Spectral Complexity Reduction of Music Signals for Cochlear Implant Users based on Subspace Tracking

Johannes Gauer
Institute of Communication Acoustics
Ruhr-Universität Bochum
Bochum, Germany
johannes.gauer@rub.de

Ekaterina Krymova
Faculty of Mathematics
Universität Duisburg-Essen
Essen, Germany
ekaterina.krymova@uni-due.de

Denis Belomestny
Faculty of Mathematics
Universität Duisburg-Essen
Essen, Germany
denis.belomestny@uni-due.de

Rainer Martin
Institute of Communication Acoustics
Ruhr-Universität Bochum
Bochum, Germany
rainer.martin@rub.de

***Abstract*—Spectral complexity reduction can be used to emphasize the leading voice or melody and attenuate the competing accompaniment of music pieces. This method is known to facilitate music perception in cochlear implant (CI) users as spectrally less complex signals are perceived as being more pleasant. In this paper we investigate a method to obtain a reduced-rank approximation for the desired complexity reduction that extends the established projection approximation subspace tracking methods (PAST, CPAST) with an additional sparsity constraint. We evaluate our method with the instrumental SIR and SAR measures as well as an auditory distortion measure (ADR) on a database of 110 classical chamber music pieces. While the resulting signal quality is found to be comparable to existing methods the iterative structure and the reduced computational complexity of our method make it suitable for real-time and low-latency on-line applications.**

***Index Terms*—Subspace Tracking, Music Signal Processing, Cochlear Implants, Sparse Eigenvectors**

I. Introduction

Cochlear implants (CI) have led to a remarkable improvement of speech intelligibility in severely hearing impaired or deaf people: more than 300,000 patients have been implanted by now [1]. However, the perception of music signals remains difficult because of several technical and physiological restrictions: the implants can only transmit limited information on pitch, timbre and dynamics of music signals which is caused amongst others by the restricted number of electrodes (≤ 22) and place-pitch mismatches at the interface to the hearing nerve [2], [3]. Thus, CI listeners report unnatural and distorted hearing impressions when listening to music. They frequently prefer easier accessible music genres like country or pop music over e.g. classical music [4], [5]. Also, it has been shown that CI users appreciate an emphasis on vocals, bass and drums and an attenuation of the accompaniment [6].

Recently several studies have been published that investigate different approaches to improve the music enjoyment of CI users by reducing the complexity of music signals. In [7] CI listeners rated recordings of music pieces with a smaller number of instruments more enjoyable than the original versions. Hence, in a first group of methods, complexity reduction is achieved by reducing the number of competing musical voices or instruments, e.g., by manually re-engineering multi-track recordings. However, these separated recordings are usually not available. Therefore, also (blind) source separation techniques like non-negative matrix factorization (NMF) [8] or deep learning [9] can be used to decompose the signals and remix them with the desired weights. In another approach, the authors make use of a harmonic/percussive sound separation (HPSS) method to emphasize drums and other strong rhythmic elements from the percussive portion of a signal and thereby de-emphasize its harmonic components [10].

The spectral complexity of a music signal cannot only be reduced by discarding or attenuating particular voices or instruments but also by modifying the overtone series of individual notes directly. In [11] multiple versions of a monophonic piece were created where the harmonic series of each tone was reduced to 5 different degrees using custom-fit low-pass filters. Among those signals CI and normal-hearing (NH) listeners with CI simulation rated the version containing only the fundamental frequency F_0 of the melody as being most pleasant.

While source separation methods preserve the original full spectrum of each separated voice, dimensionality reduction techniques lead to a spectral complexity reduction as they affect both the mixture of voices and the harmonic series of individual voices. These techniques are based on the assumption that the most prominent elements of the signal spectrum correspond to strong partial tones of the melody or leading voice. In a music piece where a predominant leading voice or melody is accompanied by one or several instruments as, e.g., in classical chamber music, these most prominent elements can be identified by principal component analysis (PCA) [12]. Preserving them and discarding less prominent spectral components leads both to a reduction of the overtones and to an attenuation of the accompaniment [13]. Listening experiments with CI users showed significant preference ratings for these spectrally simplified signals in comparison to the unprocessed versions and to score-informed and NMF-based approaches [14]. Exploiting binaural signals can further improve the complexity reduction outcome [15].

