A FEATURE-BASED LINEAR REGRESSION MODEL FOR PREDICTING PERCEPTUAL RATINGS OF MUSIC BY COCHLEAR IMPLANT LISTENERS

Anil Nagathil[1], *Jan-Willem Schlattmann*[1], *Katrin Neumann*[2], *and Rainer Martin*[1]

[1]Institute of Communication Acoustics, Ruhr-Universität Bochum, Bochum, Germany
[2]Abt. für Phoniatrie und Pädaudiologie, Klinik für Hals-, Nasen- und Ohrenheilkunde, Kopf- und Halschirurgie, St. Elisabeth-Hospital, Ruhr-Universität Bochum, Germany
email: {firstname.lastname}@rub.de

ABSTRACT

While speech quality and intelligibility prediction methods for normal-hearing and hearing-impaired listeners have found a lot of attention as a cost-saving complement to listening tests, analogous procedures for music signals are still rare. In this paper a method is proposed for predicting perceptual ratings of music as obtained by cochlear implant (CI) listeners. For this purpose a listening test with CI listeners was conducted, who were asked to provide their ratings for music excerpts on different scales. It is shown that principal component regression (PCR) is a suitable tool to model and accurately predict the median ratings of the CI listeners using timbre and pitch related signal features as predictor variables. These features describe signal characteristics such as high-frequency energy, spectral bandwidth and roughness. The proposed prediction model is a first step towards an instrumental evaluation procedure for music processing algorithms in hearing devices.

***Index Terms*—** Music, cochlear implants, principal component analysis, regression, predictive models

1. INTRODUCTION

Cochlear implants (CI) restore the hearing ability of profoundly hearing-impaired or deaf people via an array of electrodes in the cochlea, which stimulate the auditory nerve using trains of signal-controlled electrical pulses. However, since the spectral resolution of the encoded acoustic signals is poor, CI listeners are not able to achieve the same degree of frequency selectivity as normal-hearing (NH) listeners. This leads to a distorted perception of pitch and timbre, whereas rhythmic information can generally be recognized well [1, 2, 3]. To improve the quality of music transmission in the presence of cochlear hearing loss, music preprocessing schemes have been proposed recently, which highlight vocals and percussive elements in popular music based on a harmonic/percussive sound separation [4, 5] or reduce the spectral complexity of classical chamber music using dimensionality reduction techniques [6].

In view of the current development of music preprocessing schemes for CIs, evaluation methods are required which can predict the perceived music quality in CI listeners and thus complement time-consuming listening tests. Frequently used measures to assess speech enhancement algorithms include the signal-to-noise ratio (SNR), the signal-to-distortion ratio (SDR), the signal-to-inference ratio (SIR), and the signal-to-artifacts ratio (SAR) [7], which account for different kinds of signal improvements or distortions. Other measures are based on data obtained from listening tests or auditory models and thus are capable of predicting speech and audio quality more precisely, e.g. the perceptual evaluation of speech quality (PESQ) [8], PEMO-Q [9], or the perceptual objective listening quality assessment (POLQA) [10]. Methods which estimate the amount of speech intelligibility are, for instance, the short-time objective intelligibility (STOI) measure [11] or measures based on mutual information [12]. Furthermore, a feature-based prediction model of noise annoyance for hearing-impaired listeners was proposed in [13]. In contrast, music-related evaluation metrics for signal enhancement algorithms in hearing devices have only gained prominence recently. In [14] a measure was proposed which predicts the sound quality of music after being processed by an adaptive feedback canceler in a hearing aid. In [15] a metric was introduced which uses changes in envelope modulation, temporal fine structure, and long-term spectral shape to predict music quality for NH and hearing-impaired listeners. Furthermore, in [16] a method was proposed which estimates musically relevant properties such as note onsets times, pitch, and music instruments from the output of an auditory model which may also include hearing impairments. However, these metrics are designed to evaluate hearing aid algorithms and do not take electric hearing into account. In this work we therefore present an approach for predicting music quality ratings obtained from CI listeners. The method is based on a linear regression model which uses acoustic signal features related to pitch and timbre cues to predict the music quality ratings.

