox which codes sound signals into spike-trains of the auditory nerve. We have previously developed a model of the intact inner ear, which we have complemented with detailed models of a CI speech processor, channel crosstalk and spiral ganglion neuron models. With our toolbox we present qualitative comparisons of neurograms elicited by different coding strategies with the situation in the healthy inner ear. Moreover, we conducted quantitative evaluations using two methods: i) With the framework of automatic speech recognition we evaluated speech discrimination using a noisy database. ii) With the methods of information theory we quantified the transmitted information coded in neuronal spike trains, which allows us to evaluate especially well how well temporal information is coded. The major advantage of our approach is that we are able to evaluate both spectral- and temporal aspects of novel coding strategies before we conduct extensive clinical studies.

Model of acoustic and electric hearing & interface to ASR



The upper figure shows the schematic outline of our models for acoustical (intact inner ear) and electrical (inner ear with CI) stimulation. Our inner ear model separates sound signals into 100 frequency channels. In each channel, a sensory cell converts the band-passed and compressed signal into spike trains of multiple auditory nerve fibers (ANFs). For the implanted ear, auditory nerve fibers are modeled with single- or multi-compartment models with Hodgkin-Huxley like ion channels. Here we focus on a model of spiral ganglion type I neurons with Hodgkin-Huxley type ion channels, which are also found in cochlear nucleus neurons (HPAC, Kht, Klt). Their large time constants might be responsible to explain adaptation to electrical stimulation (Negem 2008). We corrected conductances and time-constants to a body temperature of 37° and solved the differential equations in the time domain with the Crank–Nicolson method and an exponential Euler rule. Depending on the task, we model the neurons at different levels of detail.

The electrode was modeled as an array of 12 current point sources (I) at a distance of 0.5 mm | (x y z) | from the spiral ganglion cells (SGC). The coupling between electrode and excitation of the neuron is described by the activation function (second derivative of the extracellular potential (V) with respect to the neuron's path x). Cannel cross-talk is modeled by a field equation based on an anisotropic medium, described by a resistivity tensor ρ_{xyz}.

The channel crosstalk results from a lower resistance ρ_y along the electrode, in this way the field will be distorted so that we obtain a symmetrical spread of excitation with a slope of approximately 1dB/mm.

$$\frac{d^2V}{dx^2}=\frac{I}{4\cdot\pi}\left(\frac{3\sqrt{\rho_x\rho_y\rho_z}\;\rho_x^2x^2}{\left(\rho_x x^2+\rho_y y^2+\rho_z z^2\right)^{\frac{5}{2}}}-\frac{\sqrt{\rho_x\rho_y\rho_z}\;\rho_x}{\left(\rho_x x^2+\rho_y y^2+\rho_z z^2\right)^{\frac{3}{2}}}\right)$$

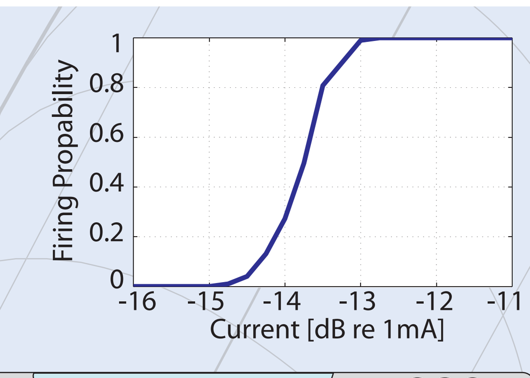
We analyze the quality of coding with the framework of automatic speech recognition (ASR). Features are extracted by temporal and spectral downsampling and fed to a multi-layer perceptron (MLP), which passes posterior probabiliteis to a Hidden Markov (HMM) recognition engine.

