INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Introducing temporal rate coding for speech in cochlear implants: A microscopic evaluation in humans and models

Anja Eichenauer[1], *Mathias Dietz*[1,2], *Bernd T. Meyer*[3], *Tim Jürgens*[1]

[1] Medizinische Physik and Cluster of Excellence Hearing4all, University of Oldenburg, Germany
[2] National Centre for Audiology, The University of Western Ontario, London, Ontario, Canada
[3] Center for Speech and Language Processing, Johns Hopkins University, Baltimore, MD, USA
`anja.eichenauer@gmail.com`

Abstract

Standard cochlea implant (CI) speech coding strategies transmit formant information only via the place of the stimulated electrode. In acoustic hearing, however, formant frequencies are additionally coded via the temporal rate of auditory nerve firing. This study presents a novel CI coding strategy ("Formant Locking (FL)-strategy") that varies stimulation rates in relation to extracted fundamental and formant frequencies. Simulated auditory nerve activity resulting from stimulation with the FL-strategy shows that the FL-strategy triggers spike rates that are related to the formant frequencies similar as in normal hearing, and greatly different than in a standard CI strategy. Vowel recognition in seven CI users via direct stimulation of their electrode array shows that the FL-strategy results in significantly increased scores of the vowels /u/ and /i/ compared to a standard CI strategy. However, at the same time, a decrease in scores for /o/ and /e/ occurred. A microscopic speech intelligibility model involving an automatic speech recognizer reveals good agreement between modeled and predicted confusion matrices for the FL-strategy. This suggests that microscopic models can be used to test CI strategies in the development phase, and gives indications which cues might be used by the listeners for speech recognition.

Index Terms: vowel recognition, auditory model, automatic speech recognition, confusion matrix, cochlear implants

1. Introduction

Accurate perception of speech frequency content such as fundamental frequency (F0) and formant frequencies (F1, F2) is one important aspect for speech intelligibility. In the acoustically stimulated auditory system frequencies are precisely coded via two mechanisms: place code and temporal rate code. Place code is determined by the maximum excitation along the cochlear partition, its best frequency (BF). Temporal rate code is determined by the synchronization of auditory nerve (AN) cells to the phase of the stimulus, especially for low frequencies. Temporal AN discharge patterns (neurograms) were found to phase-lock to formant frequencies when vowels are presented acoustically [1] and scalp-recorded frequency following responses (FFR) from humans show preservation of this neural code up until the brainstem [2]. Place code and temporal rate co-vary with the stimulus frequencies in acoustic hearing.

In cochlear implant (CI) electrode arrays, place code and temporal rate code can be separated via the stimulation electrode and stimulation rate, respectively. Most state-of-the-art CI coding strategies use constant stimulation rates. Thus, information is transmitted via place coding of temporal amplitude modulations. Due to electric field spatial spread, the desired tonotopic stimulation pattern is less precise and spectral resolution is restricted. As a result, confusions between vowels with similar first or second formant (F1 and F2) frequently occur [3].

Experimental coding strategies for example apply additional amplitude modulation with respect to F0 on pulse trains with constant stimulation rate [4, 5]. Others extract information from auditory models to analyze frequency dependent information per electrode and relate those to variable stimulation rates [6, 7].

In this study, a new CI coding strategy was developed that aims at reproducing AN discharge patterns for vowels due to acoustic stimulation by varying the stimulation rates with respect to F0 and formant frequencies. It is hypothesized that additional temporal rate coding can improve vowel perception in CI users. An auditory model [8] was utilized to simulate and evaluate neurograms due to CI stimulation. Vowel perception was tested with actual CI listeners and a microscopic model of CI user's speech perception was applied. Predicted confusion matrices were compared to measured results.

