of about 10 dB between 1 and 3 kHz. Accordingly, the change of the own voice due to the hearing aid is less than the change of other voices, e.g. a voice by a significant other. Therefore, own voice problem cannot solely be explained by great alteration of voice characteristics due to change of spectral content but small changes can, for some subjects, result in severe disturbance of own voice perception.

REFERENCES

Carle, R., Laugesen, S., and Nielsen, C. (**2002**). “Observation on the relation among occlusion effect, compliance, and vent size” J. Am. Acad. Audiol. **13**, 25-37.

Dillon, H. (2001). *Hearing Aids*. New York, Thieme.

Gnewikow, D., and Moss, M. (**2006**). “Hearing aid outcomes with open and closed canal fittings” Hear. J. **59**, 66, 68-70, 72.

Holube, I., Fredelake, S., Vlamming, M., and Kollmeier, B. (**2010**). ”Development and analysis of an international speech test signal (ISTS)” Int. J. Audiol. **49**, 891-903.

Kiessling, J., Brenner, B., Thunberg Jespersen, C., Groth, J., and Dyrlund Jensen, O. (**2005**) ”Occlusion effect of earmolds with different venting systems” J. Am. Acad. Audiol. **16**, 237-249.

Kochkin, S. (**2010**). “MarkeTrack VIII: Consumer satisfaction with hearing aids is slowly increasing” Hear. J. **63**, 19-20, 22, 24, 26, 28, 30-32.

Kuk, F., Keenan, D., and Lau, C. (**2005**). “Vent configurations on subjective and objective occlusion effect” J. Am. Acad. Audiol. **16**, 747-762.

Kuk, F., Keenan, D., and Lau, C. (**2009**). “Comparison of vent effects between solid earmold and hollow earmold” J. Am. Acad. Audiol. **20**, 480-491.

Laugesen, S., Sorgaard Jensen, N., Maas, P., and Nielsen, C. (**2011**). ”Own voice qualities (OVQ) in hearing-aid users: There is more than just occlusion” Int. J. Audiol. **50**, 226-236.

Reinfeldt, S., Östli, P., Håkansson, B., and Stenfelt, S. (**2010**). “Hearing one’s own voice during phoneme vocalization – transmission by air and bone conduction” J. Acoust. Soc. Am. **128**, 751-762.

Stenfelt, S., and Reinfeldt, S. (**2007**). “A model of the occlusion effect with bone-conduction stimulation” Int. J. Audiol. **46**, 595-608.

Stenfelt, S., Wild, T., Hato, N., and Goode, R. (**2003**) “Factors contributing to bone conduction: The outer ear” J. Acoust. Soc. Am. **113**, 902-913.

Stenfelt, S. and Goode, R. L. (**2005**) “Bone conducted sound: Physiological and clinical aspects” Otol. Neurotol. **26**, 1245-1261.

Stenfelt, S. (**2006**) “Middle ear ossicles motion at hearing thresholds with air conduction and bone conduction stimulation” J. Acoust. Soc. Am. **119**, 2848-2858.

Taylor, B. (**2006**). “Real-world satisfaction and benefit with open canal fittings” Hear. J. **59**, 74, 76, 78, 80-82.

Von Békésy, G.(**1949**). “The structure of the middle ear and the hearing of one's own voice by bone conduction” J. Acoust. Soc. Am. **21**, 217-232.

Comparative evaluation of cochlear implant coding strategies via a model of the human auditory speech processing

TAMAS HARCZOS[1,2], STEFAN FREDELAKE[3], VOLKER HOHMANN[3], AND BIRGER KOLLMEIER[3]

[1] Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany

[2] Institute for Media Technology, Faculty of Electrical Engineering and Information Technology, Ilmenau University of Technology, Germany

[3] Medical Physics, Institute of Physics, Faculty of Mathematics and Science, Carl von Ossietzky University of Oldenburg, Germany

