Evaluation of cues for horizontal-plane localization with bilateral cochlear implants

Tamás Harczos[1,2], Stephan Werner[2], Gero Szepannek[3], and Karlheinz Brandenburg[1,2]

[1] *Fraunhofer Institute for Digital Media Technology IDMT, Ehrenbergstr. 31, 98693 Ilmenau, Germany*

[2] *Institute for Media Technology, Faculty of Electrical Engineering and Information Technology, Ilmenau University of Technology, Helmholtzplatz 2, 98693 Ilmenau, Germany*

[3] *Department of Statistics, University Dortmund, Vogelpothsweg 87, 44227 Dortmund, Germany*

Over the past decade, cochlear implants (CIs) have become a widely accepted alternative for treatment of people with severe to profound hearing loss. Now, bilateral implantation is offered to a growing number of individuals in order to provide benefits of binaural hearing. One of these benefits is the ability to localize sound sources.

Current CIs are not able to fully transmit the spectro-temporal information that is available for normal-hearing listeners. Experiments have shown however, that even in the absence of the fine temporal structures, CI-users are able to localize sounds in the horizontal-plane to some extent.

We present a simulation study dealing with possible causes of the limited localization capabilities of CI-subjects. Concentrating on the most commonly used strategy, advanced combination encoder (ACE), we elaborate on the effects of left-right device asynchrony and that of changing the pulse rate and the number of selected spectral bands.

We simulate the CI-processing up to the actuation of the electrodes complemented with a model of current spreading in the endolymph. Relying on this data, we verify localization possibilities based on interaural time differences using a generalized cross-correlation method.

INTRODUCTION

The purpose of this study is to find the bottleneck in the ACE strategy based CI signal processing in terms of sound source localization ability, furthermore, to propose parameterization for ACE, with which localization may work best.

First, results from recent studies with CI-recipients are discussed. Then, basics of localization and the impact of head-related transfer functions (HRTFs) are outlined. Next, the design considerations for the tests and the function of each processing step are explained. Finally, results and outcomes, complimented by a discussion and outlook are presented.

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation, 2010.

PERFORMANCE OF CI-RECIPIENTS

While normal-hearing (NH) listeners perform horizontal-plane localization tests with about 3-5° mean error, bilaterally implanted CI-users are reported to achieve much worse results. Grantham *et al*. (2007) report approx 21-24° (depending on the signal to be localized) mean error for postlingually deafened adults, the performance highly varying among testees. Litovsky *et al*. (2006) found that children (between 3 and 16 years) obtain slightly better results (approx 20° mean error), and that some of the young patients were able to discriminate directions (at least left and right side) with only one CI turned on. Dunn *et al*. (2004) presented a study with 4 bilaterally implanted patients and showed that localization was possible in all cases, independently of the used speech processing strategy. Seeber *et al*. (2004) report, that there was a subject in their test group, who in fact could localize sounds with an average error of only 6.2°.

All these studies are very detailed, still, there are so many unknowns (starting from the exact settings of the CIs), that the results of them can hardly be compared. This paper tries to deliver comparable data at least up to the level of data provided by the CIs.

LOCALIZATION BASICS AND ROLE OF THE HRTF

The main cues to localize an audio source are the interaural time difference (ITD) and interaural level difference (ILD). Though ITD and ILD cues normally get merged in listening situations, this paper elaborates more on the ITDs transmitted by cochlear implants.