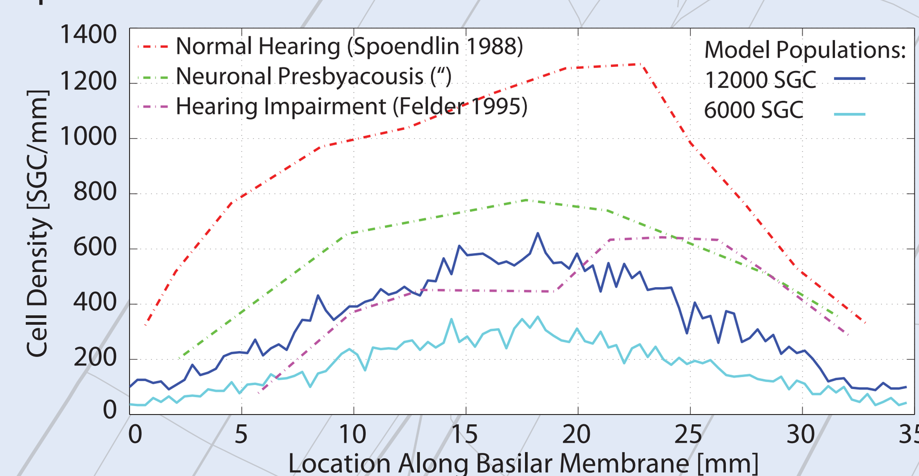
Individual model patient

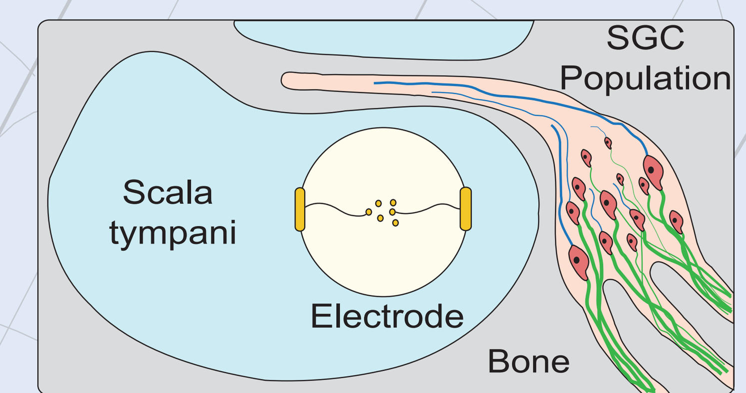
The dynamic range of the SGC population (DR_{pop}) comprises of the channel noise from an individual neuron (DR_n), the different axon diameters (DR_d) and the different distances between electrodes and cells in a channel (DR_s).

$$DR_{pop} = DR_s + DR_n + DR_d$$

The dynamic range due to channel noise is about 2.5 dB, variations of the diameter add up to 6 dB, the distance between electrode and SGC up to 12.5 dB.