Traditional cochlear implant (CI) coding strategies present some information about the waveform or spectral features of the speech signal to the electrodes. However, neither of these approaches takes the cochlear traveling wave or the auditory nerve cell response into account, though these are given in acoustic hearing. Therefore, a new CI coding strategy based on an auditory model including the above mentioned properties of the healthy cochlea was evaluated and compared with an n-of-m-coding strategy, in which n electrodes out of m possible electrodes are stimulated in each stimulation cycle. The selection of the n electrodes is based on the n highest spectral maxima of the momentary signal. Simulated electrical output of both CI coding strategies served as input to a model of the electrically stimulated auditory system, which consisted of an auditory nerve cell population. The nerve cells generated delta pulses as action potentials in dependence on the spatial and temporal properties of the electric field produced by the electric stimuli. This model is used to predict CI user performance in terms of speech intelligibility and pitch discrimination for both coding strategies. Furthermore, an additional model of normal hearing is presented, the output of which is compared to the neural representation resulting from the modeled CI stimulation. We will show under which circumstances and to what extent an auditory model based coding strategy may outperform a traditional CI speech coding algorithm.

INTRODUCTION

Speech recognition performance in noise as well as the ability to discriminate pitch exhibits a high variation in cochlear implant (CI) users. The most probable origins of these differences are degenerative functional changes of the auditory nerve and dissimilarities between the used speech coding strategies.

To describe the quantitative relation between parameters of the auditory processing and speech perception with CIs, a model of the electrically stimulated auditory nerve (Hamacher, 2004), has been modified, and speech intelligibility is simulated for a

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

large range of model parameters with an approach similar to that proposed by Jürgens *et al.* (2010). The same model is used as a processing stage for evaluating cues for place and temporal pitch.

The model is used to predict CI user performance in terms of speech intelligibility and pitch discrimination for two strategies: a common n-of-m and SAM (see below). Furthermore, an additional model of normal hearing is employed, the output of which is compared to the neural representation resulting from the modeled CI stimulation.

METHODS

The SAM strategy

SAM (Stimulation based on Auditory Modeling) is a novel CI speech processing strategy (Harczos *et al.,* 2011), incorporating active cochlear filtering (basilar membrane and outer hair cells) along with the mechanoelectrical transduction of the inner hair cells. An overview of SAM is shown in Fig. 1. Through its functional design several psychoacoustic phenomena like compression, adaptation and realistic cochlear delays are accounted for inherently. The coder, unlike in common strategies, is not restricted by a pre-defined channel stimulation rate and it activates stimulating electrodes in a stochastic manner.

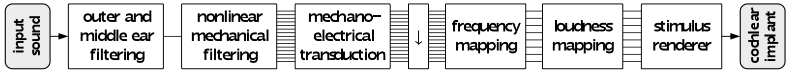

input sound
outer and middle ear filtering
nonlinear mechanical filtering
mechano-electrical transduction
↓
frequency mapping
loudness mapping
stimulus renderer
cochlear implant

Fig. 1: SAM system overview and signal path.

Model of the auditory processing

A model of the electrically stimulated auditory nerve by Hamacher (2004) has been adapted. Electric pulses, encoding a speech signal with a simulated CI, are multiplied with a spatial current function modeling the current spread inside the cochlea. A population of auditory nerve cells, which are based on the leaky-integrate-and-fire model, generates spikes, which are further processed to internal representations by modeling convergence and adaptation. An overview is presented in Fig. 2. The model includes parts of both the peripheral and central auditory processing, which are abbreviated as PAP and CAP, respectively, throughout this paper. The joint model of auditory processing (PAP+CAP) will be referred to as MAP.

In order to simulate different neural degeneration, the number of auditory nerve cells (N) is decreased, while the variable (λ) of the spatial current function is increased in a way that the total number of action potentials (APs) is kept constant for a given input current amplitude. For threshold current level, the total number of APs was arbitrarily set to approximately 30 and for most comfortable level to 300, respectively. The PAP model configurations used in this study are listed in Table 1.