This paper is organized as follows. In Section 2 we describe details about a listening test we conducted with CI listeners. The results obtained in the listening test are analyzed and discussed in Section 3. Using signal-based features describing timbre and pitch characteristics of music, a linear regression model is derived in Section 4 which predicts the median ratings among the CI listeners. The work is concluded in Section 5.

2. LISTENING TEST

2.1. Test participants

Eleven CI listeners (age: 54.1 ± 15.0 years) participated in the listening test. Details about the CI listeners are summarized in Table 1. All listeners were profoundly hearing-impaired on both ears. Except for listener CI-09, all participants used unilateral or bilateral MED-EL implants with the FSP/FS4 or CIS coding strategies. If a hearing aid (HA) was used additionally, the participants were requested to turn it off during the test as we confined our investigation to music perception through electric hearing. At the time of the listening test

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

Table 1. Information about CI listeners who participated in the listening test. All age-related details are counted in years.

	CI-01	CI-02	CI-03	CI-04	CI-05	CI-06	CI-07	CI-08	CI-09	CI-10	CI-11
Age	73	53	48	55	73	55	42	65	50	61	20
Age at HI diagnosis	53	41	38	48	45 - 50	37	20	3	5	36	1
Age at CI implantation	72	52	46	50	71	48	41	62	40	59	4
Hearing device (left)	CI	HA	HA	-	-	HA	-	CI	CI	CI	HA
Hearing device (right)	CI	CI	CI	CI	CI	CI	CI	CI	CI	HA	CI
CI brand	MED-EL	MED-EL	MED-EL	MED-EL	MED-EL	MED-EL	MED-EL	MED-EL	Cochlear	MED-EL	MED-EL
Speech processor (left)	Opus2	-	-	-	-	-	-	Opus2	CP910	Opus2	-
Speech processor (right)	Opus2	Opus2	Opus2	Opus2	Opus2	Opus2	Opus2	Opus2	CP910	-	Opus2
Coding strategy	FSP	FS4	FS4	FSP	FSP/FS4	FSP	FS4	FS4	ACE	FSP	CIS

the participants had been accustomed to the CI for at least one year and were able to understand speech in quiet environments. Note, that listeners CI-08, CI-09, and CI-11 were diagnosed with a profound hearing loss during infancy.

2.2. Selection of music stimuli

For our listening test we selected 12 music excerpts of stereo CD recordings, which similarly as in studies by [17, 18] encompassed different harmonic complexity levels ranging from melodies played by single instruments (electric bass, cello, brass, acoustics/electric guitar, piano, woodwind, saxophone, bagpipe) to orchestral music (strings). The passages had a duration of 7s to 15s and included self-contained melody lines with a fixed number of instruments, a constant rhythm, and a uniform style of playing, respectively. Since music perception of CI listeners is facilitated by means of rhythmic cues [2], the set of stimuli also varied in terms of the degree of clearly accentuated note onsets. Furthermore, as high-pitched notes are often perceived as more scattered, i.e. noisy, than low-pitched notes [19], the excerpts also differed in terms of their pitch characteristics. Percussions and vocals were not considered to narrow down the focus in this investigation. Obviously, a higher number of stimuli would provide a higher number of ratings and hence more information on how CI listeners perceive music. However, since we were interested in obtaining ratings on several different scales, a trade-off between the number of the stimuli and the cognitive load for the CI users during the test had to be considered.

2.3. Rating scales and test setup

The participants were asked to rate the stimuli on 9 continuous bipolar ordinal scales which described contrary pairs of sound properties or attitudes towards the stimuli. We used the word pairs 'dull vs. sharp', 'scattered vs. compact', and 'empty vs. full' which were found to describe 88% of timbre variance in [19]. These timbre descriptions were amended by the word pairs 'blurred vs. distinct', 'distorted vs. clean', 'artificial vs. natural', 'unpleasant vs. pleasant', 'hard to follow vs. easy to follow', and 'complex vs. simple'. Similar scales were used for questionnaires with CI users in [20] and [21]. Note, that these scales reflect the degree of music acceptance, where the first word of the word pair can be associated with a negative attitude towards a music piece. In addition, the participants were asked to rate the sound quality on a scale which measures a continuous version of the mean opinion score (MOS).