Subspace tracking methods like the projection approxima-

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

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

tion subspace tracking (PAST) [16] have been employed in various signal processing applications such as data compression, image separation or direction-of-arrival estimation. They are based on the assumption that the clean signal and the interference can be regarded as separate subspaces. Therefore, those methods are also a promising means to separate the preponderant spectral components like the leading voice of a music signal from an undesired interference like the broadband accompaniment with reduced computational effort.

In this work we present a preprocessing scheme for a spectral complexity reduction of music signals which is based on the well-established and computationally less complex PAST method [16] to track the desired spectral components instead of PCA [12]. Starting from the enhanced constrained projection approximation subspace tracking (CPAST) method [17] which constraints the signal subspace to be spanned by an orthonormal basis, we utilize a thresholding method to enforce the sparsity of the obtained spectral components.

The remainder of this paper is organized as follows: In Section II we will first review both the PCA-based spectral complexity reduction of music signals and the subspace tracking methods PAST and CPAST. Then we present the application of the sparse constrained projection approximation subspace tracking (SCPAST) method [18] to promote spectral sparsity. In Sections III and IV the experimental setup and the experimental results are presented and discussed. Conclusions are drawn in Section V.

II. Spectral Complexity Reduction using subspace tracking methods

A. Spectral Analysis

In order to represent music signals we consider a model where the discrete-time signal $s(n) = t(n) + i(n),\ n \in \mathbb{N}^0$ is a mixture of the target signal $t(n)$ containing the leading voice and the interfering signal $i(n)$ containing the accompaniment of a music piece. We split the signal into L overlapping segments $s(n,l) = s(n+lR)$ of length N with segment index $l \in \{1,\ldots,L\}$ and segment shift R. Any appropriate spectral transform like the discrete Fourier transform (DFT) or the constant-Q transform (CQT) can be employed to compute a short-time spectrogram representation $\mathbf{x}(l)$ of $s(n,l)$.

The harmonic relations of the spectral components have an essential relevance for music signals. Therefore we choose the CQT [19], as it features a frequency analysis grid $f_\kappa = f_{\min} \cdot 2^{\frac{\kappa}{12b}}$ which perfectly matches the geometrically spaced frequency grid of the scales that usually underly western music. The parameters $\kappa \in \{1,\ldots,K\}$ and $b \in \mathbb{N}$ denote the frequency index and the spectral resolution in terms of spectral bins per semitone, respectively.

B. Spectral Complexity Reduction

The spectral complexity reduction method is based on the assumption that the spectrum of a music piece shows the strong partial tones of its melody or leading voice as its most prominent components. These components need to be preserved while other less prominent components are expected to carry information about the accompaniment and thus will be dropped. To identify these elements the authors in [13] applied principal component analysis (PCA) to blocks $\mathbf{U} = [\mathbf{x}(l_1),\ldots,\mathbf{x}(l_M)] \in \mathbb{C}^{K\times M}$ of M short-time signal spectrum segments $\mathbf{x}(l_m)$. Solving the eigenvalue problem $\mathbf{U}\mathbf{U}^H\mathbf{w}_k = \lambda_k\mathbf{w}_k$ leads to an orthonormal set of base vectors $\mathbf{w}_k$ which are sorted in descending order of their corresponding eigenvalues, i.e. $\lambda_1 \geq \ldots \geq \lambda_k \geq \ldots \geq \lambda_K$. Hence, the first eigenvectors carry most of the overall variance and thus represent the most prominent spectral bands of each segment. To reduce the spectral complexity, only a selected number $\hat{k}$ of leading base vectors is retained while those for $\hat{k} < k \leq K$ are dismissed.