In a single dual-receiver system, ITDs mostly depend on source angle α, distance of the two receivers d, and sonic speed v_s. Then, the maximum time difference $\Delta t_{max}=d/v_s$ for ±90° angle of incidence. For sampled systems (like for CIs), sampling frequency F_s and Δt_{max} are critical parameters for the localization abilities. The maximum number L of separable azimuthal directions between 0° and 90°, and the directions α_k that can be best localized are given as:

$$L = \left\lfloor \frac{F_s \cdot d}{v_s} \right\rfloor + 1, \qquad \text{(Eq. 1)}$$

$$\alpha_k = \arcsin\left(\frac{(k-1) \cdot v_s}{F_s \cdot d} \right), k = 1 .. L. \qquad \text{(Eq. 2)}$$

The distance between a microphone pair of left and right CI is approx 16.4 cm. However, experiments on the head-related impulse response (HRIR) data provided by Kayser *et al*. (2009) showed that in the frequency range matching with the lowest frequency bands of a CI, sound has to travel about 27 cm in the case of total

lateralization. For this reason *d=0.27* meters will be assumed all through this paper, unless HRIRs are directly used for the calculations.

DESIGN OF THE EXPERIMENTS

This section describes the design of the experiments done for this study. The framework for the tests was built in MATLAB (version 7.6.0.324, R2008a), calculations were done under Linux on a Dell Precision WorkStation 490. An overview of the processing steps is presented in Fig. 1.

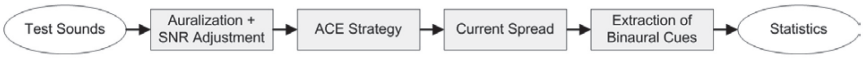

Fig. 1: Overview of processing steps.

Test sounds

A natural sound and a synthetic sound are used as test signals to be localized. Both kind of test signals have 16-bit resolution, are sampled at 16 KHz and are normalized to 0 dB FS peak value.

The natural sound is a 1.34 seconds long part of a randomly picked wave file from the TIMIT Acoustic-Phonetic Continuous Speech Corpus, where a male speaker utters "to open the store by eight" (train/dr4/mbma0/sx232, 0.77-2.11 seconds). This test signal will be referred to as the *Timit* signal. For more information on TIMIT, see Zue *et al*. (1990).

The second test signal is very similar to that used during the listening tests by Seeber *et al*. (2004). It contains 15 white noise pulses (pulse duration is 25 ms with 2 ms linear fade-in and fade-out) divided into 3 groups (pause between groups is 65 ms, pause between pulses within a group is 35 ms). This signal will be referred to as the *Pulses* signal through the rest of this paper.

Simulated environment

For the experiments relying on ITDs, simulated room impulse responses (RIRs) are employed. RIRs are generated based on the source and ear positions in a virtual box-shaped room. The dimensions of the virtual room are chosen to be four times as wide and five times as long as the distance between the microphone mid-point and the sound source. The algorithm uses the image method and virtual sources to calculate the RIRs, and is well documented in McGovern (2009).

An important parameter for generating the RIRs is the reflection coefficient (r). For this paper two different conditions were made up simulating anechoic (r=0.001) and office-like (r=0.7) environments. Reverberation time (T60) for the latter case is around 550 ms for a source to microphone mid-point distance of s=3.0.

To be able to estimate the effects of realistic interaural level differences (ILD), for some setups HRIR-based simulations of the actual test sounds were created, too.

Noise conditions

Tests are repeated five times, simulating five different noise conditions. In the first case, no noise is added to the auralized test signal, and hence it is referred to as the *clean* condition. In the 2nd and 3rd cases uncorrelated stereo white noise (WN) is added to the signal, resulting in 20 and 5 dB signal-to-noise ratios (SNR), respectively. In the 4th and 5th cases a stereo recording of a crowded train station (TS) is used as background noise, reaching the same SNR levels as in the previous cases. The recording had been de-correlated to avoid additional ITD cues, and was dynamic-compressed to yield a smooth envelope.

CI signal processing

To be able to simulate binaural hearing with CIs, ACE algorithm was implemented from scratch in MATLAB as illustrated in Fig. 2. Its input is a stereo wave file, and calculations end at the point, where at each time instant, an electrode to be activated and the corresponding stimulation level are determined.