All scales except for the MOS scale were labeled with the attributes 'slightly', 'quite', and 'extreme' at equidistant scale positions towards each direction, which were represented numerically by ±1, ±2, and ±3, respectively. Note, that negative numbers pointed towards the first word of the word pair. The marks were complemented by a neutral level with its numerical representation 0. On the MOS scale distinct marks corresponding to a bad, poor, fair, good, and an excellent sound quality were provided along with their respective numerical representations 1, 2, 3, 4, and 5. In order to encourage the participants to consider the whole range of the rating scale, the slider bars were extended at both ends as suggested in [22]. However, the ratings were confined by the values ±3 for the first nine scales and by the values 1 and 5 for the sound quality scale.

The stimuli were presented in an audiometric booth via GENELEC 2029B loudspeakers which were positioned in a stereo setup one meter in front of the participants. The participating CI listeners provided their ratings by using slider bars in a graphical user interface. Before starting the actual test session the participants were requested to listen to two representative stimuli in order to set the volume to a comfortable level. Furthermore, they familiarized themselves with the graphical user interface and the test procedure by a training in which they were presented two additional stimuli. During the test there were no repeated trials, i.e. each stimulus was rated only once on each scale. However, the CI listeners were allowed to listen to the stimuli as often as desired while setting the slider bars. The total duration of the test amounted to approximately 30 to 45 minutes.

3. TEST RESULTS

In Figure 1 the aggregated CI listener ratings are visualized separately for each stimulus and each scale by means of boxplots. In addition to the scale names the figure provides a numbering from 1 to 10. Note, that only the ratings of the postlingually deafened listeners CI-01 to CI-10 are included in the boxplots. The ratings of listener CI-11 who was implanted due to a prelingual deafness are shown by black squares. The results show that in most cases listener CI-11 assigned higher ratings to the stimuli than the other CI listeners, which points towards a higher degree of music preference. This can be attributed to the prelingual stage of implantation which can facilitate speech understanding and increase the enjoyment of music with a CI [23, 24].

For the other listeners we observe that the ratings often exhibit a large variance, which is due to the strongly subjective nature of rating music. In addition, the presence of a hearing loss is likely to increase the rating variability [17]. Nevertheless, in particular stimuli 4, 5, and 11 are perceived as more scattered, blurred, distorted, artificial, unpleasant, difficult to follow and complex. These stimuli are played by electric guitars with delay and distortion effects (stimuli 4 and 11) and a bagpipe (stimulus 5), respectively. In contrast, stimuli 1 and 6 exhibit ratings which tend towards the opposite ends of the scales. These stimuli are played by an electric bass and acoustic guitars, respectively, which are characterized by accentuated note onsets or a pronounced rhythm.

To find scales for which the listeners were able to perceive subtle differences between the considered stimuli, we performed a scale-wise analysis of variance (ANOVA) of the ratings at significance

