BINAURAL SPECTRAL COMPLEXITY REDUCTION OF MUSIC SIGNALS FOR COCHLEAR IMPLANT LISTENERS

Johannes Gauer, Anil Nagathil, and Rainer Martin

Institute of Communication Acoustics, Ruhr-Universität Bochum, Bochum, Germany
email: {firstname.lastname}@rub.de

ABSTRACT

An emphasis on the leading voice or melody is known to facilitate music perception in cochlear implant (CI) listeners while a competing accompaniment is perceived as disturbing. In this paper we present the extension of a monaural music complexity reduction scheme for CI users towards a binaural application. The scheme aims at an attenuation of the accompaniment in music signals and relies on a reduced-rank approximation by means of principal component analysis (PCA). In the proposed binaural system the PCA is only performed for the melody dominated ear and its eigenvectors are used for a reduced-rank representation of both ear signals. We use SIR and SAR measures for evaluation and show that with binaural processing a further attenuation of the accompaniment can be achieved in comparison to separate bilateral processing of both ear signals. At the same time neither additional artifacts are introduced to the reconstructed melody signals nor the binaural cues accessible to CI users are considerably harmed.

Index Terms— Music signal processing, cochlear implants, principal component analysis, binaural hearing

1. INTRODUCTION

With more than 300,000 implanted patients, cochlear implants (CI) have become a widespread means to restore the hearing ability of deaf or severely hearing-impaired patients by electrical stimulation of the auditory nerve via an electrode array implanted in the cochlea [1]. The currently used stimulation strategies are successfully optimized for regaining speech perception, as speech communication is an important facet of social life. Music, however, remains less accessible for the majority of CI users [2]. Due to technological and physiological restrictions information is lost in the spectral, fine-temporal, and dynamic range representation [3]. Hence the hearing impressions provided by CIs appear to be unnatural and distorted. The limited number of electrodes (≤ 22) and a mismatch of the place-pitch mapping between the electrodes and the stimulated hearing nerve fibers [4] result in a poor spectral representation of the stimulation signals. This leads to distortions of pitch, timbre, and melody whereas rhythmic information can be accessed by CI users almost as good as by normal hearing (NH) listeners [5]. To make listening to music a more enjoyable experience for CI users, recently several approaches have been proposed that tackle the constraints of music perception in CI users [3] by reducing the spectral complexity of music signals. It has been shown, that an emphasis on vocals, drums and bass in pop music recordings and an attenuation of the other instruments is appreciated by CI users [6]. In [7] the authors propose a music pre-processing scheme based on a harmonic/percussive sound separation (HPSS) that performs a spectral complexity reduction by attenuating the harmonic portion of a signal while preserving strong rhythmic elements like drums represented by the percussive portion. In order to preserve and promote the vocals that are typically placed in the center of the stereo image during the production, also spatial information from stereo music recordings was used [8].

Complexity reduction can also be achieved by reducing the number of musical instruments involved in a performance. CI users rated versions of a country music piece that were manually re-engineered from multitrack recordings more enjoyable than the original recording [9]. The multitrack recordings required for such remixing approaches are generally not available, hence in [10] the application of source separation methods using non-negative matrix factorization (NMF) is investigated. It is shown that CI users often do not benefit from general mixing presets, but rather need individual mixes depending on the musical piece and the individual subject.

Different to full source separation with subsequent remixing the estimated sources, dimensionality reduction techniques also reduce the spectral complexity of a music signal [11]. This approach is based on the assumption that a predominant leading voice is accompanied by one or several instruments as e.g. in classical chamber music. The spectrum will show the strong partial tones of the melody as its most prominent elements. These are identified by principal component analysis (PCA) [12] and preserved while less prominent spectral components are dismissed and thus attenuated. For this method significant preference ratings in comparison to unprocessed signals were found in listening tests with CI users [13]. Thinning out the series of overtones belonging to individual tones in a melody can also reduce the spectral complexity of a corresponding music signal. In [14], the harmonic series of each tone in a monophonic music piece played by 7 different instruments was individually reduced to 5 different harmonic levels by means of custom-fit low-pass filters. When presented with the melody tone's fundamental frequency F_0 only, CI users and NH listeners with CI simulation rated the signal most pleasant. In summary, a reduction of the spectral content in a music signal and an emphasis on the leading voice, on vocals, and on low frequency and percussive portions are found to increase the enjoyment of music in CI users.