In ACE the digitized signal is processed in frames. The first step is to filter through a filter-bank. Typically, a 128-point fast Fourier transform (FFT) is used at F_s=16 KHz. The spectral analysis ignores phase information, but keeps magnitudes (excluding DC-component). The envelope detector reshapes the magnitude information into M bands (M is the total number of available electrodes and is typically 22 for ACE-driven systems) yielding quasi-logarithmically spaced center frequencies, as presented in table 1.

Next, the "Sampling & Selection" block in Fig. 2 selects the N largest out of M envelopes ($N < M$) for stimulation. For this paper integer numbers from 1 to 8 will be allowed for N, which is also the range of typically used values.

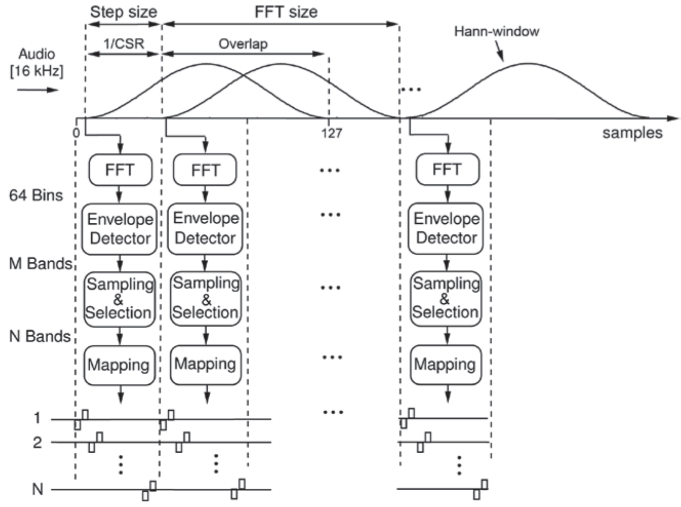

Fig. 2: Overview of the ACE strategy, source: Nogueira (2008).

Channel	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
CF [Hz]	250	375	500	625	750	875	1000	1125	1250	1437	1687
Channel	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
CF [Hz]	1937	2187	2500	2875	3312	3812	4375	5000	5687	6500	7437

Table 1: Center frequencies of the simulated electrodes.

Finally, the envelope amplitudes are mapped to the corresponding electrodes by compressing values into the subject's dynamic range between threshold level (*THL*) and most comfortable level (*MCL*) utilizing the so-called loudness growth function (LGF). The LGF is a logarithmically-shaped function that maps the acoustic envelope amplitude to an electrical magnitude. Its three parameters (base level L_b, saturation level L_s, and steepness ρ) are responsible for both the input dynamic range (IDR) available without clipping and the degree of dynamic compression. While in real situations the *THL* and *MCL* values for each CI channel are always fitted to the patients' needs, these settings play no role in the simulations, and they will be set to 50/256 and 170/256, respectively, for all simulated channels during all tests. LGF parameters are set to L_b=4/256, L_s=150/256, ρ=416.2, which allows for around 30 dB IDR. Input sound files are scaled to a level to approximately fit into the available IDR.

For each frame of the audio signal, *N* electrodes are stimulated sequentially, which, with other words, is one complete stimulation cycle. The requested channel stimulation rate (*CSR*) thus determines the number of cycles/second and also the step size for the spectral analysis. During this study, typical *CSR* values of 720, 900, 1200, 1800, 2400,

3200, and 3500 are tested. Time resolution *dt* of the output becomes $(CSR \cdot N)^{-1}$, where $CSR \cdot N$ is also called total stimulation rate.

The ACE© algorithm presented in this section is the strategy used in the Nucleus® and Freedom® devices of Cochlear Ltd. For more details on ACE see e.g. Nogueira (2008).

Asynchrony of the simulated CIs

If left and right cochlear implant processors are not synchronized, they may be in different states of the ACE strategy at a given time instant. This means, for example, that while the left device has just selected *N* spectral peaks and starts a new stimulation cycle on the corresponding electrodes, the right device may still be in a previous stimulation cycle not yet having processed the data for the new spectral peaks. Since one processing cycle takes CSR^{-1} seconds, the maximum binaural slip $-\frac{1}{2} \cdot CSR^{-1} < t_{slip} < \frac{1}{2} \cdot CSR^{-1}$ holds. Some distinguished situations, where t_{slip} is a multiple of F_s^{-1}, will be discussed later in this paper.

