Preclinical Stress Test for Coding Strategies in Cochlear Implants

Michele Nicoletti[1], Christian Wirtz[2] and Werner Hemmert[1]

[1] *Technische Universität München, 85748 München, E-Mail: michele.nicoletti@tum.de*
[2] *MED-EL Deutschland GmbH, Starnberg, Germany*

Introduction

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 a toolbox was developed which calculates the excitation of the auditory nerve to sound stimuli. The toolbox includes detailed models of the CI speech processor, electrical field spread in the cochlea and the electrical excitation of spiral ganglion neurons. The toolbox allowed qualitative and quantitative comparison of neurograms elicited by different coding strategies with auditory nerve responses of the healthy inner ear. Quantitative evaluations are provided with two methods: i) automatic speech recognition to evaluate speech discrimination with a noisy speech database. ii) a binaural model which evaluates interaural time differences and quantifies the temporal precision of cochlear implant coding strategies. The major advantage of this approach is that we are able to evaluate both spectral and temporal aspects of novel coding strategies before we conduct extensive clinical studies.

Material and Methods

Figure 1 shows the schematic out-line of the model for electrical (inner ear with CI) stimulation. Sound signals are processed by the CI speech processor and biphasic electrical pulse trains are delivered to the electrode implanted in the inner ear. The electrode was modelled as an array of 12 current point sources at a distance of 0.5 mm from the spiral ganglion cells (SGC), which comprise the auditory nerve.

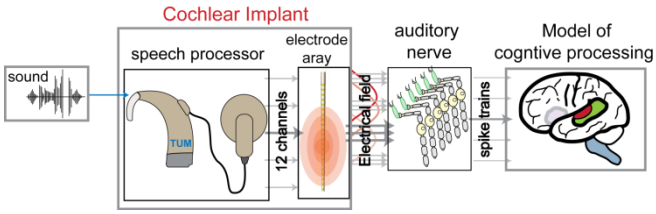

Figure 1: Schematic figure of our model for electrical stimulation of the auditory nerve.

The coupling between electrodes and excitation of the SGCs is described by the activation function (second derivative of the extracellular potential (V) along the neuron's path x) [1]. Cannel cross-talk is modelled by a field equation. Large channel crosstalk results from a wide current spread along scala tympani. It was modelled with symmetrical spread of excitation function with a slope of approximately 1dB/mm [2-4].

For the implanted ear, SGCs are modelled with single- or multi-compartment models with Hodgkin-Huxley like ion channels, which are also found in cochlear nucleus neurons (HPAC, K_{HT}, K_{LT}). Their large time constants might be responsible to explain adaptation to electrical stimulation (Negm 2008 [5]). Conductances and time-constants were corrected to a body temperature of 37°. Differential equations were solved in the time domain with the Crank–Nicolson method and an exponential Euler rule.

The dynamic range of the SGC population is due to the channel noise of an individual neuron, different axon diameters and the different distances between electrodes and neurons in a channel. When we select appropriate SGC populations, we can model patients with dynamic ranges between 3 to 21 dB. With increasing dynamic range, which require larger SGC populations, computation times increase. Distance of the cells to the electrode and SGCs along the basilar membrane are randomly distributed (Figure 2) [6, 7].

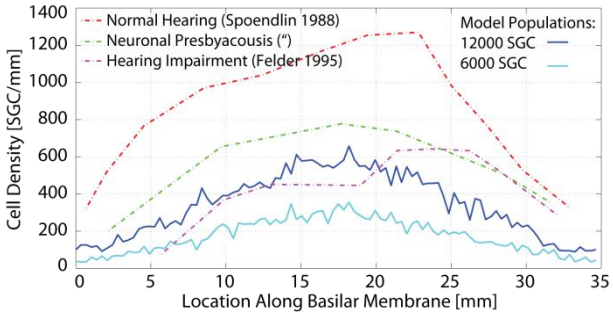

Figure 2: Distribution of SGC along the cochlea

The quality of speech coding was analysed with an automatic speech recognition system (ASR). Features were extracted by temporal and spectral down-sampling and fed to a multi-layer perceptron (MLP), which passes posterior probabilities to a Hidden Markov (HMM) recognition engine.