The preceding works on spectral complexity reduction mostly considered monaural signals only. The aforementioned pre-processing scheme based on HPSS exploits stereo input signals but its output signal remains also monaural [8, 15]. However, an increasing number of CI users is bilaterally implanted and there is strong evidence, that bilateral stimulation can improve speech perception and source localization ratings in CI users [1]. In this work we present a pre-processing scheme for a spectral complexity reduction of binaural (stereo) music signals which is based on the dimensionality reduction approach [11]. It relies on the assumption that the sources of the leading and accompaniment voices of a music piece are spatially

This work is funded by the German Research Foundation (DFG), Collaborative Research Center 823, Subproject B3.

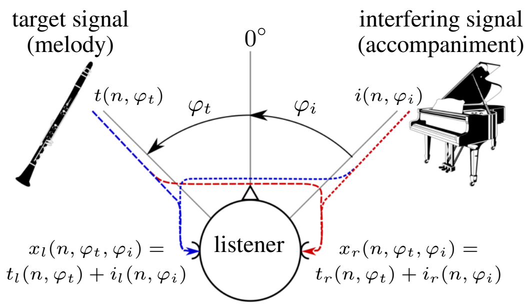

Fig. 1: Incidence angles of melody and accompaniment signals in relation to the listener's position and composition of the ear signals.

distributed around the listener's position so that their respective sound signals arrive from different incidence angles. We show that with a common, binaural processing of the left and right ear signals both the attenuation of the accompaniment can be improved and a reduction of the processing artifacts is achieved. To this end, PCA is only performed for the ear signal facing the melody source and the resulting basis vectors are also used for the ear signal on the contralateral side. The proposed method also reduces computational costs and preserves important binaural cues. It is evaluated in terms of established signal quality measures such as signal-to-interferer-ratio (SIR) and signal-to-artifacts-ratio (SAR).

The remainder of this paper is organized as follows: In Section 2 we will first review the spectral complexity reduction by means of principal component analysis (PCA) for monaural signals and subsequently extend this method to be efficiently used for binaural signals. In Sections 3 and 4 the experimental setup and the results are presented and discussed. Conclusions are drawn in Section 5.

2. SPECTRAL COMPLEXITY REDUCTION WITH PCA

2.1. Monaural spectral complexity reduction

The spectral complexity reduction is performed on classical chamber music pieces with a melody instrument playing the leading voice and one or several other instruments playing the accompaniment. A music piece can be represented by a monaural or binaural discrete-time signal $x(n) = t(n) + i(n)$ of length N, where the target signal $t(n)$ contains the melody, the interfering signal $i(n)$ contains the accompaniment, and n denotes the discrete time index. In order to perform a block-wise spectral complexity reduction, the signal $x(n,\lambda) = x(n + \lambda R)$ is split into overlapping segments and for each of these segments a short-time spectrogram representation $X(\kappa, \lambda)$ is computed. The segment index, the segment shift, and the frequency index are denoted by λ, R, and κ respectively. These spectrograms can be computed by any appropriate spectral transformation like a short-time Fourier transform (STFT) or a constant-Q transform (CQT). The CQT provides a frequency analysis grid f_κ with a variable bandwidth $\Delta f_\kappa = f_\kappa/Q$ and it is well suited to describe the geometrically spaced frequencies of the scales usually used in western music. Therefore, we prefer CQT to the STFT in this work. Its spectral resolution is adjusted via the quality factor $Q = f_\kappa/\Delta f_\kappa = 1/(2^{\frac{1}{12b}} - 1)$, where the parameter $b \in \mathbb{N}$ specifies number of spectral bins per semitone. The CQT is defined as