2. CI coding strategy

2.1. Signal Processing

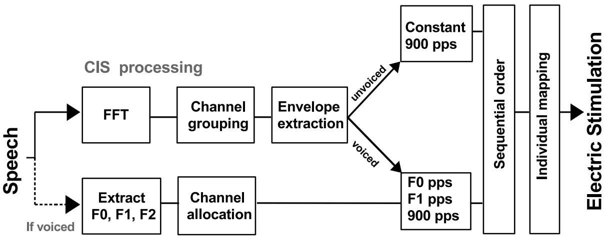


Figure 1: Flow chart of the FL-coding-strategy for CIs

The novel "formant locking" (FL) strategy is based on the widely used Continuous Interleaved Sampling (CIS) strategy [9]. A flowchart of the strategy's signal processing is shown in Fig. 1. The incoming speech sound is processed with a 12 channel filter bank (similar to a MED-EL CI system) using FFT, followed by an extraction of the Hilbert envelope [10]. In addition, the broadband input signal is analyzed for voiced and unvoiced parts, performed by a combination of voice activity detection and zero-crossing algorithm. Signal parts that are 10 dB below the average root-mean-square (RMS) of the signal, or whose average time distance of zero crossings is less than 5 ms are declared as unvoiced. The FL-strategy only applies rate coding

`http://dx.doi.org/10.21437/Interspeech.2016-267`

to voiced segments, unvoiced segments are stimulated using a constant stimulation rate of 900 pps (pulses per second, as used in CIS). For voiced signal parts every 32 ms F0 (via autocorrelation), F1 and F2 (both via LPC analysis [11]) are extracted and the channels to be stimulated using these frequencies are determined: Channel 1 (located apical) up to the channel closest to F1 stimulate with a rate corresponding to F0. Higher channel numbers up to the channel closest to F2 stimulate with a rate corresponding to F1. The remaining channels constantly stimulate with 900 pps. This channel allocation introduces temporal rate information and boundaries between coherently stimulating electrodes that both contain formant frequency information, inspired by physiological studies of [1, 2]. Pulses are applied sequentially in order to avoid electrical field overlap. If two pulses should occur simultaneously, the pulse in the more apical channel is temporally shifted. Finally, the extracted envelope magnitudes are mapped in the individual dynamic range of each patient using a logarithmic function [12].

2.2. Peripheral modeling

2.2.1. Setup

For CIS and the FL-strategy a neurogram was simulated using a model of electric stimulation of the AN [13, 8] and compared against a simulated neurogram generated with the Meddis' model of acoustic stimulation of the AN [14] which was published in [15]. The electric model contains the respective CI coding strategy, a function to replicate spatial spread of the electric field and integrate-and-fire AN models. The current spread is simulated by a double-sided, one-dimensional, exponentially decaying function defined by the spread constant λ = 0.5 mm. In agreement with [15], 500 stimulus repetitions of a synthesized /da/ stimulus were used and responses of 178 AN fibers equally distributed along the cochlear partition were summed for each neurogram. In the CI coding strategies the phase duration is set to 30 μs and inter phase gap to 2.1 μs. Electrode positions were chosen to match MED-EL's short electrode array ($Flex^{24}$) with electrode 1 located at 8.5 mm from the apex and a constant electrode spacing of 1.9 mm. BFs corresponding to these positions were calculated using Greenwood's formula [16]. For the simulation threshold (T)- and comfortable (C)-levels (defined as "loud, but not unpleasant") were taken from one random participant of this study. The broad band RMS of the speech signal was calibrated to be 20 dB above the level that is mapped to the T-level of each electrode, making optimal use of the dynamic range of speech.

2.2.2. Neurograms

Fig. 2 displays the first 45 ms of the simulated neurograms in response to the synthesized /da/ stimulus when the acoustic model [14] (left panel) or the electric model [8] with FL-strategy (middle panel) or CIS (right panel) were used. The acoustic neurogram (left panel) shows action potentials (APs) that occur periodically with respect to the frequencies of F1, F2 and F3 (marked in the graph). In the very bottom part the interval between two APs equals the reciprocal of F0. The neurogram of the FL-strategy (middle panel) shows APs that have periods which precisely match F0, F1 or 900 pps (F2-related stimulation rate). The synchrony of APs across areas of BFs is well visible. In comparison to the simulated acoustic neurogram (left panel), broader areas of fibers respond to the same stimulation rate, especially with respect to F0 and the pattern seems more coarse, which is partly due to the electrode positions (outlined as circles at the very right end of the figure). In the very top part of the middle neurogram, the AN shows periodicities again corresponding to F0, similar as in the acoustic neurogram. The neurogram of CIS (right panel) shows constant stimulation with a rate of 900 pps across all BFs. Also here, the AN response to the stimulation rate is precise but does not contain any information about formants. The modeling shows that formant frequencies can be temporally represented in the AN discharge pattern via rate coding.