Current spread model

A simple model of current spread is used to simulate the electrode-to-auditory nerve interface by approximating the excitation along the cochlea, as shown in Nogueira (2008). An exponential decay function models the current density in the proximity of a stimulating pulse:

$$E_z(x) = e^{\frac{-|X(z)-x|}{\lambda}}, \qquad \text{(Eq. 3)}$$

where *z* is the number of the stimulating electrode, *X(z)* is the position in millimeter of the electrode, *x* is the position on the cochlea where current density is to be measured, and λ represents the extent of current spread (in millimeter).

Since the center frequencies of the simulated electrodes are known, positions along the cochlea can be calculated using the inverse of the place-frequency equation given by Greenwood (1990). Tests discussed in this paper are repeated three times, with different current spread settings, λ=0 (no spread), λ=0.5, and λ=2.0.

Finding binaural cues

Once the calculations are done, result matrices for left and right channels (also called left and right auditory images, AI_S-L and AI_S-R for the simulated RIR-based and AI_H-L and AI_H-R for the HRIR-based calculations, respectively) are compared.

For the ITD cues AI_S-L and AI_S-R are cross-correlated on a window-by-window basis (window size is 30 ms, no overlapping is allowed) using generalized cross-correlation (GCC) without any spectral weighting. Maximum lags are determined by applying

equation 1. The localized degree is determined for each data window by finding the maximum in the cross-channel-summed cross-correlation function and indexing equation 2 with the "winner" lag.

ILD cues are simply approximated by summing up all stimuli in AI_H-L and AI_H-R for the whole duration of the test signal.

Statistics

For each trial, statistics in terms of mean, standard deviation, and extremes of the localized degrees among data windows are gathered and stored along with the found ILD cues and calculation time. The most interesting outcomes will be presented during the next section.

RESULTS

During the first tests, localization accuracy for various parameterization of ACE was inspected, while test file, reverberation, noise and current spread conditions were fixed.

Figure 3a shows how localized directions vary as a function of source direction. Mean and standard deviation (STD) values are calculated from the localization results of the 30 ms windows. The step-like alignment of the mean values corresponds to the directions, which can be represented best with the given sampling rate F_s (which from the localization's point of view equals to *CSR*) and receiver distance *d* settings. STD gets large for degrees, where ambiguity is high during the localization task.

Panels b, c, and d of Fig. 3 represent mean errors and mean of STD for all tested directions between 5° and 90° (5° steps, the trivial 0° case is excluded). An interpretation of these figures can be as follows. In general, higher stimulation rates perform better (yield lower mean error). With the typical setting of N=8, the lowest rate where localization gets possible is *CSR*=1200. *CSR*=1800 can be a good choice for all *N*. Even though its mean error is a bit higher than that of *CSR*=1200, standard deviation is moderate, which ensures lower ambiguity and more robust side discrimination. N=4 seems to be a good compromise for all CSR ≥1200; large N at lower rates seem to render localization impossible. $N<4$ at lower rates, however, seem to add some fuzziness to the auditory images, so that localization (with very high ambiguity) may become feasible again.

Left panel in Fig. 4 deals with the effects of changing the signal type, grade of reflection, and noise conditions (WN and TS stand for white noise and train station noise, respectively, while the dB values show signal to noise ratios). It can be stated that a reverberating environment decreases localization robustness largely, but pulse-like sounds suffer from more added localization error than speech. Furthermore, in the case of speech input signal and babble-like (TS) noise, each dB of lost SNR seems to add approximately 1° to the mean localization error.