$$X_{\mathrm{cqt}}(\kappa,\lambda) = \frac{1}{N_\kappa} \sum_{n \in \mathcal{N}_\kappa} x(n,\lambda) w_\kappa(n) \exp\left(-j\frac{2\pi Q n}{N_\kappa}\right) \tag{1}$$

for a signal segment of length $N_0 = Qf_s/f_0$ [16]. To avoid spectral leakage, Hann analysis windows with a frequency-dependent length of $N_\kappa = f_s/\Delta f_\kappa = Qf_s/f_\kappa$,

$$w_\kappa(n) = \begin{cases} 0.5\left(1 - \cos\left(\frac{2\pi\left(n - \frac{N_0 - N_\kappa}{2}\right)}{N_\kappa - 1}\right)\right) & \forall\, n \in \mathcal{N}_\kappa \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

with $\mathcal{N}_\kappa = \left\{\frac{N_0 - N_\kappa}{2}, \frac{N_0 - N_\kappa}{2} + 1, \ldots, \frac{N_0 + N_\kappa}{2} - 1\right\}$ are applied. The obtained short-time CQT spectra

$$\mathbf{X}_{\mathrm{cqt}}^{(\lambda)} = [X_{\mathrm{cqt}}(0,\lambda), X_{\mathrm{cqt}}(1,\lambda), \ldots, X_{\mathrm{cqt}}(K-1,\lambda)]^{\mathrm{T}}$$

are combined to blocks consisting of B_m frames:

$$\mathbf{U}^{(m)} = [\mathbf{X}_{\mathrm{cqt}}^{(\lambda_m)}, \mathbf{X}_{\mathrm{cqt}}^{(\lambda_m+1)}, \ldots, \mathbf{X}_{\mathrm{cqt}}^{(\lambda_{m+1}-1)}]^{\mathbf{T}} \in \mathbb{C}^{B_m \times K}.$$

The block index m will be omitted in the following for notational convenience.

Performing PCA [12] on the original spectral blocks solves the eigenvalue problem $\mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{w}_k = d_k \mathbf{w}_k$ and delivers the eigenvalues d_k of the covariance matrix $\mathbf{C}_{uu} \sim \mathbf{U}^H\mathbf{U}$ and the corresponding eigenvectors $\mathbf{w}_k$ which span an orthogonal basis $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_k, \ldots \mathbf{w}_K] \in \mathbb{C}^{K \times K}$. The index of the principal components is denoted by $k \in 1, 2, \ldots K$, where K corresponds to the number of spectral components. PCA returns the eigenvectors $\mathbf{w}_k$ sorted in descending order of their corresponding eigenvalues, i.e. $d_1 \geq d_2 \geq \ldots \geq d_K$. Hence, the first eigenvectors carry the highest portion of the overall variance and represent the most prominent spectral bands of the respective block. Projecting the original spectral block $\mathbf{U}$ onto its basis $\mathbf{W}$ yields the coefficient or score representation

$$\mathbf{T} = \mathbf{U}\mathbf{W}, \tag{3}$$

where $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots \mathbf{t}_k, \ldots \mathbf{t}_K] \in \mathbb{C}^{B_m \times K}$ comprises the coefficient vectors belonging to each principal component. The dimensionality reduction is achieved by retaining a reduced basis $\widehat{\mathbf{W}} = [\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_k, \ldots \mathbf{w}_{\hat{k}}] \in \mathbb{C}^{K \times \hat{k}}$. It consists only of a selected number $\hat{k} \leq K$ of basis vectors and spans a subspace of the complete representation. The reduced-rank approximation $\widehat{\mathbf{U}}$ of the original spectrogram block is obtained by

$$\widehat{\mathbf{U}} = \widehat{\mathbf{T}}\widehat{\mathbf{W}}^{\mathrm{H}} = \mathbf{U}\widehat{\mathbf{W}}\widehat{\mathbf{W}}^{\mathrm{H}},\ \widehat{\mathbf{U}} \in \mathbb{C}^{B_m \times K}. \tag{4}$$