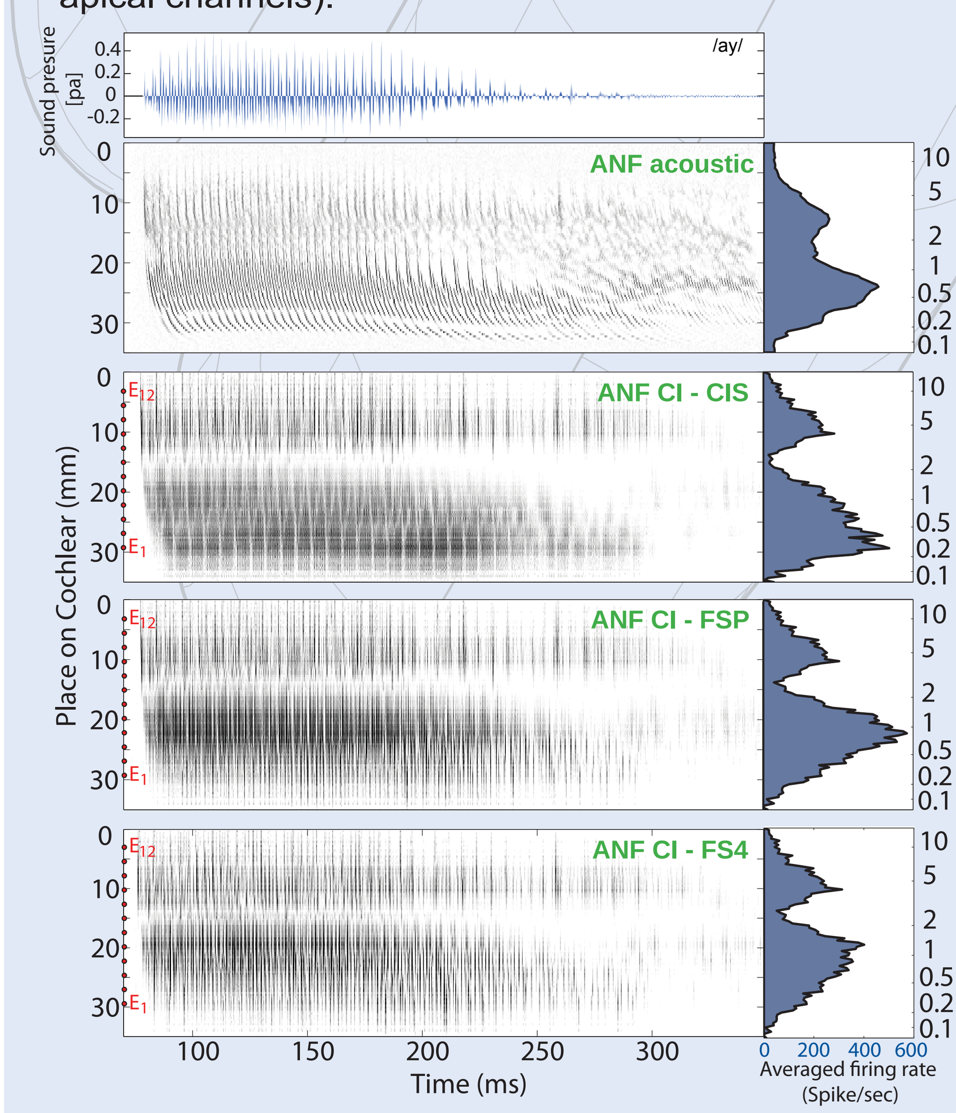


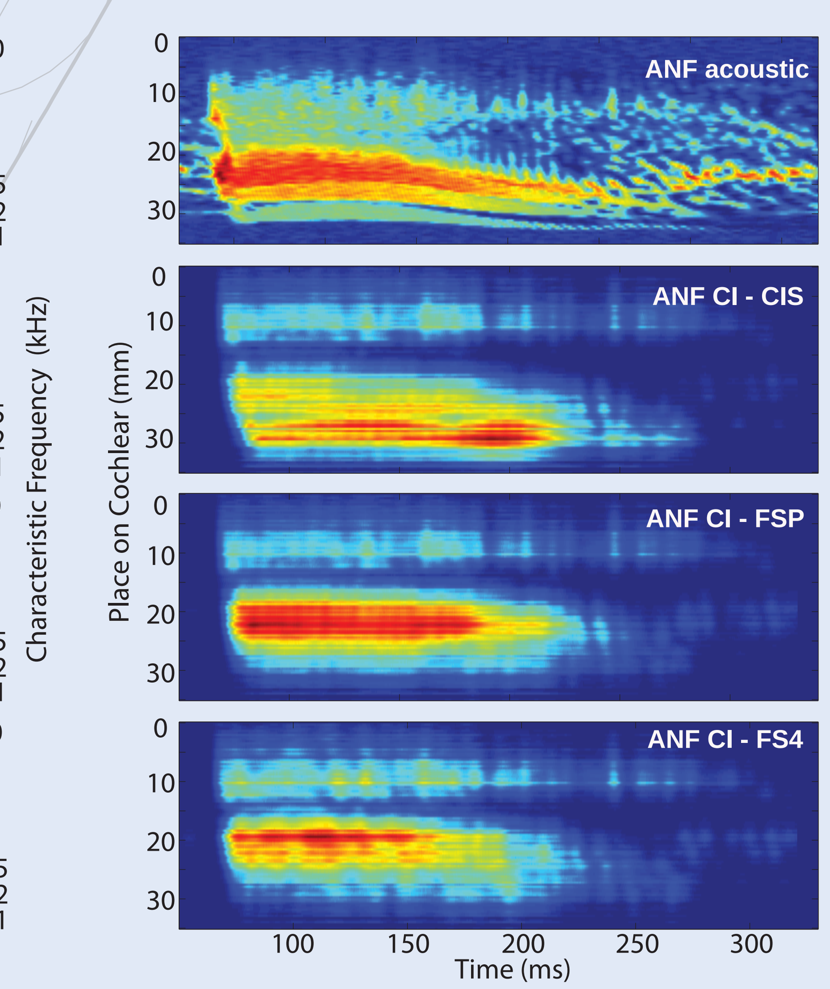


When we select appropriate SGC populations, we can model patients with dynamic ranges between 3 to 21 dB. With increasing dynamic range (requires a larger population), the computing time strong increases. The distribution of the cells along the basilar membrane (left Figure) and the distance of the cells to the electrode are generated with a random generator. For the speech recognition task we used the ISOLET database with noise added, which consists of 7800 spoken letters.