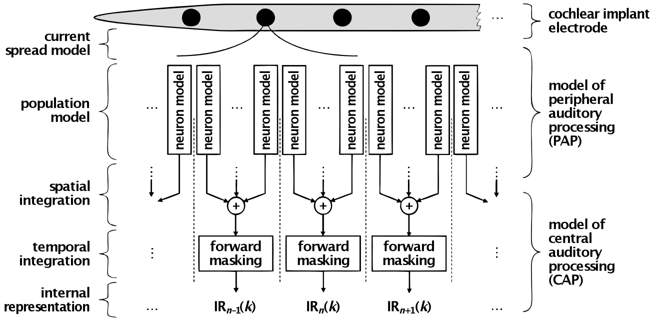
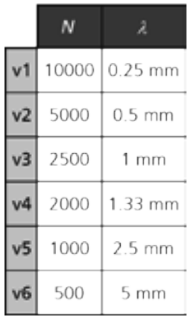


Fig. 2: Overview of the model of auditory processing. Adapted from Hamacher (2004).



	N	λ
v1	10000	0.25 mm
v2	5000	0.5 mm
v3	2500	1 mm
v4	2000	1.33 mm
v5	1000	2.5 mm
v6	500	5 mm

Table 1: Model configurations.

Normal hearing model

A model of normal hearing (NH) has been developed for the purpose of having a basis of comparison. It joins the models of the following parts of the auditory system: peripheral ear, basilar membrane and outer hair cells (Baumgarte, 1999), inner hair cells (IHC), synaptic clefts and auditory nerves (AN) (Sumner et al., 2002). Ten thousand IHC-AN complexes are distributed equally along the simulated active cochlea. They are clustered into groups, the activity of which is then averaged over time. A comparison of the various internal data representations is given in Fig. 3.

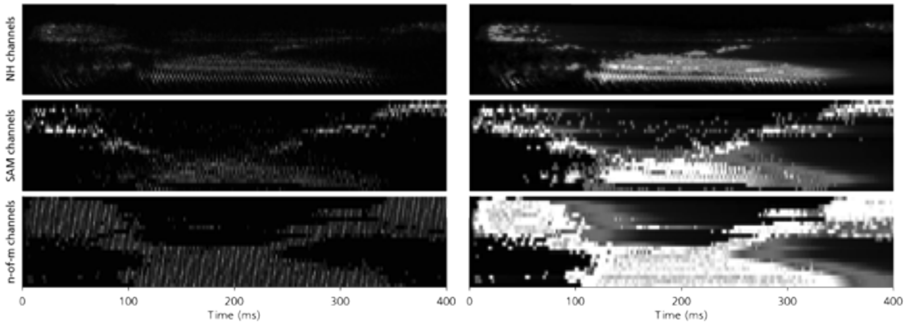


Fig. 3: Overview of the internal data representations for the short utterance "choice". Rows from top to bottom: normal hearing model (NH), SAM strategy, n-of-m strategy. Left column: raw output of the given processing unit, right: data processed by PAP (for NH) or MAP (for CI strategies).

SRT estimation via a DTW speech recognizer

For the speech reception threshold (SRT) estimation, the OLSA (Oldenburg Sentence Test), see Wagener *et al*. (1999), in stationary noise (*olnoise*) is used with a limited vocabulary (50 words). The background noise level is fixed at 65 dB SPL and single words are mixed with *olnoise* with SNRs ranging from -15 to 25 dB in 5 dB steps. For each word and each SNR the internal representations (IRs) are calculated via MAP. Afterwards, the IRs are classified with a dynamic-time warping (DTW) algorithm and therefrom the speech intelligibility function with the parameters SRT and slope s is calculated.

The DTW speech recognizer has a speech memory (consisting of pre-calculated IRs, called response alternatives). For every input word it calculates the "perceptive distance" between the unknown IR and each response alternative. A multiplicative internal noise (σ_{int}) simulates limited resolution ("cognitive factor"). An overview is presented in Fig. 4.

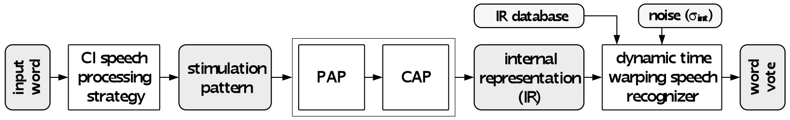

IR database
noise (σ_{int})
input word
CI speech processing strategy
stimulation pattern
PAP
CAP
internal representation (IR)
dynamic time warping speech recognizer
word vote

Fig. 4: SRT estimation via a DTW speech recognizer.