In order to yield the rank-$\hat{k}$ approximation of the mixed signal, the reduced full-length signal is reconstructed block-wise via the overlap-add method from signal segments obtained by the inverse CQT as proposed in [17]. Given the linearity of PCA, it can be described as a sum of the rank-$\hat{k}$ approximations of the target and interfering signals, respectively:

$$\hat{x}_{\hat{k}}(n) = \hat{t}_{\hat{k}}(n) + \hat{i}_{\hat{k}}(n). \tag{5}$$

2.2. Binaural spectral complexity reduction

In realistic listening situations (concert situation or stereophonic playback with the listener in the "sweet spot") the melody signal $t(n)$ and the accompaniment signal $i(n)$ of a music piece arrive at the listener's position at angles of φ_t and φ_i respectively (see Figure 1). Therefore, the left and right ear signals, $x_l(n) = t_l(n) + i_l(n)$ and $x_r(n) = t_r(n) + i_r(n)$, consist of different mixtures of the melody and accompaniment signals. The spectral complexity reduction could thus be computed for both ear signals separately and independently. Besides the doubling of computational costs in comparison to the monaural case, this might lead to an undesirable behavior when the accompaniment predominates the melody on one ear. Hence, we propose a common binaural processing of both ear signals where the PCA basis matrix $\mathbf{W}_t$ is computed only for the melody-dominated side but is used for the spectral complexity reduction on both sides. Thereby we assume that the melody signal's direction of arrival in relation to the listener is known. The knowledge about the exact incidence angles is not necessary. The predominant melody side can e.g. be determined by comparison of the respective signal energies on each side or by measuring spectral sparsity. To illustrate our approach, let us assume that the melody source is located on the left hand side of the listener (see Figure 1). In this case only the corresponding basis matrix $\mathbf{W}_t = \mathbf{W}_l$ needs to be computed, and it is used to obtain the reduced-rank spectra for both sides:

$$\widehat{\mathbf{U}}_{\mathbf{l}} = \mathbf{U}_l \widehat{\mathbf{W}}_t \widehat{\mathbf{W}}_t, \quad \widehat{\mathbf{U}}_{\mathbf{r}} = \mathbf{U}_r \widehat{\mathbf{W}}_t \widehat{\mathbf{W}}_t. \tag{6}$$

After reconstruction we obtain the reduced-rank approximations of the mixed ear signals as

$$\hat{x}_{l,\hat{k}}(n) = \hat{t}_{l,\hat{k}}(n) + \hat{i}_{l,\hat{k}}(n) \text{ and } \hat{x}_{r,\hat{k}}(n) = \hat{t}_{r,\hat{k}}(n) + \hat{i}_{r,\hat{k}}(n). \tag{7}$$

As only the basis vectors for the melody-dominated side are used for reconstruction, the computational costs are reduced, and as shown below, a further attenuation of the accompaniment in the contralateral right ear signal is achieved.