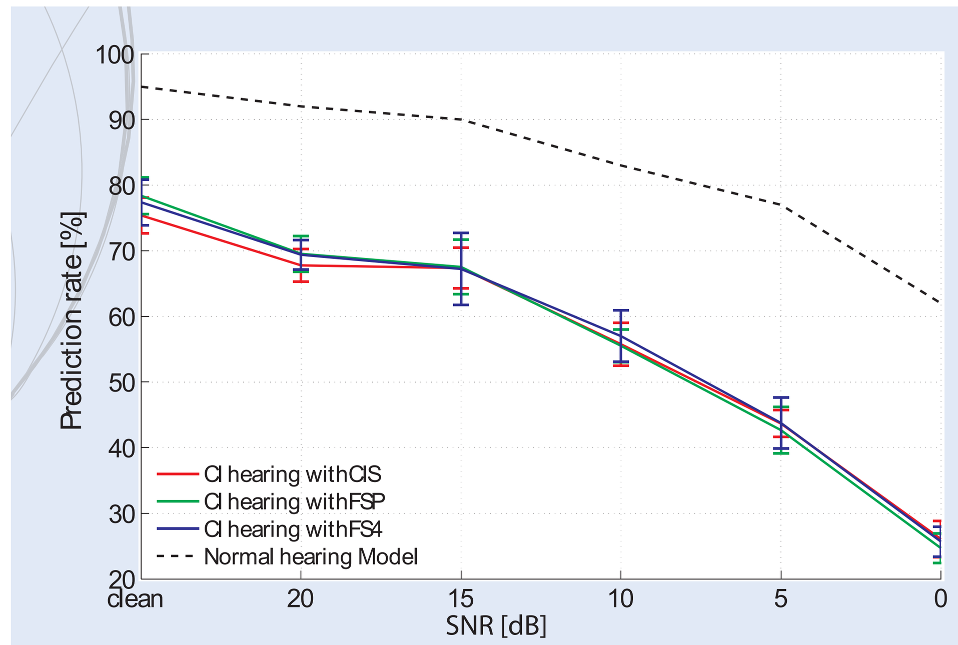
Comparison of Coding Strategies

The figure below shows response patterns of 6000 auditory nerve fibers in response to the spoken utterance /ay/ from the ISOLET database (female speaker fcmc0-A1-t, upper trace, 72.8 dB(A)). Upper panel: acoustic signal. Second panel: intact ear, 60 high-spontaneous-rate ANFs per frequency channel (right column: averaged firing rates). Third panel: response of a population of 6000 SGN to electric stimulation, CIS strategy. Fourth panel: same with FSP strategy and lower panel: FS4 strategy. The right column shows responses averaged with a 10 ms Hamming window, which emphasizes speech relevant spectral cues. Red circles on the left y-axis represent the positions of the stimulation electrodes in the cochlea. Electrical crosstalk (here: 1 dB/mm) limits the resolution of electrical stimulation. Both FS and FS4 strategies provide additional temporal information (phase-locking to fundamental frequency as well as to the first harmonics of the sound signal) in the most apical channels (3 respectively 4 most apical channels).