Extraction of pitch cues

To extract cues that possibly support pitch perception, a temporal and place analysis (TPA) for pure tones and sung vowels is done.

The temporal analysis is carried out via discrete-time Fourier transform (DTFT) on the up-sampled input data. While the final statistics use the DTFT magnitude of the whole input signal, the TPA plots (Fig. 7 and Fig. 8) are based on overlapping windows of about 50 ms.

The place analysis accumulates only the input signal magnitudes according to the places on the basilar membrane, which have the characteristic frequencies of typical CI electrode channels.

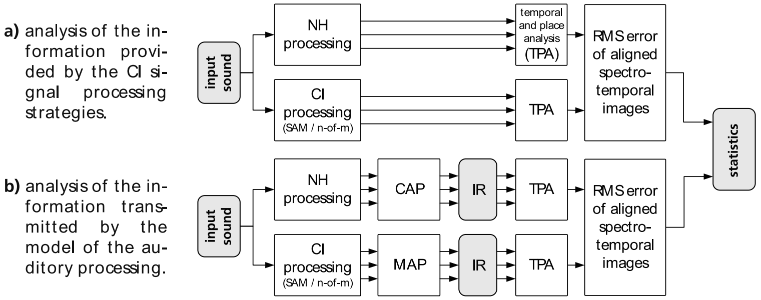

a) analysis of the information provided by the CI signal processing strategies.
b) analysis of the information transmitted by the model of the auditory processing.
input sound
NH processing
CI processing (SAM / n-of-m)
temporal and place analysis (TPA)
TPA
RMS error of aligned spectro-temporal images
statistics
input sound
NH processing
CAP
IR
TPA
CI processing (SAM / n-of-m)
MAP
IR
TPA
RMS error of aligned spectro-temporal images

Fig. 5: Overview of the temporal and place analysis methods.

Pure tones are 65 dB SPL sines with 300 ms linear fade-in. Sung vowels have been created with a concept similar to that proposed by Vandali *et al.* (2005) for pitch ranking tests. Sustained "I" and "Λ" were recorded from a female (pitch key tones: C4-F5) and a male singer (key tones: G2-A#3) with half-tone steps. The two vowels were selected because they differed significantly in spectral shape.

Resulting statistics are based on the difference (root mean square error, RMSE) between the NH-based and CI-processing-based TPA output. An overview of the processing pathways for pitch cue extraction is given in Fig. 5.

RESULTS

The outcomes of the simulation study in terms of the modeled SRT and speech intelligibility versus SNR is summarized in Table 2 and in Fig. 6, respectively.

Model configuration	σ_{int}=0		σ_{int}=0.15		σ_{int}=0.25		σ_{int}=0.35	
	SAM	n-of-m	SAM	n-of-m	SAM	n-of-m	SAM	n-of-m
v1	**-1.9 dB**	-1.6 dB	**-2.1 dB**	-1.9 dB	**-2.2 dB**	-2.0 dB	-1.6 dB	-1.7 dB
v2	**-1.9 dB**	-1.7 dB	**-2.2 dB**	-2.1 dB	**-2.4 dB**	-2.0 dB	**-1.8 dB**	-1.3 dB
v3	-2.5 dB	-2.9 dB	**-2.9 dB**	-2.9 dB	**-2.7 dB**	-1.8 dB	**-1.2 dB**	0.6 dB
v4	-2.9 dB	-3.3 dB	**-3.0 dB**	-3.0 dB	**-2.5 dB**	-1.7 dB	**0.1 dB**	2.1 dB
v5	-3.3 dB	-3.9 dB	-3.1 dB	-2.8 dB	**-1.2 dB**	1.0 dB	**7.0 dB**	10.1 dB
v6	-3.6 dB	-4.4 dB	-2.5 dB	-1.1 dB	8.6 dB	10.2 dB	28.2 dB	25.2 dB

Table 2: Modeled SRT values for the OLSA for various model configurations, strategies and standard deviations σ_{int} of the internal noise. SRT values in bold indicate benefit or equal performance with SAM against the n-of-m strategy.