3. Measurement methods

3.1. CI listeners

Seven postlingually deaf MED-EL CI implant users (4 female, 3 male) aged between 18 and 76 years (mean age 53 years) participated in this study. Participants were native German speakers and had more than 9 months experience with their CI (more details in Table 1). Ethical approval was granted by the Ethics commission of the Medizinische Hochschule Hannover.

Table 1: Participant details: Gender, age, status (CI = Cochlear Implant, NH = Normal Hearing, HA = Hearing Aid, EAS = Electro-Acoustic Stimulation), measured side, CI experience

ID	gender	age	left	right	measured	duration CI use
1	female	52	CI	NH	left	1 year
2	female	60	deaf	CI	right	12 years
3	male	73	CI	CI	right	8 years
4	male	76	CI	HA	left	1 year
5	male	65	HA	CI	right	1 year
6	female	29	EAS	CI	right	4 years
7	female	18	CI	NH	left	2 years

3.2. Apparatus

Tests were conducted at the University of Oldenburg. If participants had acoustic hearing on the contralateral ear, a soundproof booth was used. CI coding strategies were implemented in Matlab (The MathWorks) and transferred to the subjects' electrode array via the Research Interface Box (RIB II, University of Innsbruck).

3.3. Speech material

Speech tokens of a male speaker from the OLLO (OLdenburg LOgatom speech corpus) [17] were processed by the CI coding strategies and presented to the subjects. The OLLO was particularly designed to investigate phoneme recognition in humans and automatic speech recognizers (ASR). Three-phoneme logatomes (consonant-vowel-consonants) were used with the centered vowel serving as the target phoneme. Eight outer consonants (initial and final identical) were possible: /d/, /t/, /f/, /g/, /k/, /s/, /p/, /b/, which embedded five vowels: /a/, /e/, /i/, /o/, /u/. Thus, in total 40 different logatomes were used, such as /tat/ or /tet/. Recordings from OLLO were chosen to have *slow* speaking rate, *normal* vocal effort and speaking style *statement*.

3.4. Procedure

The test procedure started with a determination of the compliance limit (maximal possible current transmitted by each electrode). Subsequently, a pulse train at 100 pps and current just below the smallest compliance limit was used to measure an individual phase duration at this electrode. The individual phase duration is the smallest phase duration at which subjects can reach a C-level with this stimulus. If they could not reach their

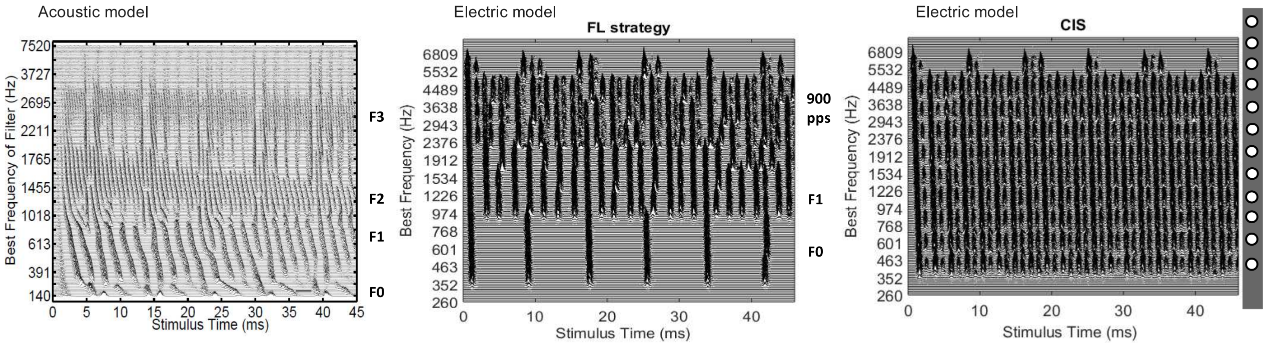


Figure 2: Instantaneous AN firing rates as a function of time for a range of best frequencies (Hz): Simulated neurograms of the first 45 ms of a synthesized /da/ speech stimulus generated by an acoustic model (left, adopted from [15]) and an electric model when the FL-strategy (middle) or CIS is used (right). On the very right a sketch of the electrode array is displayed.