3. EXPERIMENTAL SETUP

The proposed binaural spectral complexity reduction scheme is applied on a database containing 110 MIDI files of chamber music phrases, each with a leading and an accompanying voice and a length of $T = 10\,\mathrm{s}$ [11]. The audio files are synthesized with high quality samples based on recordings of real instruments using *Native Instruments Komplete*[1]. Prior to further processing, the signals $t(n)$ and $i(n)$ are normalized to 0 dB input SIR. To obtain a reasonably natural sound impression, also a slight amount of artificial room reverb was added to the synthesized audio fragments. Additional room reverberation was not considered as it might blur the directional effects. For higher amounts of reverberation, we expect a convergence towards the monaural case as this leads to a more diffuse sound field. To evaluate the influence of different incidence angles on the spectral complexity reduction, a set of 91 ear signals with combinations of 7 incidence angles φ_t for the target or melody signal and 13 incidence angles φ_i for the interfering or accompaniment signal is created for each phrase in the database. The incidence angles in the azimuthal plane range between $0^\circ \leq \varphi_t \leq 90^\circ$ and $-90^\circ \leq \varphi_i \leq 90^\circ$ respectively, varying with a step size of $\Delta\varphi_{t,i} = 15^\circ$. This angular resolution appears to be sufficient as in CI users smaller minimal audible angles (MAA) ranging from 4° to 8° can only be found in the frontal and dorsal quadrant while the MAA in the lateral quadrants typically exceeds 30° [18]. The database signals $t(n)$ and $i(n)$ are convolved with head-related impulse responses (HRIRs) in order to obtain the clean target and interfering ear signals $t_l(n, \varphi_t)$, $t_r(n, \varphi_t)$, $i_l(n, \varphi_i)$, and $i_r(n, \varphi_i)$ (see Figure 1). These HRIRs are taken from a multichannel database [19] and were measured with the center microphone of three-channel behind-the-ear (BTE) hearing aids installed at a human head and torso simulator. The clean ear signals are finally mixed, resulting in $x_l(n, \varphi_t, \varphi_i) = t_l(n, \varphi_t) + i_l(n, \varphi_i)$ and $x_r(n, \varphi_t, \varphi_i) = t_r(n, \varphi_t) + i_r(n, \varphi_i)$. The indices $\hat{k}$, φ_t, and φ_i will be dropped in the following unless needed.

[1] http://www.native-instruments.com

4. EXPERIMENTAL RESULTS

4.1. Accompaniment attenuation and melody distortion

The spectral complexity reduction is performed for separate bilateral processing ("sep") and binaural processing ("bin") in parallel to evaluate the differences between both cases. For that purpose the clean melody and accompaniment ear signals, which are available in the synthesized MIDI database, are also subjected to spectral complexity reduction using the PCA basis determined for the mixed ear signals. This procedure allows the evaluation of both the desired attenuation of the accompaniment and distortions of the leading voice in the reconstructed melody signal by means of the SIR and SAR measures. In analogy to [11, 20] and with $||a||^2 := \sum_{n=1}^{N} a^2(n)$ denoting the energy of a signal $a(n)$ with N samples they are computed for the left (l) and right (r) side as

$$\mathrm{SIR}_{l,r} = 10 \log_{10} \left(\frac{||t_{l,r}||^2}{||\hat{i}_{l,r}||^2} \right) \tag{8}$$

$$\mathrm{SAR}_{l,r} = 10 \log_{10} \left(\frac{||t_{l,r}||^2}{||\hat{t}_{l,r} - t_{l,r}||^2} \right). \tag{9}$$

In Figure 2 (a) we depict the mean SIR difference between binaural and separate bilateral complexity reduction

$$\Delta\mathrm{SIR} = \mathrm{SIR}_{bin} - \mathrm{SIR}_{sep} = 10 \log_{10} \left(\frac{||\hat{i}_{sep}||^2}{||\hat{i}_{bin}||^2} \right), \tag{10}$$

at the side of the listener facing the accompaniment source over all 110 pieces of the database and for opposite incidence angles ($\varphi_t = -\varphi_i$). As an attenuation of the interfering accompaniment is desired, a higher $\Delta\mathrm{SIR}$ means a better suppression of the accompaniment with binaural processing compared to separate processing. Besides small numbers of retained PCA components ($\hat{k} \leq 10$), incidence angles in the range of $\varphi_{t,i} = \pm 60 \ldots 75^\circ$ lead to the highest accompaniment attenuation. Listening tests with CI users and the monaural complexity reduction scheme [11, 13] show, that the reduced-rank approximations obtained significant preference ratings. Thus the SIR improvement described by $\Delta\mathrm{SIR}$ correlates well with subjective preference. Nevertheless, for strong dimensionality reduction with small $\hat{k}$, the reconstructed signals also bear a higher amount of artifacts generally resulting in a lower SAR. Figure 2 (b) shows the mean SAR difference between binaural and separate complexity reduction