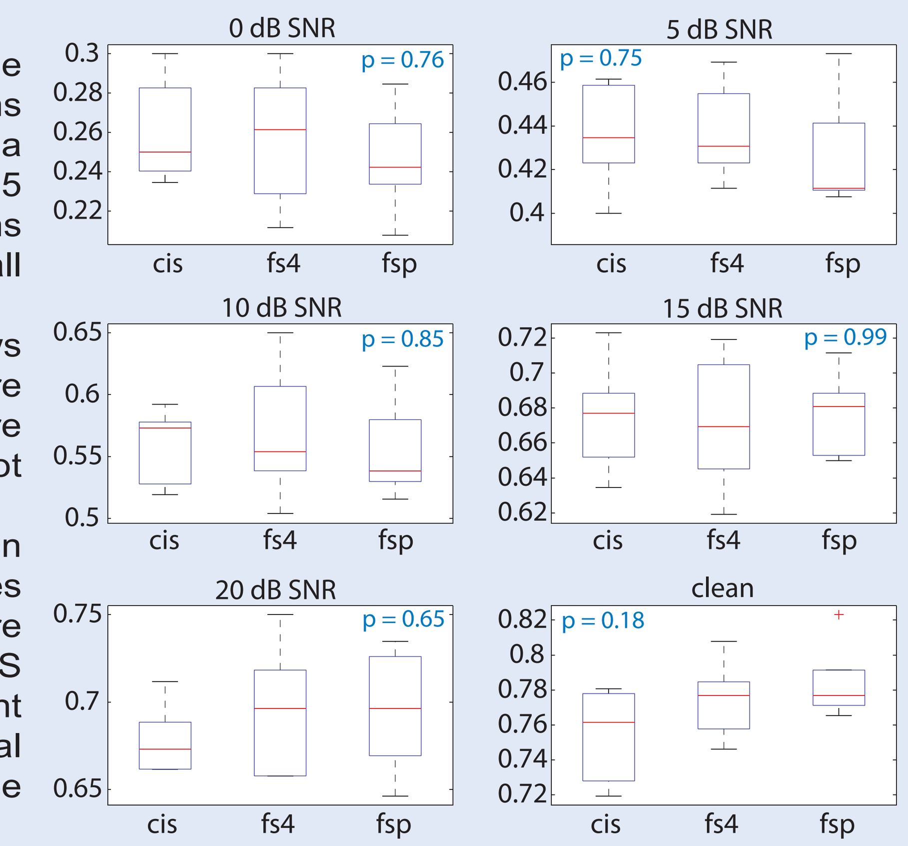




Speech recognition scores



The fine structure (FS) coding strategy is inspired by the FSP coding strategy in the MED-EL MAESTRO cochlear implant system and encodes temporal fine structure (TFS) on a few apical channels, while the remaining more basal channels carry CIS stimuli. Electrical stimulation with a CIS strategy degrades considerably in noise: recognition scores were between 75% and 80% in clean conditions (no noise added) and between 20% and 30% at 0 dB SNR.



To investigate if the differences between the strategies are significant, we used the 5 partitions of the ISOLET database and performed a one-way ANOVA for comparing the means over 5 trials of the prediction rates. The function returns the p value under the null hypothesis that all strategy's have the same mean's.

The comparison between the strategies shows not differernce in noise. For clean speech, there is a tendency that fine-structure strategies are slightly superior, however, this trend is not significant.

Please note that in this case, the recognition engine extracts spectral features only and does not take advantage of the temporal fine structure coded by the FS. We conclude that the FS strategies do not degrade the speech relevant spectral coding significantly. The potential improvement by the temporal coding has to be evaluated yet.

Conclusion

- Electric hearing is able to restore rate-place coding with a precision sufficient for speech perception in clean conditions, which however degrades severely in noise.
- Automatic speech recognition provides a quantitative tool to test the performance of coding strategies
- Strategies, which code temporal fine structure in low-CF channels seem to provide more natural stimulation of the auditory nerve.
- Model calculations of information transmission (data not shown) for electrical stimulation predict very high values, however, it is unclear if this information can be decoded.

References

- Holmberg, M., Gelbart, D. and Hemmert, W. (2007), Speech Communication, 49(12), pp. 917-932.
- Negm, M. and Bruce, I. (2008). EMBS 2008, pp 5539-5542.
- Shepherd, R. K. & Javel, E. (1997). Hear Res. , 108, pp. 112-144
- Wilson, B. S.; Finley, C. C.; Lawson, D. T.; Wolford, R. D.; Eddington, D. K. & Rabinowitz, W. M. (1991). Nature 352, pp. 236-238
- Kral, A.; Hartmann, R.; Mortazavi, D. & Klinke, R. (1998). Hearing Research, 121, pp.11-28
- Spoendlin, H. & Schrott, A. (1988). Acta Otolaryngol, 105, pp. 403-410
- Felder, E. & Schrott-Fischer, A. (1995). Hear Res., 91, pp.19-32

Acknowledgment

This work was funded by a grant from MED-EL Innsbruck and within the Munich Bernstein Center for Computational Neuroscience by the German Federal Ministry of Education and Research (reference number 01GQ1004B).