C-level, the phase duration was set to 40 μs. T- and C-levels were measured for 250 pps, 500 pps and 900 pps using this individual phase duration.

Prior to the OLLO test participants were asked to adjust the volume control for both algorithms until the speech (using one sentence from [18]) was comfortably loud. These individual volume settings were then also used for the OLLO. Five response alternatives were displayed on a touch screen only differing in the (middle) vowel and the participant was asked to press the response alternative that was understood. Feedback was given after each response. Three sets of 40 logatomes each were tested per strategy in a row. The order of logatomes was randomized.

4. Microscopic modeling

4.1. Central model stage

For the microscopic model of CI user's speech intelligibility neurograms of the same OLLO stimuli were extracted using the same model as described in Section 2.2.2 but now with 2000 AN fibers equally distributed along the cochlear partition. These neurograms were further processed using the central auditory processing stage of [8]. This stage includes spatial grouping (36 groups, each with a width of 0.95 mm) and temporal integration, as well as multiplication with internal Gaussian noise. The internal noise standard deviation (σ_{CN}) was chosen to be 0.05. The internal representation, i.e., the output of the central stage, is used as input for the ASR backend.

4.2. ASR backend

The ASR experiment for phoneme recognition was performed with a Hidden Markov Model (HMM) in HTK [19]. Three states plus one additional start and end state were used to model monophones. The 36-dimensional internal representation for either CIS or the FL-strategy served as feature. 680 logatomes produced by one speaker were used for training, 40 different logatomes (same speaker) were used for the test procedure. Similar to the test with the CI listeners, only confusions between central phonemes were analyzed. For both strategies the modeling was performed 7 times and the results were averaged.

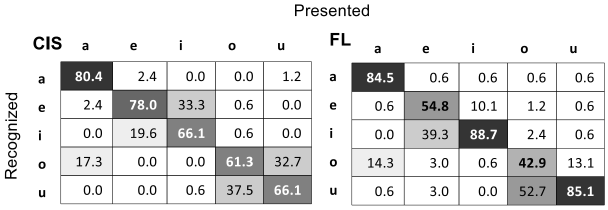
4.3. Evaluation of model outcome

[20] presented a standardized procedure for microscopic speech intelligibility model evaluation. This procedure uses log-likelihoods to assess the goodness of the model's ability to predict phoneme recognition. In this study, the model is evaluated in terms of confusion frequency prediction (i.e., the general probability that a phoneme is confused) and full confusion prediction (i.e., with which phoneme it is confused). The evaluated log-likelihood is compared against two references: the oracle (perfect prediction) and the random prediction (20 % for each response alternative). The closer the likelihood of the evaluated model is to the oracle result, the better the prediction.

5. Results and comparison

The recognition score of CIS averaged across all seven participants and all vowels is 70.4 %, individual results range between 52 % and 78 %. The average recognition score of the FL-strategy is 71.2 % with individual results between 49.2 % and 94.2 %. Averaged results show no significant differences between algorithms according to overlap of 95 % confidence intervals [21]. When individual results of the FL-strategy are compared against CIS, participants six and seven achieved significant improvement with a benefit of 15 % and 18.4 %, respectively. The highest decrease in recognition score occurred for participant four (-18.3 %). Fig. 3 displays measured vowel confusion matrices averaged across all participants of CIS (left panel) and of the FL-strategy (right panel), each column being a presented vowel and each row a recognized vowel. Numbers indicate percentages. Prominent confusions occur mainly between /e/ and /i/ as well as between /o/ and /u/ for both strategies. Recognition of /i/ and /u/ is significantly higher using the FL-strategy than using CIS, recognition of /o/ and /u/ is significantly lower, and recognition of /a/ only changes slightly. For the FL-strategy, participants confused a presented /e/ more often with /i/ and a presented /o/ more often with /u/, i.e. tended to responses with lower F1.