$$\Delta\mathrm{SAR} = \mathrm{SAR}_{bin} - \mathrm{SAR}_{sep} = 10 \log_{10} \left(\frac{||\hat{t}_{sep} - t||^2}{||\hat{t}_{bin} - t||^2} \right) \tag{11}$$

over all 110 pieces in the database which considerably improves. Thus, binaural processing leads to less distortion of the leading voice. The highest $\Delta\mathrm{SAR}$ improvement with binaural processing is again achieved for $45^\circ \leq \varphi_t \leq 75^\circ$, $-45^\circ \geq \varphi_i \geq -75^\circ$, and $5 \leq \hat{k} \leq 10$ retained PCA components.

Figure 2 (c) shows that a good trade-off between accompaniment attenuation and melody distortion can be achieved for a number of retained components in the range of $5 \leq \hat{k} \leq 10$ that hardly depends on the incidence angles. These values correspond to the results from listening experiments with CI users for monaural spectral complexity reduction, where $\hat{k} = 8$ lead to the highest preference scores in comparison to $\hat{k} = 13$ and the unprocessed signal [13].

Listening to both binaurally and separately processed signals from the MIDI chamber music database in comparison reveals effects on the accompaniment portion that range between a reduction

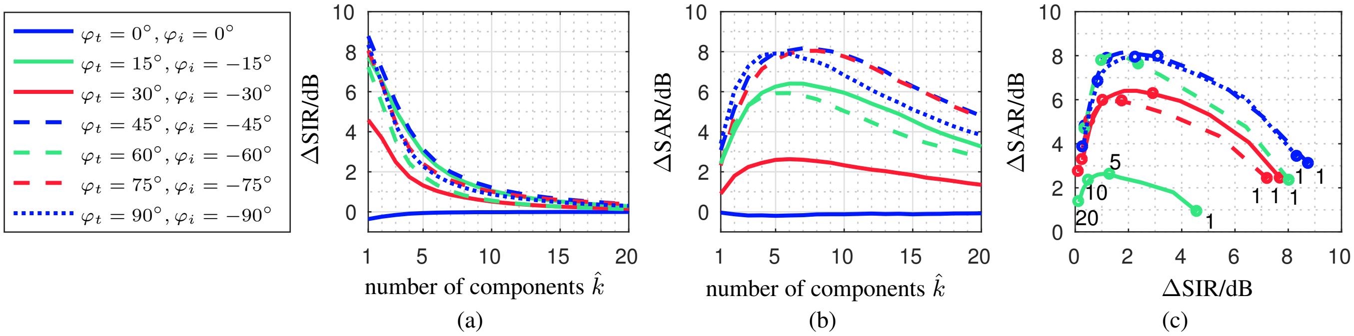

Fig. 2: Mean ΔSIR (a) and ΔSAR (b) results for the MIDI database with opposite incidence angles ($\varphi_t = -\varphi_i$), and ΔSAR over ΔSIR results (c) parameterized by $\hat{k} = \{1, 5, 10, 20\}$ growing from right to left.

of the overtones and a smearing of the attacks to a noticeable suppression of the accompaniment. The presented SIR improvements are thus clearly perceptible in the reconstructed processed signals. The degree of these effects depends both on the chosen number of retained components $\hat{k}$ and on the incidence angles φ_t and φ_i. We also analyzed the outcome for equal incidence angles ($\varphi_t = \varphi_i$), a case that mostly resembles to the monaural situation as the same HRIR is applied to both melody and accompaniment signals. In the range of $5 \le \hat{k} \le 15$ retained PCA components, which is particularly relevant for practical application, ΔSIR values between 1 and 2 dB are obtained for the opposite ear while ΔSAR also does not exceed 2 dB. In consequence, although in this case the accompaniment attenuation can also be slightly improved with binaural processing, the overall performance does neither benefit nor suffer significantly from binaural processing. We expect a similar outcome also for more reverberant signals.