To perform PCA, the computationally expensive eigenvalue problem needs to be solved for each block of signal segments and thus in a repeated fashion. Therefore, instead of performing PCA on each block we propose an iterative subspace tracking method that consists of the following steps: First, we iteratively estimate the spectral covariance matrix. We then compute an orthogonal estimator of the subspace spanned by the first $\hat{k}$ leading eigenvectors using the CPAST method. In addition, with the proposed version of the SCPAST method, we apply a threshold prior to the orthogonalization step to further enforce the (spectral) sparsity of the eigenvectors. Finally, we compute the reduced-rank approximation.

C. Constrained Subspace Tracking (CPAST)

We assume that the spectral covariance matrix $\mathbf{C}(l) \in \mathbb{C}^{K\times K}$ only varies slowly in time. Hence for $\mathbf{C}(l)$ we use the recursive estimator [16]

$$\widehat{\mathbf{C}}(l) = \sum_{u=0}^{l} \gamma^{l-u}\mathbf{x}(u)\mathbf{x}^H(u) = \gamma\widehat{\mathbf{C}}(l-1) + \mathbf{x}(l)\mathbf{x}^H(l) \quad (1)$$

with the so-called memory parameter $0 \leq \gamma \leq 1$ and $\widehat{\mathbf{C}}(l) = \mathbf{0}$ for $l < 0$. As subspace tracking methods like PAST [16] usually aim to estimate the eigen subspace spanned by the leading eigenvectors of the covariance matrix $\mathbf{C}(l)$ we apply the following singular value decomposition (SVD)

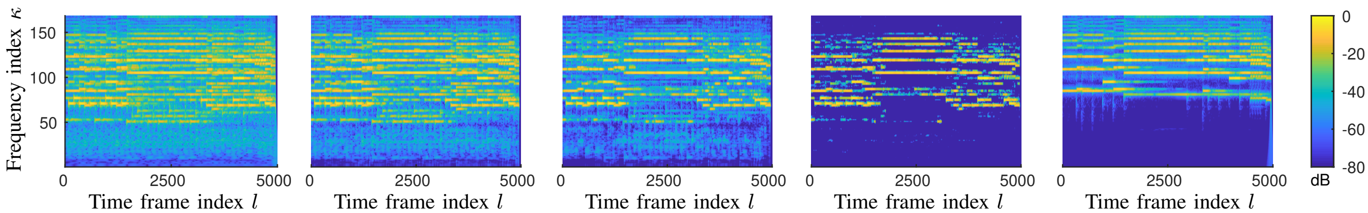
$$\widehat{\mathbf{C}}(l) = \mathbf{V}(l)\mathbf{\Lambda}_d(l)\mathbf{V}^H(l) \quad (2)$$

to the estimated covariance matrix $\widehat{\mathbf{C}}(l)$ which in general has rank $d \leq K$. The matrices $\mathbf{V}(l) = [\mathbf{v}_1(l), \mathbf{v}_2(l), \ldots, \mathbf{v}_d(l)] \in \mathbb{C}^{K\times d}$ and $\mathbf{\Lambda}_d(l) = \mathrm{diag}\{\lambda_1(l), \lambda_2(l), \ldots \lambda_d(l)\} \in \mathbb{C}^{d\times d}$ would then contain the orthonormal eigenvectors and the eigenvalues of $\widehat{\mathbf{C}}(l)$ respectively. For spectral complexity reduction we only retain the first $\hat{k} \leq d \leq K$ eigenvectors.