The microscopic model predicted the recognition score (averaged across seven runs and all vowels) of CIS to be 55.7 % and of the FL-strategy to be 82.1 %. Averaged results are significantly different between the two algorithms according to [21]. The predicted score of CIS is lower and of the FL-strategy higher than the measured scores. Fig. 4 displays predicted vowel confusion matrices of CIS (left panel) and of the FL-strategy (right panel). Both confusion matrices exhibit most confusions between /e/ and /i/ as well as between /o/ and /u/. Especially the prediction with the FL-strategy shows a good replication of measured confusions. Similarly to the measurements, predicted scores of /i/ and /u/ are also higher than of /e/ and /o/.



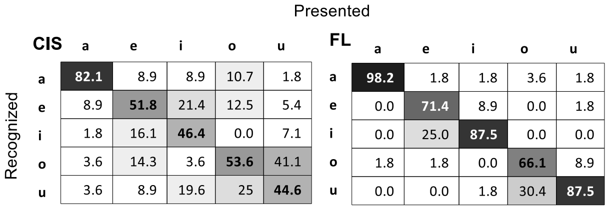
Presented

CIS	a	e	i	o	u
a	80.4	2.4	0.0	0.0	1.2
e	2.4	78.0	33.3	0.6	0.0
i	0.0	19.6	66.1	0.6	0.0
o	17.3	0.0	0.0	61.3	32.7
u	0.0	0.0	0.6	37.5	66.1

FL	a	e	i	o	u
a	84.5	0.6	0.6	0.6	0.6
e	0.6	**54.8**	10.1	1.2	0.6
i	0.0	39.3	88.7	2.4	0.6
o	14.3	3.0	0.6	**42.9**	13.1
u	0.6	3.0	0.0	52.7	85.1

Recognized

Figure 3: Measured confusion matrices averaged across all seven participants. Results indicate % correct.



Presented

CIS	a	e	i	o	u
a	82.1	8.9	8.9	10.7	1.8
e	8.9	**51.8**	21.4	12.5	5.4
i	1.8	16.1	**46.4**	0.0	7.1
o	3.6	14.3	3.6	**53.6**	41.1
u	3.6	8.9	19.6	25	**44.6**

FL	a	e	i	o	u
a	98.2	1.8	1.8	3.6	1.8
e	0.0	71.4	8.9	0.0	1.8
i	0.0	25.0	87.5	0.0	0.0
o	1.8	1.8	0.0	**66.1**	8.9
u	0.0	0.0	1.8	30.4	87.5

Recognized

Figure 4: Predicted confusion matrices averaged across 7 runs. Results indicate % correct.

Table 2 shows the results of the model evaluation according to [20]. The left two columns display the confusion frequency log-likelihoods: Modeled phoneme recognition with CIS is closer to random than to the oracle reference, whereas the results of the FL-strategy display log-likelihoods close to the oracle reference. The right two columns of Table 2 present the log-likelihoods for full confusions. Again, usage of the FL-strategy is evaluated to be more precise than of CIS, with log-likelihoods now closer to oracle than to random, indicating overall good prediction of occurring confusions.

Table 2: Microscopic model evaluation for CIS and the FL-strategy according to [20]. Results display log-likelihoods.

	Confusion frequency		**Full confusion**	
	CIS	**FL**	**CIS**	**FL**
Oracle model	-2.2	-2.4	-4.2	-7.3
Proposed model	-49.7	-6	-48.5	-22.6
Random model	-57.0	-48.8	-165.0	-170.6