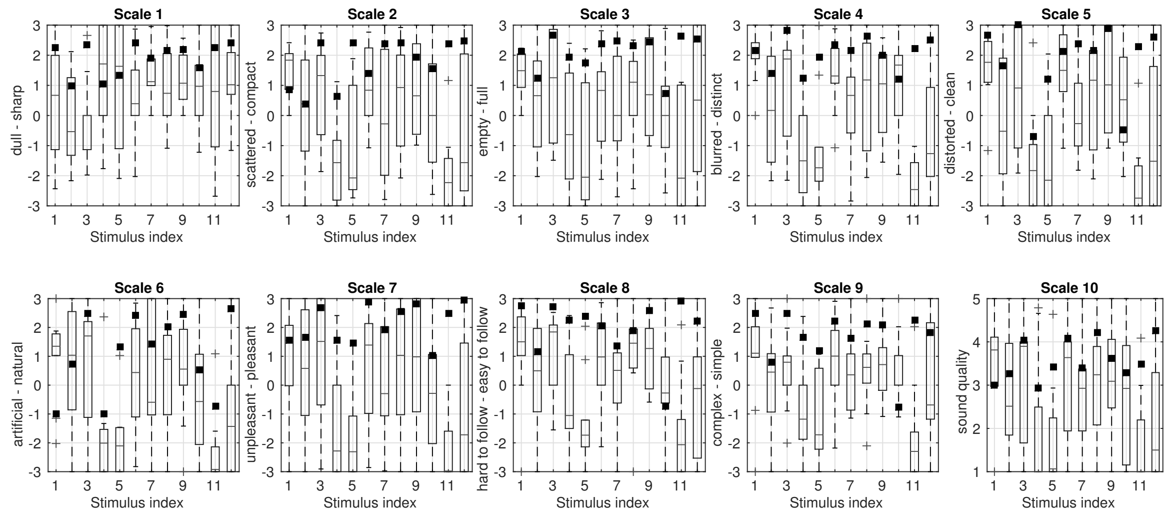


Fig. 1. Boxplots of CI participant ratings vs. stimulus index per scale. The boxplots only include ratings of participants CI-01 to CI-10. The ratings of participant CI-11 (implanted due to prelingual deafness) are depicted by black squares.

Table 2. Numbers of scale-wise significant rating differences for stimulus pairs. In total there were 66 possible pairs.

Scale index	1	2	3	4	5	6	7	8	9	10
# sign. different ratings	-	6	1	10	10	7	5	5	6	-

level $\alpha = 0.05$. It resulted in nonsignificant differences between stimuli ratings for scale 1 (dull vs. sharp) ($p = 0.459$). Significant differences were found for scale 3 (empty vs. full) ($p = 0.01$) and scale 10 (sound quality) ($p = 0.02$), respectively. For the remaining scales the test revealed highly significant rating differences ($p < 0.001$). To identify pairs of stimuli with significant rating differences, a Tukey-Kramer post-hoc test was applied to the ratings on scales 2 to 10, respectively (with $\alpha = 0.05$). The quantity of pairs with significantly different ratings are listed in Table 2. Note, that in total there were 66 possible pairs. Obviously, the highest number of significantly different ratings was obtained on scales 4 and 5. Thus, we can argue that the CI listeners were able to perceive subtle nuances between different stimuli on these scales. In contrast, on scales 1 and 10 the listeners did not provide significantly different ratings for pairs of different stimuli.

4. PREDICTION OF MUSIC RATINGS

Instead of predicting the ratings of individual CI listeners we considered median ratings across the postlingually implanted listeners CI-01 to CI-10. This reduces variations due to external conditions. Listener CI-11 was not considered since electric hearing obviously does not degrade his music perception. Further, we only took the preference-related scales 4 to 10 into account. However, scale 10 was excluded since ratings did not show significant differences for pairs of stimuli.

4.1. Feature extraction

The music signals were resampled at $f_s = 16$ kHz, converted to mono, and segmented into overlapping frames using a frame length of $N = 512$ and a frame shift of $R = 256$. Then, short-term low-level features were extracted from the 12 stimuli using the MIR toolbox [25]. Since studies on CI hearing suggest that timbre and pitch properties are the main influential factors for degraded music perception as opposed to rhythmic cues, we only used features which account for such characteristics. These were the zero-crossing rate, spectral roll-off, high-frequency energy, spectral centroid, spectral spread, spectral skewness, spectral kurtosis, spectral flatness, spectral entropy, roughness, spectral regularity, chromagram, and mel-frequency cepstral coefficients. As these features are computed on a short-term basis, the feature series of a complete stimulus was aggregated by computing its mean and standard deviation. Since these operations, however, ignore the temporal evolution of the feature series, the flux of the feature series was computed (i.e. the difference between feature values of successive temporal segments). The flux series was aggregated by computing its mean and standard deviation as well. Hence, in total $K = 74$ features were obtained.