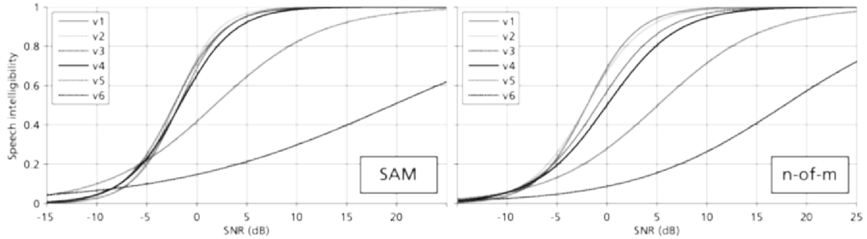


Fig. 6: Modeled speech intelligibility for the OLSA plotted against the SNR for the different model configurations and σ_{int} = 0.30.

Table 2 reveals for both strategies increasing SRTs with increased σ_{int}. The SRT increases with worse cognitive performance especially in model configurations with few auditory nerve cells. Bold values indicate lower or equal SRTs for SAM in comparison to n-of-m. However, differences are only little in most cases. Nevertheless, σ_{int}=0.35 led to lower SRTs for SAM than for the n-of-m strategy in model configurations v2–v5. Model configuration v6, however, could not be well fitted, because the SRT exceeded the highest SNR of 25 dB.

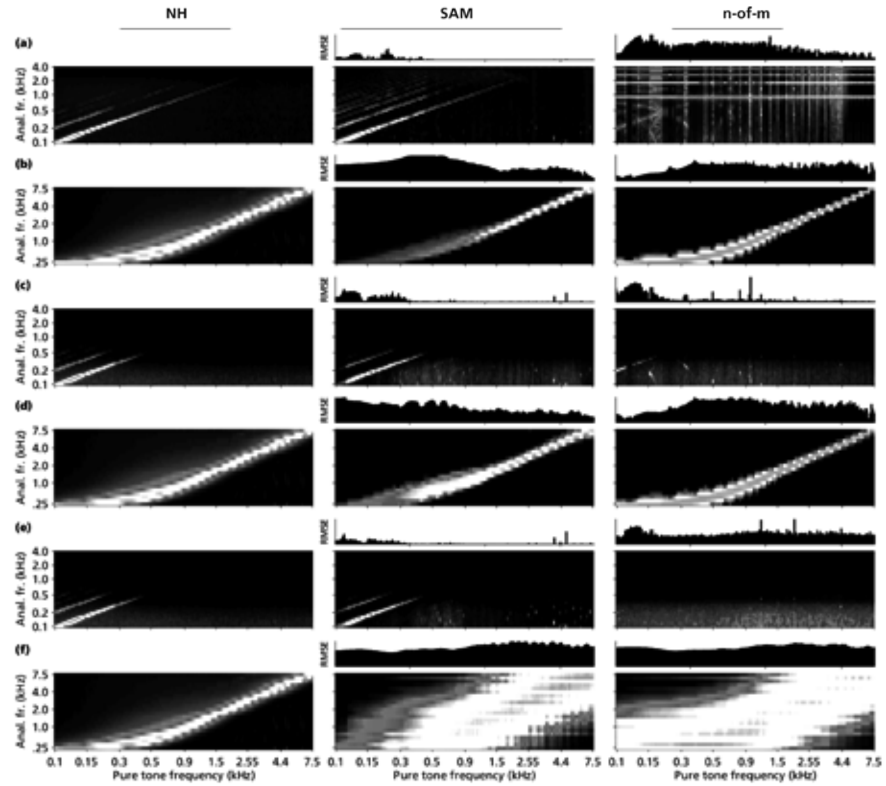


Fig. 7: Temporal (**a**, **c**, **e**) and place (**b**, **d**, **f**) analysis of the information provided (**a**, **b**) by the NH model, SAM and the n-of-m strategy, and of the information transmitted (**c**, **d**, **e**, **f**) by the model, for pure tones. Panels (**c**) and (**d**) show analysis results by using model configuration v1, and panels (**e**) and (**f**) show results with model configuration v6. RMSE shows difference between the actual plot and the corresponding NH plot.