6. Discussion

In this study a CI coding strategy was developed that introduces additional temporal coding via variable electric pulse rates. State of the art CI coding strategies, like CIS, only use constant pulse rates, i.e., frequency information is solely transmitted via place coding. Simulated AN responses show that the FL-strategy can trigger AN spikes with the periodicities of F0 and formant frequencies and therefore generates AN spiking patterns that are closely related to acoustically triggered neurograms. Neurograms of CIS and the FL-strategy differ greatly as the FL-strategy allocates specific stimulation rates to selected electrodes. The measured vowel recognition scores of this study show that for both coding strategies confusions mainly occur between vowels with similar F1-F2 combinations, i.e., /e/ and /i/ as well as /o/ and /u/, in agreement with [3]. /a/ with its unique position in the vowel space has a high recognition score. The allocation of different coherently stimulating groups of electrodes provides an additional cue that highlights formant frequencies, which contain the characteristic F1-F2 combination of each vowel. Furthermore, the temporal stimulation rates in the selected channels reflect F0 or F1, indicating another cue. With the FL-strategy, averaged measured vowel recognition increases by 0.8 % compared to CIS. The improvement is small, but nevertheless promising, which may be due to acute testing without longer periods of accommodation. Future research should investigate if CI listeners can use these new temporal rate cues after longer familiarization periods. This may be possible because listeners have the ability to adapt phonetic cue labeling by weighting their importance differently [22]. The two participants with the highest recognition benefit were both aged under 30 years, indicating that younger CI users are faster in adopting new stimulation patterns.

The predicted results highlight the potential of the FL-strategy with an increase in average recognition score from 55.7 % (CIS) to 82.1 % (FL-Strategy). Using the evaluation method according to [20], the results of the microscopic speech intelligibility model are in very good agreement with the measured results for the FL-strategy, particularly when the clustering of /e/-/i/ and of /u/-/o/ are considered. The prediction is less precise when CIS is used. With CIS, the internal representations lack the additional rate-cue that the model can use to improve speech recognition. However, the participants make better use of place coding with CIS than the model. This can be seen in the strongly clustered confusions that were determined in the measurements for CIS. For the FL-strategy /e/ was often confused as /i/, and /o/ was more often identified as /u/ than for CIS. This was determined for both, measured and predicted results. Vowels /i/ and /u/ have a lower F1 compared to /e/ and /o/. The reason why those recognition shifts occurred is currently unclear. Since the ASR backend with few temporal states does not explicitly extract AN timing, it is likely that the boundaries between coherently stimulating electrodes are mainly responsible for the effect.

The auditory model applied in this study does not replicate data of individual CI users, but only uses one set of physiological parameters. Modeled variation only comes from the internal noise that is re-generated for every run. More model predictions with other sets of physiological parameters could be done in order to investigate which speech cues are accessible in each strategy in relation to these parameters. These could also involve modeling differences to perceive temporal rate changes, which has shown to vary considerably across individuals [23, 24, 25].

7. Conclusions

This study presents a novel CI coding strategy (”FL-strategy”) that varies stimulation rates in relation to spectral speech characteristics. The simulation of neurograms was a vital part in the development process of the FL-coding-strategy. Simulations show that the novel temporal rate information is present in the AN responses, which brings them closer to acoustically evoked AN responses than elicited by CIS. Measured confusion matrices display that mainly vowels with similar positions in the vowel space are confused. The model can exploit rate channel allocations that highlight formant frequencies. Future research should include further tests with longer accommodation times for participants. Measurements of words and sentences but also formant frequency difference limens are of great interest.

8. Acknowledgements

This work was supported by the DFG Cluster of Excellence EXC 1077/1 ”Hearing4all”. Special thanks to Stefan Strahl (MED-EL) for support on the RIB II.

9. References

[1] H. E. Secker-Walker and C. L. Searle, "Time-domain analysis of auditory-nerve-fiber firing rates," *The Journal of the Acoustical Society of America*, vol. 88, no. 3, pp. 1427–1436, 1990.

[2] A. Krishnan, "Human frequency-following responses: representation of steady-state synthetic vowels," *Hearing research*, vol. 166, no. 1, pp. 192–201, 2002.

[3] B. Laback, W. A. Deutsch, and W.-D. Baumgartner, "Coding of vowellike signals in cochlear implant listeners," *The Journal of the Acoustical Society of America*, vol. 116, no. 2, pp. 1208–1223, 2004.

[4] T. Green, A. Faulkner, and S. Rosen, "Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants," *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2298–2310, 2004.