4.2. Preservation of binaural cues

Bilaterally implanted CI users mainly rely on interaural level differences (ILDs) to localize sound sources [21]. The ILD is defined as

$$\text{ILD} = 10\log_{10}(P_l) - 10\log_{10}(P_r) = 10\log_{10}\left(\frac{P_l}{P_r}\right), \quad (12)$$

where P_l and P_r denote the power of the left and right ear signals [22]. As in the present case the attenuation of the interfering accompaniment signal is desired as a matter of principle, the processed and the unprocessed clean melody signals $\hat{t}(n)$ and $t(n)$ need to be compared to determine modifications of binaural cues introduced by the signal processing. This yields the ILD difference

$$\begin{aligned} \Delta\text{ILD} &= \text{ILD}_{bin} - \text{ILD}_{ear} \\ &= 10\log_{10}\left(\frac{||\hat{t}_{bin,l}||^2 \cdot ||t_r||^2}{||\hat{t}_{bin,r}||^2 \cdot ||t_l||^2}\right). \end{aligned} \quad (13)$$

The difference between the ILDs for the processed and the unprocessed clean melody signals $\hat{t}(n)$ and $t(n)$, ΔILD, averaged over all pieces in the database is depicted in Figure 3. ΔILD values close to zero indicate a good preservation of binaural cues in terms of ILDs as no further level differences between the left and the right ear signals are introduced during the spectral complexity reduction. Only for very small numbers of retained components ($\hat{k} < 5$) a considerable deviation from zero can be observed. As mentioned before, the spectral complexity reduction scheme introduces signal distortions and artifacts for small numbers of components, so that this parameter range is not well suitable for application in practice. Furthermore,

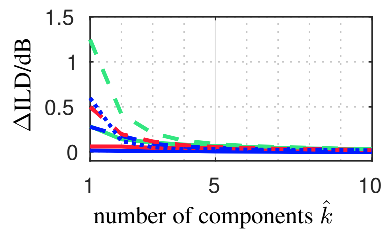

Fig. 3: ILD difference ΔILD between the binaurally processed and the original clean melody ear signals.

the just noticeable differences (JNDs) in bilateral CI listeners range between 1-5 dB input level compared to 1 dB in normal hearing listeners [21]. Even for less than $\hat{k} < 5$ retained components, ΔILD does not even exceed the JND range of NH.

5. CONCLUSIONS

In this paper we investigated the performance of a PCA-based music complexity reduction scheme for CI listeners in a binaural context. The comparison between the experimental results for binaural and separate bilateral processing demonstrate, that with the proposed binaural processing scheme both a further attenuation of the disturbing accompaniment on the side facing the accompaniment source and a reduction of the distortion introduced to the reconstructed melody signal can be achieved. At the same time, the computational costs can be reduced by up to 50 %, as the PCA only needs to be performed once. In a practical application, the amount of data exchanged by interacting bilateral CIs can be reduced as only the low-rank PCA basis matrices for the current segment would be transmitted instead of high resolution audio data. In addition it could be demonstrated that ILDs, which represent the chiefly used binaural cues for CI users, are hardly harmed by the proposed processing scheme. The spectral reduction scheme based on PCA up to now relies on attributes of music signals: in the chamber music excerpts used for this evaluation, a leading voice and an accompaniment can be clearly distinguished. Hence, in future works the application to more general music signals, also containing percussive elements, will be investigated. Furthermore, the influence of additional room reverberation will be taken into account. As hearing is highly subjective, our favorable experimental results will be eventually validated by listening experiments with CI users.

6. REFERENCES

[1] J. Wouters, H. J. McDermott, and T. Francart, "Sound coding in cochlear implants: From electric pulses to hearing," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 67–80, Mar. 2015.

[2] H. J. McDermott, "Music perception with cochlear implants: A review," *Trends in Amplification*, vol. 8, no. 2, pp. 49–82, Jun. 2004.