Binaural evaluations were conducted with a Lindemann model [8, 9], which was extended to work with action potentials. The simulated listening setup consisted of a loudspeaker, which was circling around the listener's head (once per second). The distance between the ears was 15 cm. An emitted wavefront reached both ears at different times and invoked interaural time differences (ITD). Furthermore, the simulation assumed linear air-attenuation and also generated interaural level differences (ILD). The binaural Lindemann model extends the correlation delay line of the Jeffress model by inhibitory elements, thus adding ILD-sensitivity to the model. A positive Lindemann correlation time-delay indicates a sound source originating from right hand side and vice versa. A sound source with 1 Hz circling frequency will lead to a deviation of max. 0.441 ms.

Results

Comparison of neural representations

Figure 3 shows the response patterns of 6000 auditory nerve fibres in response to the spoken utterance /ay/ from the ISOLET database (female speaker fcmc0-A1-t, upper trace, 72.8 dB(A)). The upper panel shows the acoustic input signal. The second panel shows the response of an intact ear model with 60 high-spontaneous-rate auditory nerve fibres per frequency channel (right column: averaged firing rates over whole sequence). The third panel shows the response of a population of 6000 SGCs to electric stimulation with a continuous interleaved sampling (CIS) strategy. In the fourth panel, the SGC response to a fine-structure coding strategy (FSP) strategy is shown. Responses were 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.

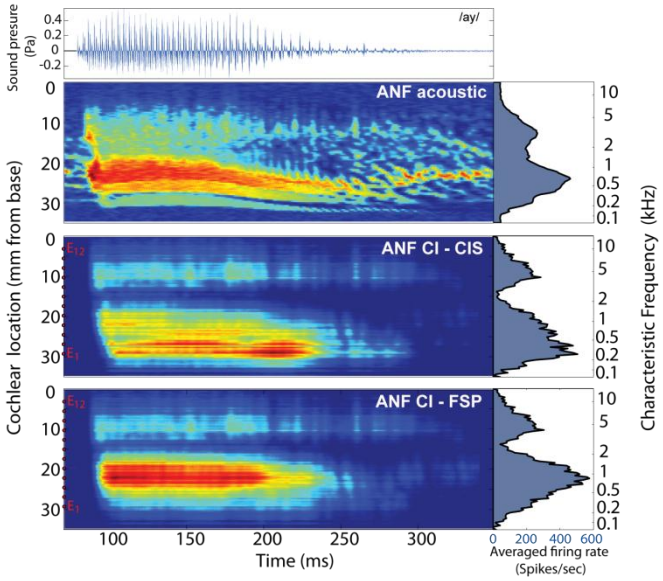

Figure 3: Response patterns of auditory nerve fibres in response to the spoken utterance /ay/ from the ISOLET database (female speaker fcmc0-A1-t, upper trace, 72.8 dB(A)). Upper panel: acoustic signal. Second panel: intact ear, 60 high-spontaneous-rate auditory nerve fibres per frequency channel averaged with a 10 ms Hamming window (left) and averaged firing rate (right). Third panel: response of a population of 6000 SGN to electric stimulation, CIS strategy. Fourth panel: same with FSP strategy.

Figure 4 illustrates responses with high temporal resolution. In contrast to the CIS strategy, the FSP strategy provides additional temporal information (phase-locking to fundamental frequency as well as to the first harmonics of the sound signal) in the most apical channels.

Comparison of speech recognition scores

To investigate if speech coding between the strategies was different, we used an ASR system and evaluated 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 strategies have the same means. The comparison between the CIS, FSP and FS4 strategies showed no significant difference in noise.

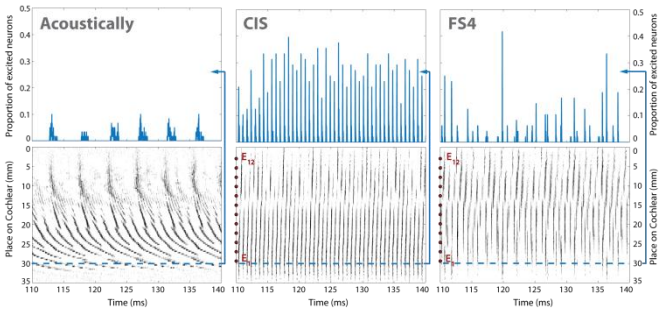

Figure 4: Response pattern at high temporal resolution (0.4 ms time bins). (Left) intact ear, (middle) implanted ear with CIS, (right) with FS4 coding strategy. Upper traces show the response probabilities of a population of neurons at the position of the most apical electrode in a time bin.