4.2. Principal component regression model

A regression model was obtained which predicts the scale-wise median ratings of the CI listeners based on the extracted features. To prevent overfitting effects, we applied the leave-one-out strategy such that training sets consisting of $I = 11$ stimuli were selected in a round-robin fashion for obtaining a regression model and the remaining 12th stimuli was used for testing. Note, that by following this strategy each stimulus rating is predicted using a different regression model. However, it ensures that the rating to be predicted and the corresponding feature values are not used in computing the regression coefficients. To avoid an ill-conditioned regression model, we did not perform an ordinary least-squares regression (OLS) but opted for a principal component regression (PCR) model. To this end, the centered feature matrix of the training set, $\mathbf{X} \in \mathbb{R}^{I \times K}$ is mapped onto its principal component (PC) score matrix $\mathbf{S} = \mathbf{XW}$, which leads to an orthogonal feature representation. The columns of the matrix $\mathbf{W} \in \mathbb{R}^{K \times I}$ contain the eigenvectors of the feature covariance matrix $\mathbf{C} \sim \mathbf{X}^{\mathrm{T}}\mathbf{X}$. Note, that typically a high percentage of the total variance in the original features can be preserved in the

Table 3. (a) RMSE values and (b) Pearson correlation coefficients between real and predicted median ratings for different numbers j of retained principal components.

(a)

#PC j	1	2	3	4	5	6	7	8
Scale 4	0.91	1.18	0.68	**0.53**	0.66	0.75	0.81	0.71
Scale 5	1.04	1.11	1.01	0.54	**0.50**	0.82	0.90	0.81
Scale 6	0.65	0.67	**0.47**	0.61	0.48	0.54	0.56	0.63
Scale 7	0.81	0.89	0.69	0.82	**0.68**	0.83	0.85	1.28
Scale 8	0.65	0.61	0.62	**0.55**	0.60	0.80	0.80	1.08
Scale 9	0.60	0.50	0.72	0.52	0.49	0.43	0.43	**0.41**

(b)

#PC j	1	2	3	4	5	6	7	8
Scale 4	0.80	0.68	0.90	**0.94**	0.90	0.88	0.86	0.89
Scale 5	0.72	0.69	0.81	0.94	**0.95**	0.89	0.88	0.92
Scale 6	0.90	0.90	**0.95**	0.93	**0.95**	0.93	0.93	0.93
Scale 7	0.87	0.84	0.92	0.91	**0.93**	0.92	0.92	0.88
Scale 8	0.86	0.88	0.88	**0.90**	0.88	0.78	0.79	0.61
Scale 9	0.84	0.89	0.84	0.92	0.93	0.94	0.94	**0.95**

first j PC scores (with $j < I$), which are contained in the columns of the matrix $\mathbf{S}_j = \mathbf{X}\mathbf{W}_j$. Here, $\mathbf{W}_j \in \mathbb{R}^{K\times j}$ denotes a matrix containing only j eigenvectors. Hence, using a low number of PC scores as predictor variables in an OLS regression model instead of the original features leads to a regularized regression model which solves the problem of feature collinearity. The vector of regression coefficients in the PC space is obtained by $\boldsymbol{\gamma}_j = \left(\mathbf{S}_j^{\mathrm{T}}\mathbf{S}_j\right)^{-1}\mathbf{S}_j^{\mathrm{T}}\mathbf{y}$, where $\mathbf{y} \in \mathbb{R}^{I\times 1}$ is a centered vector containing the median ratings of the training stimuli on a specific scale. These regression coefficients can be applied to the PC scores of the feature vector of the respective test stimulus $\mathbf{x}(i)$, with $i = 1, 2, \ldots, 12$. The corresponding rating prediction $\hat{y}(i)$ is obtained by

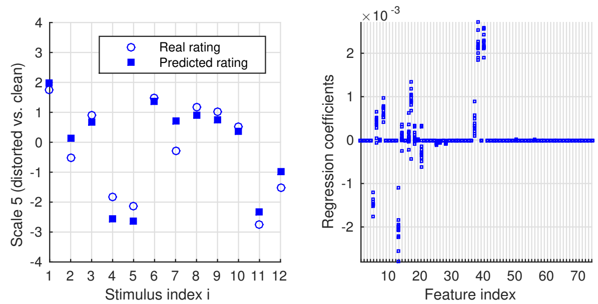
$$\hat{y}(i) = \overline{y} - \overline{\mathbf{x}}^{\mathrm{T}}\mathbf{W}_j\boldsymbol{\gamma}_j + \mathbf{x}^{\mathrm{T}}(i)\mathbf{W}_j\boldsymbol{\gamma}_j, \qquad (1)$$

where $\overline{y}$, $\overline{\mathbf{x}}$, and $\mathbf{x}(i)$ denote the average of the median ratings in the training set, the average across the rows of the feature matrix $\mathbf{X}$, and the feature vector of the test stimulus, respectively.