The CPAST method [17] however is a modification of the PAST algorithm that ensures the orthonormality of the $K \times \hat{k}$ sized estimator $\widehat{\mathbf{V}}(l)$ of the eigenvector matrix $\mathbf{V}(l)$. Using the iteratively estimated covariance matrix $\widehat{\mathbf{C}}(l)$ and some initial approximation $\widehat{\mathbf{V}}(0) = \widehat{\mathbf{V}}^0$, the CPAST method consists of two steps:

Multiplication step: we compute the $K \times \hat{k}$ matrix

$$\widehat{\mathbf{G}}(l) = \widehat{\mathbf{C}}(l)\widehat{\mathbf{V}}(l-1). \quad (3)$$



(a) Mixture, original
(b) Mixture, CPAST, $\hat{k} = 13$, $q_{\text{thr}} = 0.0$
(c) Mixture, CPAST, $\hat{k} = 5$, $q_{\text{thr}} = 0.0$
(d) Mixture, SCPAST, $\hat{k} = 5$, $q_{\text{thr}} = 0.90$
(e) Melody, original

Figure 1: CQT spectrograms of mixture signal $s(n)$ and melody signals $t(n)$ for different parameter settings. Note that SCPAST delivers a sparse representation of the most relevant harmonics of the melody voice $t(n)$.

Orthogonalization step: update the estimator $\widehat{\mathbf{V}}(l)$ of the matrix $\mathbf{V}(l)$ containing $\hat{k}$ leading eigenvectors according to

$$\widehat{\mathbf{V}}(l) = \widehat{\mathbf{G}}(l)[\widehat{\mathbf{G}}^H(l)\widehat{\mathbf{G}}(l)]^{-1/2}. \tag{4}$$

Note that the CPAST method may be regarded as an on-line version of the classical orthogonal iteration method [20].

D. Sparse Constrained Subspace Tracking (SCPAST)

The spectra of music signals and especially their first eigenvectors exhibit a rather sparse structure. Therefore we suggest to impose sparsity assumptions on the columns of the eigenvector matrix. Inspired by [21], the SCPAST method [18] introduces an additional thresholding step which further enforces this sparsity by suppressing small components of the estimated eigenvectors. It is implemented as follows:

Multiplication step: we compute the $K \times \hat{k}$ matrix

$$\widehat{\mathbf{G}}(l) = \widehat{\mathbf{C}}(l)\widehat{\mathbf{V}}(l-1) = \left[\widehat{\mathbf{g}}_1, \widehat{\mathbf{g}}_2, \ldots, \widehat{\mathbf{g}}_{\hat{k}}\right]. \tag{5}$$

Thresholding step: for each column vector $\widehat{\mathbf{g}}_i$ of $\widehat{\mathbf{G}}(l)$ we determine a signal-dependent thresholding parameter $\widehat{\beta}_i > 0$ and apply a threshold function $g(x, \beta)$ to the respective vector elements $\widehat{g}_{ji}$. Possible thresholding methods are hard (g_H) and soft thresholding (g_S) [21]:

$$g_H(x,\beta) = \begin{cases} 0, |x| \le \beta \\ x, |x| > \beta \end{cases}; \quad g_S(x,\beta) = \begin{cases} 0, & |x| \le \beta \\ x-\beta, & x > \beta \\ x+\beta, & x < -\beta \end{cases}$$

We compute $\widehat{\beta}_i$ such that all entries of the respective column vector $\widehat{\mathbf{g}}_i$ with a magnitude below the particular quantile, defined by the level $q_{\text{thr}} \in [0,1]$, are set to zero: $\widehat{\beta}_i = \hat{g}_{i(\lfloor q_{thr} \cdot K \rfloor)}$, where $\hat{g}_{i(k)}$ is the k-th largest (in absolute value) component of the vector $\hat{g}_i$. Hence, with this thresholding parameter $\widehat{\beta}_i$ we obtain the thresholded matrix

$$\widehat{\mathbf{G}}_{\text{thr}}(l) = \left[\widehat{\mathbf{g}}_{\text{thr},1}, \ldots, \widehat{\mathbf{g}}_{\text{thr},\hat{k}}\right], \ \widehat{\mathbf{g}}_{\text{thr},i} = g(\widehat{\mathbf{g}}_i, \widehat{\beta}_i). \tag{6}$$

For $q_{\text{thr}} = 0$ we fall back to CPAST as a baseline method for SCPAST as in this case $\widehat{\beta}_i$ is also set to 0 and thus thresholding is disabled.