Fig. 7 and Fig. 8 show the results of the temporal and place analysis for pure tones and sung vowels. The leftmost columns present the NH model output for the given input and is used as a basis for comparison. The center and right columns represent data based on SAM and n-of-m processing, respectively. Lighter shades of gray represent higher magnitudes. RMSE values express the difference between the respective plot (below the RMSE plot) and the associated NH plot. RMSE plots of one row are always co-normalized.

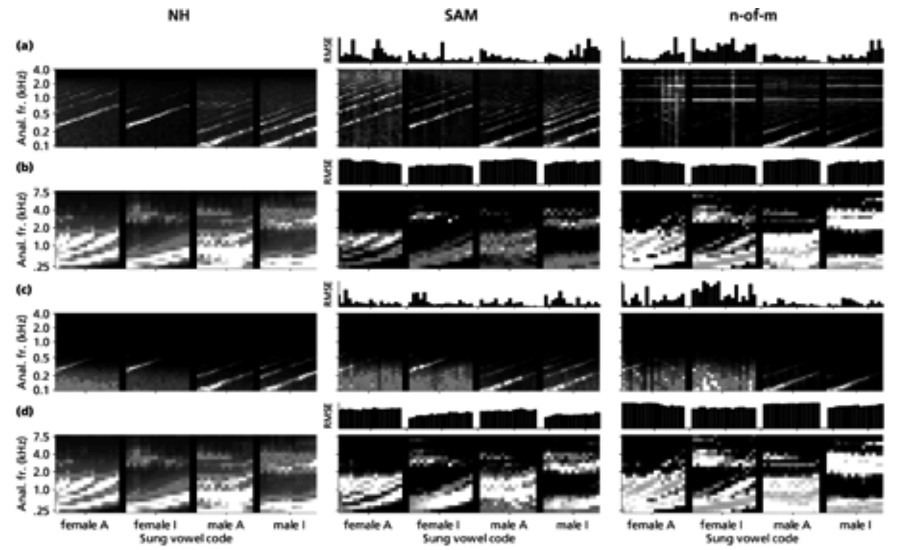


Fig. 8: Temporal (**a**, **c**) and place (**b**, **d**) analysis of the information provided (**a**, **b**) by the NH model, SAM and the n-of-m strategy, and of the information transmitted (**c**, **d**) by the model (v1) for sung vowels.

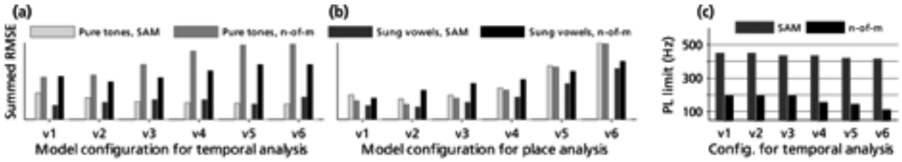


Fig. 9: Panels (**a**) and (**b**) show total RMS error summed over all stimuli with different model configurations. Panel (**a**) shows that temporal pitch cues of the NH model are more similar to those of SAM than to those of n-of-m. Panel (**c**) presents phase-locking (PL) limits for various model configurations. The temporal analysis could not find local maximum in the spectrum for a pure tone with the given frequency over the shown limits.

Place analysis yields similar RMSE with both strategies, but place code is prone to getting indefinite with increasing current spread. Temporal code is better preserved by SAM, where it also seems to be robust against current spread and degeneration of auditory nerves.

SUMMARY

The presented model of the auditory processing and the analysis methods can be used prior to clinical studies with CI users to estimate speech perception and pitch discrimination performance for various CI speech processing strategies.

It can be stated that SAM mimics normal hearing in a more realistic way than the n-of-m strategy does, even though the modeled SRTs show little differences between SAM and n-of-m. With increasing internal noise (worse simulated cognitive condition), however, SAM outperforms the n-of-m strategy especially in model configurations with fewer auditory nerve cells. While the two strategies deliver about the same amount of place pitch cues, SAM provides more temporal pitch cues, which may well contribute to pitch perception according to the modeled results.

Results, of course, needs to be verified with clinical studies.