4.3. Prediction results

We applied PCR in a leave-one-out fashion using the median stimulus ratings for scales 4 to 9 and the feature values extracted from the signals. The number of retained PCs was set to $j = 1, 2, ..., 8$, thereby ensuring that the number of scores based on which the regression coefficients are computed is less than the number of stimuli. The prediction performance was evaluated by means of the root mean squared error (RMSE) and the Pearson correlation coefficient between the real ratings and the predicted ones, respectively. The results of the RMSE and the correlation coefficient are provided in Tables 3(a) and 3(b), respectively, where the lowest RMSE values and the highest correlation values are set in bold. Taking into account both measures we observe that a number of four scores is optimal for scales 4 and 8. For scales 5 and 7 five scores result in the best performance. For scales 6 and 9 the highest degree of prediction is achieved by using three and eight scores, respectively. As an example, in Figure 2(a) the real and predicted median ratings on scale 5 are depicted using the optimal number of five PCs. Similar results are obtained for the other scales using the corresponding optimal number of PCs.

Transforming the vectors of regression coefficients $\boldsymbol{\gamma}_j$ in the principal component space back to the original feature space by $\boldsymbol{\beta}_j = \mathbf{W}_j\boldsymbol{\gamma}_j$ enables us to assess the importance of individual feature dimensions. Therefore, in Figure 2(b) the transformed regression coefficients are visualized for scale 5 using the optimal number of five principal components. Note, that in this case we used all 12 stimuli to obtain a single regression model. This plot shows that most of the regression coefficients are zero or close to zero. However, there are 12 features which differ considerably from zero across all four scales and which therefore contribute most to the prediction of the music ratings. These features are the mean, standard



(a) (b)

Fig. 2. (a) Real median ratings and their predictions for different stimuli on scale 5 using 5 PCs and (b) regressions coefficients in original features space.

deviation, and flux standard deviation of the spectral rolloff, spectral centroid, spectral spread, and roughness feature, respectively. They describe the amount of high-frequency energy in the signal, the frequency region with the highest concentration of spectral energy, the spectral bandwidth, and the degree of dissonance between pairs of spectral peaks, respectively. We found that reducing the feature set to these 12 features and performing PCR with this subset does not degrade the prediction performance. This also applies to the other scales for which the same reduced feature set can be used.

5. CONCLUSIONS

The results demonstrate a high prediction accuracy of the proposed PCR model in combination with the applied cross-validation strategy, yielding RMSE values between 0.41 and 0.68 and Pearson correlation coefficients between 0.9 and 0.95. Moreover, an analysis of the regression coefficients reveals that a reduction of the feature set to a subset of 12 dominating features does not degrade the performance. These features explain the perceptual music ratings of CI listeners in terms of high-frequency energy, spectral bandwidth and roughness of the stimuli. Since the stimuli chosen in this work stem from small, purely instrumental musical setups, the proposed measures are mainly applicable to recordings from similar acoustic scenarios. Hence, the derived regression model may not be suited for music including percussive instruments or vocals since this type of music was excluded in this work. The measures can serve as performance indicators for music processing algorithms in CIs with respect to specific perceptual dimensions.

6. REFERENCES

[1] K. Gfeller, C. Turner, M. Mehr, G. Woodworth, R. Fearn, J. F. Knutson, S. Witt, and J. Stordahl, "Recognition of Familiar Melodies by Adult Cochlear Implant Recipients and Normal-hearing Adults," *Cochlear Implants International*, vol. 3, no. 1, pp. 29–53, 2002.

[2] H. J. McDermott, "Music Perception with Cochlear Implants: A Review," *Trends In Amplification*, vol. 8, no. 2, pp. 49–82, 2004.