Orthogonalization step: update $\widehat{\mathbf{V}}(l)$ according to

$$\widehat{\mathbf{V}}(l) = \widehat{\mathbf{G}}_{\text{thr}}(l)[\widehat{\mathbf{G}}_{\text{thr}}^H(l)\widehat{\mathbf{G}}_{\text{thr}}(l)]^{-1/2}. \tag{7}$$

Note that introducing the thresholding step after the multiplication step leads to sparse estimators (see [21], [18]). We gain the reduced spectrum by projecting the original spectrum to the thresholded estimates of its rank-reduced eigen subspace

$$\widehat{\mathbf{x}}(l) = \left(\widehat{\mathbf{V}}(l)\widehat{\mathbf{V}}^H(l)\right)\mathbf{x}(l) \tag{8}$$

and finally apply the inverse CQT [22] to transform the simplified CQT spectra $\widehat{\mathbf{x}}(l)$ back to time domain and reconstruct the resulting signal $\hat{s}_{\hat{k}}(n)$ using the overlap-add method.

A detailed derivation and a convergence analysis of the SCPAST method can be found in [18]. Instead of initializing the estimated eigenvector matrix $\widehat{\mathbf{V}}^0$ using values from the first l_0 observations as proposed in [18], we set $l_0 = 0$ and obtain $\widehat{\mathbf{V}}^0 = [\mathbf{I}_{\hat{k}}, \mathbf{0}_{\hat{k},K-\hat{k}}]^T \in \mathbb{R}^{K \times \hat{k}}$ as an initial estimate of the eigenvector matrix $\widehat{\mathbf{V}}$. Then (3) and (5) each yield a projection matrix $\widehat{\mathbf{G}}(0)$ which is formed by the first $\hat{k}$ column vectors of the estimated covariance matrix $\widehat{\mathbf{C}}(0)$ of the first signal frame.

In the proposed procedure the thresholding parameter has to be fixed in advance. A fully adaptive procedure would include data-driven estimation of the quantile level e.g. using covariance and noise level estimation [23], [24]. Although (4) and (7) comprise inversion operations, the resulting computational effort is manageable as in our complexity reduction application only small matrices ($\hat{k} \times \hat{k}$) are involved. Due to its iterative structure and in contrast to the PCA-based spectral complexity reduction method, the proposed method is suitable for low-latency on-line processing of music signals and has been succesfully implemented in real-time on a standard PC.

III. Experimental Setup

To evaluate the proposed spectral complexity reduction method we process 110 excerpts, each with a length of 10 s, from a database of classical chamber music MIDI files that each consist of a leading and accompanying voices [13]. The audio signals were created using high quality samples from the *Native Instruments Komplete*[1] package, that are based on recordings of real instruments. The melody and accompaniment signals $t(n)$ and $i(n)$ of each piece were generated individually and are normalized to equal signal energy so they add up to $s(n)$ at 0 dB input SIR. This leads to both a realistic volume ratio of both voices and a simple interpretation of the source separation measures.

[1]Komplete 9, 2013, https://www.native-instruments.com

A. Signal Quality Measures

We employ established signal quality measures such as signal-to-interferer-ratio (SIR) and signal-to-artifacts-ratio (SAR) [25] to evaluate the accompaniment attenuation and the distortion of the leading voice, respectively. To this end we processed the mixture signals $\hat{s}_{\hat{k}}(n)$ as well as the clean melody and accompaniment signals $t(n)$ and $i(n)$ by projecting their respective spectra onto the rank-reduced eigen subspace estimates of the mixture. This projection is a linear operation. In this way, besides the reduced mixture signals $\widehat{s}(n)$ we also obtain the reduced melody and accompaniment signals $\widehat{t}(n)$ and $\widehat{i}(n)$ and do not need to estimate them in order to analyze interference suppression and signal distortion introduced by processing artifacts. Thus, the SIR measure is defined as

$$\mathrm{SIR} = 10\log_{10}\left(\frac{\sum_n t^2(n)}{\sum_n \hat{i}^2(n)}\right). \tag{9}$$