[5] A. E. Vandali and R. J. van Hoesel, "Development of a temporal fundamental frequency coding strategy for cochlear implants," *The Journal of the Acoustical Society of America*, vol. 129, no. 6, pp. 4023–4036, 2011.

[6] D. B. Grayden, A. N. Burkitt, O. P. Kenny, J. C. Clarey, A. G. Paolini, and G. M. Clark, "A cochlear implant speech processing strategy based on an auditory model," in *Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004. Proceedings of the 2004*. IEEE, 2004, pp. 491–496.

[7] T. Harczos, A. Chilian, and P. Husar, "Making use of auditory models for better mimicking of normal hearing processes with cochlear implants: the sam coding strategy," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 7, no. 4, pp. 414–425, 2013.

[8] S. Fredelake and V. Hohmann, "Factors affecting predicted speech intelligibility with cochlear implants in an auditory model for electrical stimulation," *Hearing research*, vol. 287, no. 1, pp. 76–90, 2012.

[9] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, no. 6332, pp. 236–238, 1991.

[10] L. Rabiner and B. Gold, *Theory and Applications of Digital Signal processing*. Prentice-Hall, Eaglewood Cliffs, NJ, 1975.

[11] R. C. Snell and F. Milinazzo, "Formant location from lpc analysis data," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 2, pp. 129–134, 1993.

[12] W. Nogueira, A. Büchner, T. Lenarz, and B. Edler, "A psychoacoustic "nofm"-type speech coding strategy for cochlear implants," *EURASIP Journal on Applied Signal Processing 18*, pp. 3044–3059, 2005.

[13] V. Hamacher, *Signalverarbeitungsmodelle des elektrisch stimulierten Gehörs*. Ph.D Thesis, RWTH Aachen, Wissenschaftsverlag Mainz in Aachen, 2004.

[14] R. Meddis, "Auditory-nerve first-spike latency and auditory absolute threshold: a computer model," *The Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 406–417, 2006.

[15] G. Brown, T. Jürgens, R. Meddis, M. Robertson, and N. Clark, "The representation of speech in a nonlinear auditory model: time-domain analysis of simulated auditory-nerve firing patterns," *Proceedings Interspeech, Florence, Italy*, pp. 2464–2468, 2012.

[16] D. D. Greenwood, "A cochlear frequency-position function for several species 29 years later," *The Journal of the Acoustical Society of America*, vol. 87, no. 6, pp. 2592–2605, 1990.

[17] T. Wesker, B. T. Meyer, K. Wagener, J. Anemüller, A. Mertins, and B. Kollmeier, "Oldenburg logatome speech corpus (ollo) for speech recognition experiments with humans and machines." in *Interspeech*, 2005, pp. 1273–1276.

[18] K. Wagener, V. Kühnel, and B. Kollmeier, "Entwicklung und evaluation eines satztests für die deutsche sprache teil 1: Design des oldenburger satztests." *Zeitschrift für Audiologie 38*, pp. 4–15, 1999.

[19] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey *et al.*, *The HTK book*. Entropic Cambridge Research Laboratory Cambridge, 1997, vol. 2.

[20] R. Marxer, M. Cooke, and J. Barker, "A framework for the evaluation of microscopic intelligibility models," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[21] G. Cumming and S. Finch, "Inference by eye: confidence intervals and how to read pictures of data." *American Psychologist*, vol. 60, no. 2, p. 170, 2005.

[22] A. L. Francis, K. Baldwin, and H. C. Nusbaum, "Effects of training on attention to acoustic cues," *Perception & Psychophysics*, vol. 62, no. 8, pp. 1668–1680, 2000.

[23] F.-G. Zeng, "Temporal pitch in electric hearing," *Hearing research*, vol. 174, no. 1, pp. 101–106, 2002.

[24] U. Baumann and A. Nobbe, "Pulse rate discrimination with deeply inserted electrode arrays," *Hearing research*, vol. 196, no. 1, pp. 49–57, 2004.

[25] Y.-Y. Kong, J. M. Deeks, P. R. Axon, and R. P. Carlyon, "Limits of temporal pitch in cochlear implants," *The Journal of the Acoustical Society of America*, vol. 125, no. 3, pp. 1649–1657, 2009.