[3] C. J. Limb and A. T. Roy, "Technological, biological, and acoustical constraints to music perception in cochlear implant users," *Hearing Research*, vol. 308, pp. 13–26, Feb. 2014.

[4] O. Stakhovskaya, D. Sridhar, B. H. Bonham, and P. A. Leake, "Frequency map for the human cochlear spiral ganglion: Implications for cochlear implants," *Journal of the Association for Research in Otolaryngology*, vol. 8, no. 2, pp. 220–233, Apr. 2007.

[5] K. Gfeller, A. Christ, K. John, S. Witt, and M. Mehr, "The effects of familiarity and complexity on appraisal of complex songs by cochlear implant recipients and normal hearing adults," *Journal of Music Therapy*, vol. 40, no. 2, pp. 78–112, 2003.

[6] W. Buyens, B. van Dijk, M. Moonen, and J. Wouters, "Music mixing preferences of cochlear implant recipients: A pilot study," *International Journal of Audiology*, vol. 53, no. 5, pp. 294–301, May 2014.

[7] W. Buyens, B. van Dijk, J. Wouters, and M. Moonen, "A harmonic/percussive sound separation based music pre-processing scheme for cochlear implant users," in *21st European Signal Processing Conference (EUSIPCO 2013)*. IEEE, 2013, pp. 1–5.

[8] ——, "A stereo music preprocessing scheme for cochlear implant users," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 10, pp. 2434–2442, Oct. 2015.

[9] G. D. Kohlberg, D. M. Mancuso, D. A. Chari, and A. K. Lalwani, "Music engineering as a novel strategy for enhancing music enjoyment in the cochlear implant recipient," *Behavioural Neurology*, vol. 2015, pp. 1–7, 2015.

[10] J. Pons, J. Janer, T. Rode, and W. Nogueira, "Remixing music using source separation algorithms to improve the musical experience of cochlear implant users," *The Journal of the Acoustical Society of America*, vol. 140, no. 6, pp. 4338–4349, 2016.

[11] A. Nagathil, C. Weihs, and R. Martin, "Spectral complexity reduction of music signals for mitigating effects of cochlear hearing loss," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 445–458, 2016.

[12] I. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 2002.

[13] A. Nagathil, C. Weihs, K. Neumann, and R. Martin, "Spectral complexity reduction of music signals based on frequency-domain reduced-rank approximations: An evaluation with cochlear implant listeners," *The Journal of the Acoustical Society of America*, vol. 142, no. 3, pp. 1219–1228, Sep. 2017.

[14] J. S. Nemer, G. D. Kohlberg, D. M. Mancuso, B. M. Griffin, M. V. Certo, S. Y. Chen, M. B. Chun, J. B. Spitzer, and A. K. Lalwani, "Reduction of the harmonic series influences musical enjoyment with cochlear implants," *Otology & Neurotology*, vol. 38, no. 1, pp. 31–37, Jan. 2017.

[15] W. Buyens, B. van Dijk, M. Moonen, and J. Wouters, "Evaluation of a stereo music preprocessing scheme for cochlear implant users," *Journal of the American Academy of Audiology*, 2017.

[16] J. Brown, "Calculation of a Constant Q Spectral Transform," *Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 425–434, Jan. 1991.

[17] A. Nagathil and R. Martin, "Optimal signal reconstruction from a constant-Q spectrum," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 349–352.

[18] P. Senn, M. Kompis, M. Vischer, and R. Haeusler, "Minimum audible angle, just noticeable interaural differences and speech intelligibility with bilateral cochlear implants using clinical speech processors," *Audiology and Neurotology*, vol. 10, no. 6, pp. 342–352, Oct. 2005.

[19] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, p. 298605, 2009.

[20] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.

[21] J. Wouters, S. Doclo, R. Koning, and T. Francart, "Sound processing for better coding of monaural and binaural cues in auditory prostheses," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 1986–1997, Sep. 2013.

[22] J. Blauert, Ed., *The Technology of Binaural Listening*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.