With a slight deviation from [25] we define the signal-to-artifacts-ratio as the logarithm of the energy ratio between the clean melody signal and residual error signal after processing:

$$\mathrm{SAR} = 10\log_{10}\left(\frac{\sum_n t^2(n)}{\sum_n (\hat{t}(n) - t(n))^2}\right). \tag{10}$$

B. Auditory Distortion Measure

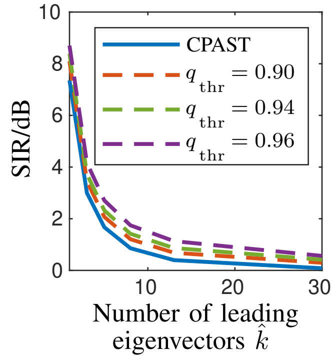
The aforementioned source separation measures describe the attenuation of the accompaniment and the emphasis on the leading voice, which had been shown to facilitate the access to music for CI users [6], [7], [8], [9], [11]. However, these measures do not account for auditory distortions which occur due to the reduced frequency selectivity of electric stimulation. Therefore, we also evaluate our method using the auditory distortion ratio (ADR) measure [13], [26]. This measure is based on a spectral smearing technique which mimicks broadened auditory filters [27]. Thus, the ADR measure is defined as

$$\mathrm{ADR}(k) = 10\log\left(\frac{\sum_n [s(n) - \tilde{s}(n)]^2}{\sum_n \left[\hat{s}_{\hat{k}}(n) - \tilde{s}_{\hat{k}}(n)\right]^2}\right)\mathrm{dB}. \tag{11}$$

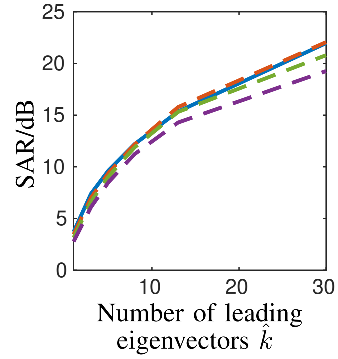
It compares the energy ratio between the errors of the original and the spectrally smeared versions $s(n)$ and $\tilde{s}(n)$ of the unprocessed signal in the numerator and the original and the spectrally smeared versions $\hat{s}_{\hat{k}}(n)$ and $\tilde{s}_{\hat{k}}(n)$ of the processed signal in the denominator, respectively. Positive ADR values indicate an improvement in auditory distortion in terms of spectral smearing. As the ADR measures the distortion of relatively weak higher-order harmonics, the resulting values in dB are rather small, e.g. in comparison to SIR and SAR values.

IV. Experimental Results

Figure 1 shows CQT spectrograms of the original and processed versions with different parameter settings for one file from the chamber music database: with a decreasing number of leading eigenvectors $\hat{k}$ the spectrograms (1b-1c) become more sparse than the original (1a). This sparsity is further increased with the thresholding quantile level q_{thr}. For $q_{\mathrm{thr}} = 0.90$ (1d) it mostly features the strongest harmonics of the leading voice (1e).



(a) SIR



(b) SAR

Figure 2: Signal-to-interference ratio (SIR) and signal-to-artifacts ratio (SAR) for memory parameter $\gamma = 0.95$ and soft thresholding, averaged over 110 pieces.