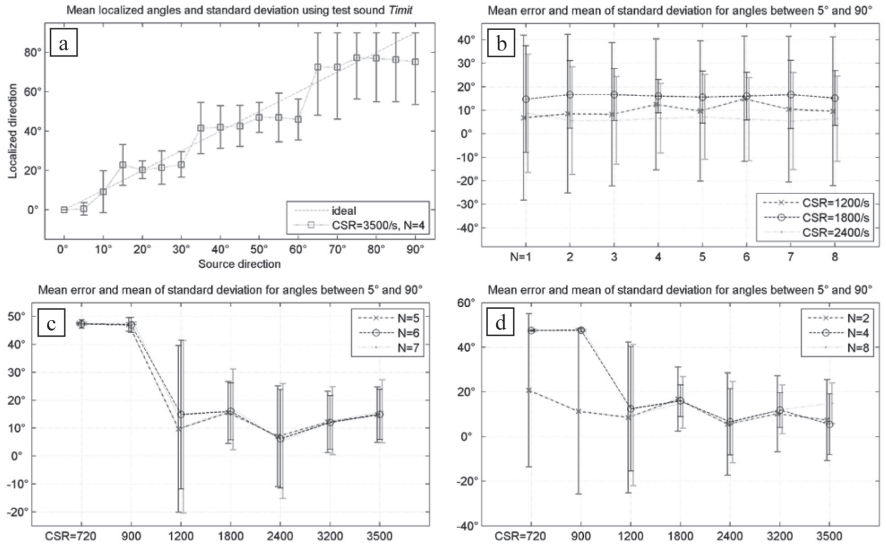

Fig. 3: Localization accuracy as a function of ACE parameters *CSR* and *N*. (Test file: *Timit*, *r*=0.001, noise condition: *clean*, *λ*=0.5.)

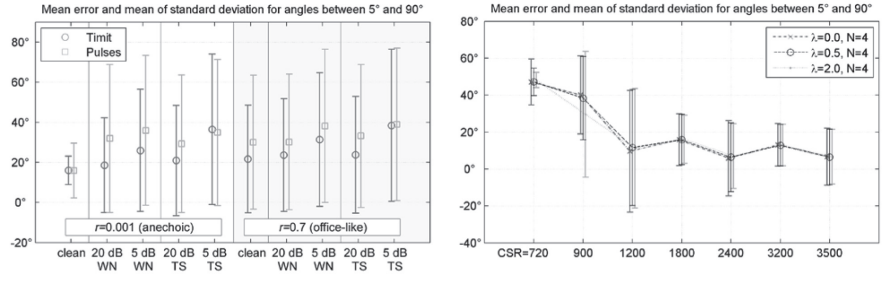

Fig. 4: Localization accuracy as a function of noise, reverberation, and current spread settings. (Left: *CSR*=1800/s, *N*=4, *λ*=0.5. Right: test file Pulses, *r*=0.001, noise condition: clean.)

Right pane of Fig. 4 shows localization quality with different levels of current spread. A larger spread seems to have a positive effect at low pulse rates (smaller STD at 720/s and smaller mean error at 900/s), but does not have impact on localization at higher *CSR*.

Localization deficits as a function of left and right CI asynchrony are depicted in left panel of Fig. 5. Effects are shown at directions, which can normally be localized best with the used settings. The influence of asynchrony may vary a lot among other directions.

Right panel of Fig. 5 illustrates how unstable ILD cues may become, too. For this HRIR-based experiment stimuli for left and right CIs during a complete localization task are aggregated and the sums are converted to a common decibel scale. It is clear to see that none of the two graphs are monotonic, plus, they have more intersections. This can be due to a combination of frequency dependency of the HRIRs, and varying channel properties and high compression of the CIs.