For clean speech, there was a tendency that fine-structure strategies were slightly superior; however, this trend was not significant. Please note that in this case, the recognition engine extracted spectral features only and did not take advantage of the temporal fine structure. We conclude that the FS strategies do not degrade the speech relevant spectral coding. The potential improvement by temporal coding has to be evaluated yet.

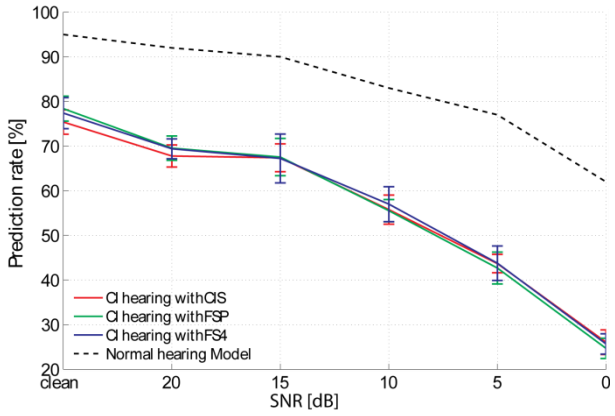

Figure 5: Recognition rates from features derived from auditory nerve spike trains from a model of normal hearing and for electrical hearing with different coding strategies.

Comparison of binaural cues

For a normal hearing model the Lindemann correlation easily tracked the sound source (compare Figure 6). The temporal resolution of a CIS strategy (stimulation rate: 1.25 kHz per channel) was not sufficient to code ITDs. When we tested MED-ELs FS4 strategy, which conveys the temporal fine-structure in its four lowest frequency channels, we could observe that the tracking of a moving source is possible with a precision of 0.2 ms. When we analyzed the electrically induced spike trains, we found a much weaker Lindemann correlation (data not shown here) due to the effects of channel crosstalk.