[3] V. Looi, H. McDermott, C. McKay, and L. Hickson, "Pitch Discrimination and Melody Recognition by Cochlear Implant Users," *International Congress Series*, vol. 1273, pp. 197 – 200, 2004.

[4] 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 Processing Conference*, 2013.

[5] 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, 2014.

[6] 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.

[7] 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, 2006.

[8] ITU-T Rec. P.862, *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs*, International Telecommunication Union, Geneva, Switzerland, 2001.

[9] R. Huber and B. Kollmeier, "PEMO-Q - A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception," *IEEE Transactions on Acoustics, Speech, and Language Processing*, vol. 14, no. 6, pp. 1902–1911, 2006.

[10] ITU-T Rec. P.863, *Perceptual Objective Listening Quality Assessment (POLQA)*, International Telecommunication Union, Geneva, Switzerland, 2011.

[11] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 7, pp. 2125 – 2136, 2011.

[12] J. Taghia and R. Martin, "Objective Intelligibility Measures Based on Mutual Information for Speech Subjected to Speech Enhancement Processing," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 22, no. 1, pp. 6–16, 2014.

[13] S. Vishnubhotla, J. Xiao, B. Xu, M. McKinney, and T. Zhang, "Annoyance Perception and Modeling for Hearing-impaired Listeners," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 161–164.

[14] A. J. Manders, D. M. Simpson, and S. L. Bell, "Objective Prediction of the Sound Quality of Music Processed by an Adaptive Feedback Canceller," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 6, pp. 1734–1745, 2012.

[15] J.M. Kates and K.H. Arehart, "The Hearing-aid Audio Quality Index (HAAQI)," *IEEE/ACM Transactions on Audio, Speech, and Language processing*, vol. 24, no. 2, pp. 354–365, 2016.

[16] K. Friedrichs, *Musikklassifikation mittels auditorischer Modelle zur Optimierung von Hörgeräten*, Ph.D. thesis, in German, TU Dortmund, 2016.

[17] K. Gfeller, A. Christ, J. Knutson, S. Witt, and Maureen 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.

[18] V. Looi, H. McDermott, C. McKay, and L. Hickson, "Comparisons of Quality Ratings for Music by Cochlear Implant and Hearing Aid Users," *Ear and Hearing*, vol. 28, no. 2, pp. 59S–61S, 2007.

[19] K. Gfeller, S. Witt, G. Woodworth, M. A. Mehr, and J. Knutson, "Effects of Frequency, Instrumental Family, and Cochlear Implant Type on Timbre Recognition and Appraisal," *Annals of Otology, Rhinology & Laryngology*, vol. 111, no. 4, pp. 349–356, 2002.

[20] K. Gfeller, A. Christ, J. F Knutson, S. Witt, K. T. Murray, and R. S. Tyler, "Musical Backgrounds, Listening Habits, and Aesthetic Enjoyment of Adult Cochlear Implant Recipients," *J. Am. Academy of Audiology*, vol. 11, no. 7, pp. 390–406, 2000.

[21] V. Looi and J. She, "Music Perception of Cochlear Implant Users: A Questionnaire, and Its Implications for a Music Training Program," *International Journal of Audiology*, vol. 49, no. 2, pp. 116–128, 2010.

[22] S. Möller, *Assessment and Prediction of Speech Quality in Telecommunications*, Kluwer Academic Publishing, Dordrecht, Netherlands, 2000.

[23] M. Manrique, F. J. Cervera-Paz, A. Huarte, and M. Molina, "Advantages of Cochlear Implantation in Prelingual Deaf Children Before 2 Years of Age When Compared With Later Implantation," *The Laryngoscope*, vol. 114, no. 8, pp. 1462–1469, 2004.

[24] C. Mitani, T. Nakata, S. E. Trehub, Y. Kanda, H. Kumagami, K. Takasaki, I. Miyamoto, and H. Takahashi, "Music Recognition, Music Listening, and Word Recognition by Deaf Children With Cochlear Implants," *Ear and Hearing*, vol. 28, no. 2, pp. 29S–33S, 2007.

[25] O. Lartillot and P. Toiviainen, "A Matlab Toolbox for Musical Feature Extraction from Audio," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, 2007, pp. 237–244.