A. Signal Quality Measures

In Figure 2a the signal-to-interference ratio (SIR) averaged over all 110 pieces in the database is depicted for a selection of four thresholding quantile levels q_{thr} with soft thresholding and a memory parameter of $\gamma = 0.95$. As higher SIR values denote a higher attenuation of the accompaniment, we obtain increased attenuation for small numbers of leading eigenvectors $\hat{k}$ as desired. It also becomes obvious that starting from $q_{\mathrm{thr}} = 0$ (CPAST) higher thresholding levels lead to a further improvement in the accompaniment attenuation up to 9 dB. However, our initial experiments also showed that with a very high thresholding ($q_{\mathrm{thr}} \geq 0.97$) a reconstruction of the reduced spectra is no longer achievable for most of the processed signals as under this condition the resulting spectra become too sparse.

The diagram in Figure 2b shows the mean signal-to-artifacts-ratio (SAR) for all 110 pieces in the database depending on the number of leading eigenvectors $\hat{k}$ and multiple threshold levels q_{thr}, again with a memory parameter of $\gamma = 0.95$ and soft thresholding. A stronger complexity reduction with smaller $\hat{k}$ implies a higher amount of artifacts as the SAR values decline towards SAR $\approx$ 3 dB. For increasing thresholding levels the improved accompaniment attenuation comes at the expense of lower SAR values and thus distortion of the melody signal. In the practically relevant parameter range of $3 \leq \hat{k} \leq 15$ only minor changes in SAR between different thresholding settings are observed.

B. Auditory Distortion Results

Figure 3 depicts the auditory distortion ratios (ADR) for the PCA, CPAST and SCPAST methods for varying numbers of leading eigenvectors $\hat{k}$ grouped by the thresholding level q_{thr}. Smaller values of $\hat{k}$ and higher thresholds q_{thr} each lead to higher ADR values as they both reduce the spectral complexity

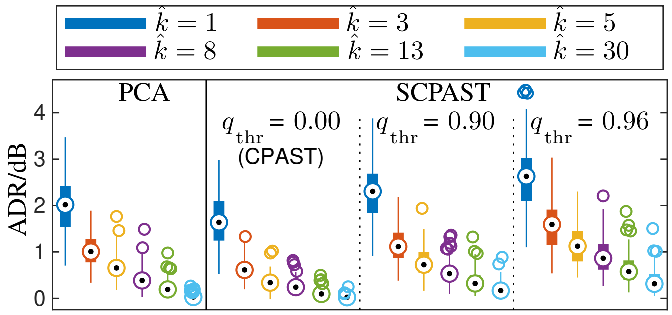


Figure 3: Auditory distortion ratio (ADR) for PCA and SCPAST methods averaged over 110 files. Colors indicate equal numbers of leading eigenvectors $\hat{k}$.

of a signal. In comparison we find that the unconstrained CPAST method ($q_{\text{thr}} = 0$) leads to slightly lower average ADR results than the PCA method. However, the sparsity-constrained SCPAST method ($q_{\text{thr}} \geq 0.9$) outperforms CPAST and also PCA in terms of the ADR measure. Listening experiments for the PCA method revealed the range of $3 \leq \hat{k} \leq 15$ to be relevant for practical use [14]. With SCPAST we gain increases in median ADR of $0.4 \ldots 0.6$ dB relative to PCA and of $0.5 \ldots 1.0$ dB relative to CPAST. These outcomes indicate an improvement in auditory distortion resulting from the additionally enforced sparseness both in comparison with the unprocessed signals (ADR = 0 dB) and with the PCA and CPAST methods.

V. Conclusions

In this paper we presented a new spectral complexity reduction scheme for music signals based on subspace tracking methods and adaptive thresholding which serves to improve music enjoyment in CI users. We demonstrated that we can achieve results comparable to previously proposed PCA-based methods both in accompaniment attenuation and sound quality. In addition, instrumental measures predict an improvement in auditory distortion due to the introduced spectral sparsity constraint. The new SCPAST-based method relies on an iterative estimation of the spectral covariance matrices and demands less complex computations. Therefore, in contrast to the PCA method, it allows real-time and low-latency on-line processing. Regarding the promising outcomes from the instrumental measures, in ongoing works we will validate the performance of the SCPAST-based spectral complexity reduction in listening experiments with CI listeners and with additional musical genres comprising also vocals and percussion instruments.