ACKNOWLEDGMENT

The underlying research reported in the paper was partially funded by the grant B514-09020 of the Thuringian Ministry of Education, Science and Culture.

REFERENCES

Baumgarte, F. (**1999**). “A Physiological Ear Model for the Emulation of Masking” ORL, **61**, 294–304.

Hamacher, V. (**2004**). *Signalverarbeitungsmodelle des elektrisch stimulierten Gehörs.* Aachener Beiträge zu Digitalen Nachrichtensystemen, Mainz, Aachen, Germany.

Harczos, T., Chilian, A., and Husar, P. (**2011**). “SAM: a novel cochlear implant speech coding strategy based on an active cochlea model” *in prep.*.

Hey, M., Hocke, T., Braun, A., Scholz, G., Brademann, G., and Müller-Deile, J. (**2010**). “Erhebung von Normativen Daten für den Oldenburger Satztest bei CI-Patienten” in Proc. 13. Jahrestagung der Deutschen Gesellschaft für Audiologie on CD-ROM.

Jürgens, T., Fredelake, S., Meyer, R., Kollmeier, B., and Brand, T. (**2010**). “Challenging the Speech Intelligibility Index: Macroscopic vs. Microscopic Prediction of Sentence Recognition in Normal and Hearing-impaired Listeners,” in Proc. Interspeech, Makuhari, Japan, 2478–2481.

Shepherd, R.K., Roberts, L.A., and Paolini, A.G. (**2004**). “Long-term sensorineural hearing loss induces functional changes in the rat auditory nerve” Eur. J. Neurosci., **20**, 3131–3140.

Sumner, C. J., Lopez-Poveda, E. A., O’Mard, L. P., and Meddis, R. (**2002**). “A revised model of the inner-hair cell and auditory nerve complex” J. Acoust. Soc. Am., **111**, 2178–2188.

Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W. D., and McDermott, H. J. (**2005**). “Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies” J. Acoust. Soc. Am., **117**, 3126–3138.

Wagener, K., Brand, T., and Kollmeier, B. (**1999**). “Entwicklung und Evaluation eines Satztests für die deutsche Sprache” Zeitschrift für Audiologie, **38**, 4–95.

Horizontal-plane localization with bilateral cochlear implants using the SAM strategy

TAMAS HARCZOS[1,2], ANJA CHILIAN[1,3], ANDRAS KATAI[1]

[1] *Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany*

[2] *Institute for Media Technology, Faculty of Electrical Engineering and Information Technology, Ilmenau University of Technology, Germany*

[3] *Institute of Biomedical Engineering and Informatics, Faculty of Computer Science and Automation, Ilmenau University of Technology, Germany*

Sound source localization capability of cochlear implant (CI) users has been a popular research topic over the past few years, because it has both social and safety implications. While it is widely accepted that unilateral implantation does not provide enough information for this task, conditions, algorithms and their parameterization for the best performance in the binaural case are still in the focus of the research.

On ISAAR 2009, we presented a simulation study revealing the theoretical limits of localization performance using the widespread ACE strategy. We also gave an example of how left-right speech processor asynchrony may influence the perceived direction.

In the present paper we give an outline of a novel, auditory model based CI speech processing strategy called SAM. Furthermore, using the framework from the previous study, we show how localization performance increases when using SAM instead of ACE. We present detailed comparisons to show how factors like pulse rate, signal to noise ratio, reverberation, etc. affect horizontal-plane localization. Finally, we give a simple explanation, why, unlike other strategies, spatial perception with SAM is robust against device asynchrony.

INTRODUCTION

Over the past decade, cochlear implants (CIs) have become a widely accepted alternative for treatment of people with severe to profound hearing loss. While bilateral cochlear implantation (BI) is offered to a growing number of individuals, not all BI-users are 100% satisfied.

One possible cause for the dissatisfaction is the missing ability to robustly localize sound sources. The trend is to use <1K/s channel stimulation rate (CSR) and ≤9K/s total stimulation rate (TSR) with n-of-m strategies like ACE, which, in fact, allows for only very limited localization performance based on temporal cues, as shown e. g. in Harczos *et al.* (2010).

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.