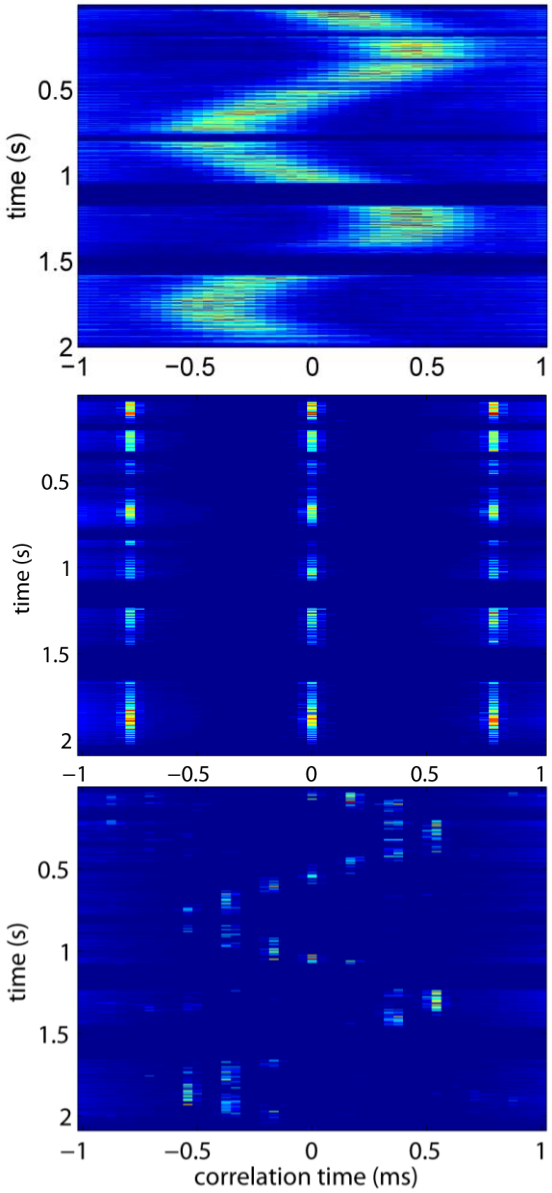

Figure 6 Lindemann correlation derived from spike trains generated by a normal hearing inner ear model (upper panel) from the stimulation signal of electrode 2 for the MED-EL coding strategy CIS (middle panel) and FS4 (lower panel). We used a virtual speaker articulating a sentence while circling around the head (1 rotation/s).

Discussion & Conclusion

Automatic speech recognition provides a quantitative tool to test the performance of coding strategies. Electric hearing is able to restore rate-place coding with a high degree of precision, sufficient for speech perception in clean conditions. However, spectral coding degrades severely in noise. Strategies which code the temporal fine structure in low-CF channels provide temporal information which provides more natural stimulation of the auditory nerve. Model calculations also showed that the temporal resolution of the FS4 strategy is already high enough to code sound localization.

In summing up, the model toolbox proposed in this chapter generates spiking auditory nerve responses and provides quantitative evaluation methods to evaluate speech coding and sound localization. The model delivers quantitative data and therefore enables comparisons between different cochlear implant coding strategies. As the model of electric excitation of the auditory nerve also includes effects such as channel crosstalk, neuronal adaptation and mismatch of electrode positions between left and right ear, its predictive power goes far beyond pure analysis of the output patterns of implants, which is how contemporary coding strategies were developed. The ability of the sound-localization model to process neuronal spike trains makes the model very versatile. It is possible to evaluate not only responses of the intact ear but also of the deaf inner ear provided with a cochlear implant. Nevertheless, up to now, this model only extends up to the level of the auditory nerve and can therefore not answer the question whether ITDs can still be processed by higher levels of the auditory pathway. Where this final evaluation always has to be done with CI users, this framework provides important answers to the question of how well monaural and binaural cues are coded at the first neuronal level. Given the long development cycles including design, fabrication, approval, implantation, and finally extensive measurements in a large group of CI users to yield statistically significant results, the benefit of this approach cannot be overestimated.

References

1. Rattay, F., *Analysis of models for external stimulation of axons.* IEEE Trans Biomed Eng, 1986. **33**(10): p. 974-977.
2. Kral, A., et al., *Spatial resolution of cochlear implants: the electrical field and excitation of auditory afferents.* Hearing Research, 1998. **121**(1-2): p. 11-28.
3. Nelson, D.A., G.S. Donaldson, and H. Kreft, *Forward-masked spatial tuning curves in cochlear implant users.* J Acoust Soc Am, 2008. **123**(3): p. 1522-1543.
4. O'Leary, S.J., R.C. Black, and G.M. Clark, *Current distributions in the cat cochlea: a modelling and electrophysiological study.* Hear Res, 1985. **18**(3): p. 273-281.
5. Negm, M.H. and I.C. Bruce, *Effects of I(h) and I(KLT) on the response of the auditory nerve to electrical stimulation in a stochastic Hodgkin-Huxley model.* Conf Proc IEEE Eng Med Biol Soc, 2008. **2008**: p. 5539-5542.
6. Spoendlin, H. and A. Schrott, *The spiral ganglion and the innervation of the human organ of Corti.* Acta Otolaryngol, 1988. **105**(5-6): p. 403-410.
7. Felder, E. and A. Schrott-Fischer, *Quantitative evaluation of myelinated nerve fibres and hair cells in cochleae of humans with age-related high-tone hearing loss.* Hear Res, 1995. **91**(1-2): p. 19-32.
8. Lindemann, W., *Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals.* J. Acoust. Soc. Am., 1986. **80**(6): p. 1608-1622.
9. Lindemann, W., *Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front.* J. Acoust. Soc. Am., 1986. **80**(6): p. 1623-1630.

Supported within the Munich Bernstein Center for Computational Neuroscience by the German Federal Ministry of Education and Research (01GQ1004B and 01GQ1004D) and MED-EL Innsbruck.