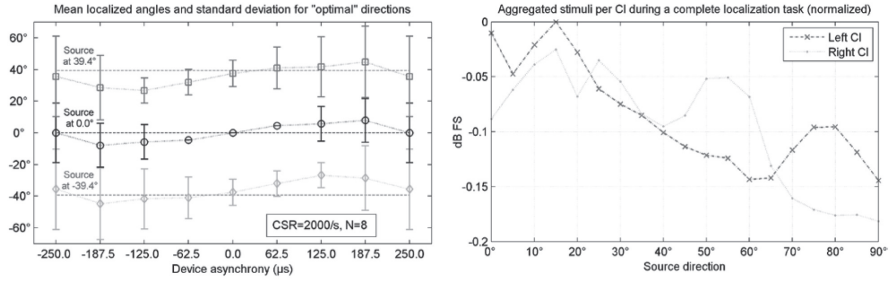

Fig. 5: Left: effects of device asynchrony (λ=0.5). Right: ILD cues (*CSR*=1800/s, *N*=4, λ=0.0). (For both calculations: test file *Timit*, *r*=0.001, noise condition: *clean*.)

DISCUSSION AND OUTLOOK

We have found that at low pulse rates a smaller number of selected spectral peaks per cycle may be of advantage. We have seen furthermore, that higher stimulation rates generally yield better localization results. Laback and Majdak (2008) state that CI listeners' sensitivity to ITDs decline at higher pulse rates, which may be due to binaural adaptation. They have found, however, that by introducing jitter in the binaurally synchronized stimulation, ITD perception can largely be recovered. Based on these findings, it may be desirable to include localization tests (as a secondary goal beyond speech perception) in binaural CI fitting sessions.

Device asynchrony has an impact on localization ability, but its effects (even if they are not the same for all directions) can partly be estimated. Since the localization error introduced by asynchrony is periodical with CSR^{-1}, ever changing asynchrony between left and right devices may indeed be a cause for some CI-recipients to feel dizzy and to dislike binaural stimulation.

A planned extension to this study is to explore the effects of divergent settings for the left and right side, and to verify some of the findings with implantees.

ACKNOWLEDGMENT

We would like to thank Prof. Birger Kollmeier for his useful comments. Also many thanks to Hendrik Kayser and colleagues for making the "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses" accessible to researchers.

REFERENCES

Dunn, C. C., Tyler, R. S., Witt, S. A. Gantz, B., and Rubinstein, J. (**2004**). "Frequency and electrode contributions to localization in bilateral cochlear implants," International Congress Series **1273**, 443-446.

Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Labadie, R. F., and Haynes, D. S. (**2007**). "Horizontal-Plane Localization of Noise and Speech Signals by Postlingually Deafened Adults Fitted With Bilateral Cochlear Implants," Ear and Hearing **28**, 524-541.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species - 29 years later," J. Acoust. Soc. Am. **87**, 2592-2605.

Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., and Kollmeier, B. (**2009**). "Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses," EURASIP J. Advances in Signal Processing, Article ID 298605, doi:10.1155/2009/298605.

Laback, B. and Majdak, P. (**2008**). "Binaural jitter improves interaural time-difference sensitivity of cochlear implantees at high pulse rates," PNAS **105**, 814-817.

Litovsky, R. Y., Johnstone, P. M., Godar, S., Agrawal, S., Parkinson, A., Peters, R., and Lake, J. (**2006**). "Bilateral cochlear implants in children: localization acuity measured with minimum audible angle," Ear and Hearing **27**, 43-59.

McGovern, S. G. (**2009**). "Fast image method for impulse response calculations of box-shaped rooms," Applied Acoustics **70**, 182-189.

Nogueira, W. (**2008**). *Design and Evaluation of Signal Processing Strategies for Cochlear Implants based on Psychoacoustic Masking and Current Steering,* PhD Dissertation (Gottfried Wilhelm Leibniz Universität Hannover).

Seeber, B. U., Baumann, U., and Fastl, H. (**2004**). "Localization ability with bimodal hearing aids and bilateral cochlear implants," J. Acoust. Soc. Am. **116**, 1698-1709.

Zue, V., Seneff, S., and Glass, J. (**1990**). "Speech database development at MIT: TIMIT and beyond," Speech Communication **9**, 351-356.