References

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

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

[3] O. Stakhovskaya, D. Sridhar, B. H. Bonham, and P. A. Leake, "Frequency map for the human cochlear spiral ganglion: Implications for cochlear implants," *J. Assoc. Res. Otolaryngol.*, vol. 8, no. 2, pp. 220–233, 2007.

[4] 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," *J. Music Ther.*, vol. 40, no. 2, pp. 78–112, 2003.

[5] V. Looi and J. She, "Music perception of cochlear implant users: A questionnaire, and its implications for a music training program," *Int. J. Audiol.*, vol. 49, no. 2, pp. 116–128, 2010.

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

[7] 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," *Behav. Neurol.*, vol. 2015, pp. 1–7, 2015.

[8] J. Pons, J. Janer, T. Rode, and W. Nogueira, "Remixing music using source separation algorithms to improve the musical experience of cochlear implant users," *J. Acoust. Soc. Am.*, vol. 140, no. 6, pp. 4338–4349, 2016.

[9] T. Gajęcki and W. Nogueira, "Deep learning models to remix music for cochlear implant users," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3602–3615, 2018.

[10] 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 *Proc. European Signal Process. Conf. (EUSIPCO)*, 2013, pp. 1–5.

[11] 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," *Otol. Neurotol.*, vol. 38, no. 1, pp. 31–37, 2017.

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

[13] A. Nagathil, C. Weihs, and R. Martin, "Spectral Complexity Reduction of Music Signals for Mitigating Effects of Cochlear Hearing Loss," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 3, pp. 445–458, 2016.

[14] 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," *J. Acoust. Soc. Am.*, vol. 142, no. 3, pp. 1219–1228, 2017.

[15] J. Gauer, A. Nagathil, and R. Martin, "Binaural Spectral Complexity Reduction of Music Signals for Cochlear Implant Listeners," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. Calgary, AB, Canada: IEEE, 2018, pp. 251–255.

[16] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, 1995.

[17] A. Valizadeh and M. Karimi, "Subspace tracking based on the constrained projection approximation approach," in *Proc. European Signal Process. Conf. (EUSIPCO)*, 2007, pp. 105–109.

[18] D. Belomestny and E. Krymova, "Sparse constrained projection approximation subspace tracking," *ArXiv181009298 Stat*, 2018.

[19] J. C. Brown, "Calculation of a constant Q spectral transform," *J. Acoust. Soc. Am.*, vol. 89, no. 1, pp. 425–434, 1991.

[20] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Johns Hopkins University Press, 1996.

[21] Z. Ma, "Sparse principal component analysis and iterative thresholding," *Ann. Stat.*, vol. 41, no. 2, pp. 772–801, 2013.

[22] A. Nagathil and R. Martin, "Optimal signal reconstruction from a constant-Q spectrum," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2012, pp. 349–352.

[23] D. Belomestny, M. Trabs, and A. B. Tsybakov, "Sparse covariance matrix estimation in high-dimensional deconvolution," *ArXiv171010870 Math Stat*, 2017.

[24] G. K. Golubev and E. A. Krymova, "Noise Level Estimation in High-Dimensional Linear Models," *Probl. Inf. Transm.*, vol. 54, no. 4, pp. 351–371, 2018.

[25] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.

[26] W. Nogueira, A. Nagathil, and R. Martin, "Making Music More Accessible for Cochlear Implant Listeners: Recent Developments," *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 115–127, 2019.

[27] B. C. J. Moore, B. R. Glasberg, and A. Simpson, "Evaluation of a method of simulating reduced frequency selectivity," *J. Acoust. Soc. Am.*, vol. 91, no. 6, pp. 3402–3423, 1992.