A Model for Music Complexity Applied to Music Preprocessing for Cochlear Implants

Wim Buyens[1 2 3], Marc Moonen[2], Jan Wouters[3] and Bas van Dijk[1]

[1]Cochlear Technology Centre Belgium, Mechelen, Belgium
[2]KU Leuven – University of Leuven, Department of Electrical Engineering (ESAT-STADIUS), Heverlee, Belgium
[3]KU Leuven – University of Leuven, Department of Neurosciences (ExpORL), Leuven, Belgium

Abstract—**Music appreciation remains challenging for cochlear implant users. In previous studies a strong negative correlation was found with cochlear implant subjects between music appreciation and music complexity. In this paper, music features that contribute to music complexity are investigated and related to a music preprocessing scheme for cochlear implants, in which a complexity reduction is achieved in an attempt to increase music appreciation. First, a complexity rating experiment is performed with pop/rock music excerpts and a linear regression model is developed to describe this (subjective) music complexity based on different music features. Subsequently, this model is used to validate the complexity reduction in the music preprocessing scheme and to provide an indication for the preferred setting for the balance between vocals/bass/drums and the other instruments for cochlear implant subjects.**

Keywords—music complexity, music appreciation, cochlear implants

I. Introduction

Music appreciation in cochlear implant (CI) users is generally poor. In [1] the correlation between music complexity and music appreciation was investigated. In normal-hearing (NH) subjects, this correlation was found to be positive, whereas for CI subjects this correlation was strongly negative. Music that was rated less complex, was appreciated more. Based on the music mixing preferences for CI subjects found in [2], a music preprocessing scheme was described in [3] to reduce music complexity in order to increase music appreciation for CI users. Complexity reduction was achieved by extracting vocals/bass/drums from stereo recordings and attenuating the other instruments with an adjustable attenuation parameter. A similar approach to reduce music complexity by remixing the music was described in [4] and [5]. In [6] a different approach to reduce (spectral) complexity was investigated in order to increase melody clarity and ease of listening for CI users in instrumental music with reduced-rank approximations of music signals. In [7] the impact of harmonic series reduction on music enjoyment was investigated with NH and CI subjects.

In the evaluation of the music preprocessing scheme with CI subjects in [3], a positive correlation was found between the (subjective) music complexity and the preferred attenuation parameter which balances vocals/bass/drums with the other instruments. Thus, with more complex music, the preferred attenuation applied to the other instruments was higher. Based on these findings, it is anticipated that a prediction of music complexity with a music complexity model might give an indication for the preferred attenuation parameter setting.

This paper is organized in three sections. First, the complexity rating experiment from [3] is summarized and discussed. Second, a linear regression model is developed based on the complexity rating results and the different features in the music excerpts. Finally, the complexity reduction of the music preprocessing scheme for CIs is validated based on the music complexity model and an indication for the preferred setting is determined.

II. Complexity Rating Experiment

In this section the complexity rating experiment from [3] is described, in which sound excerpts are rated in terms of music complexity by NH subjects. In the following subsections the sound material, the method and the results are discussed.

A. Sound material

For the complexity rating experiment, pop/rock songs were used from the top fifty songs in the all-time greatest hits list of a popular radio station in Belgium (Joe FM). Representative excerpts of the songs were selected, all including lead vocals, with an average length of 27 seconds and an average dynamic range of 10.0 dB (SD = 1.5 dB), measured with the TT Dynamic Range Meter from [8]. The song excerpts were rms-equalized and stored as mono wav-files with sampling rate of 44.1 kHz.

B. Method

A complexity rating experiment was performed with twelve NH subjects with no self-reported hearing deficit. The subjects were recruited with an internal advertisement, had diverse musical background and were familiar with most of the music excerpts. The fifty song excerpts were played in random order through headphones (Beyerdynamic DT-770 pro) in a silent room and the subjects were asked to rate the music complexity of the song on a scale from 1 to 100 with a slider in a graphical user interface on a laptop. No further definition or information was given to the subjects in order not to prime them in the experiment. Two additional song excerpts were included as training before starting the experiment. The individual results were combined and the median rating over all subjects was taken as the final complexity rating for each song. In a similar complexity rating experiment in [1] a strong positive correlation (r = 0.85) was found between the complexity ratings of NH and

This work was supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT150280) and the Cochlear Technology Centre Belgium.

CI subjects, and therefore the use of the complexity ratings of NH subjects was justified for this study.

C. *Results*

In Fig. 1 the results from the complexity rating experiment are shown. The median complexity ratings are visualized for all songs, ranked from less complex to more complex. The complexity rating over all songs was on average 50, the minimum was 19 and the maximum was 77. The most complex rated songs were heavy rock songs with prominent guitar riffs, the least complex rated songs were songs with calmer accompaniment, such as ballads. In the next section, the median complexity ratings from the complexity rating experiment are combined with the music features from the song excerpts to develop a music complexity model.

III. MUSIC COMPLEXITY MODEL

In this section a model for music complexity is developed based on the complexity ratings from the previous section and the music features from the corresponding song excerpts. The following subsections describe the selected features, the music complexity model and a discussion of the results.

A. *Music features*

To characterize the song excerpts, music features were extracted from the mono wav-files. Most of the features originated from the Music Information Retrieval Toolbox (MIRtoolbox 1.6.1), which is described in detail in [9]. For the calculation of the features with the MIRtoolbox the default settings were used, unless stated otherwise. The extracted features were all independent of the rms-value of the song excerpt. The features used in the music complexity model are as follows:

- Roughness (based on *mirroughness*): In [10] an estimation of the sensory dissonance (or roughness) was proposed related to the beating phenomenon whenever a pair of sinusoids are close in frequency. The estimation of roughness depends on the frequency ratio of each pair of sinusoids. In *mirroughness* the total roughness or sensory dissonance is estimated by computing the peaks of the spectrum, and taking the average of the dissonance between all possible pairs of peaks. A given local maximum is considered as a peak if the difference of amplitude with respect to both the previous and successive local minima (when they exist) is higher than the (default) threshold of 0.01. The overall estimate for roughness is calculated as the mean roughness of successive frames, divided by the rms-value (*mirrms*):

$$Roughness = \frac{\sqrt{mean(mirroughness)}}{mirrms} \quad (1)$$

- Crest factor: The crest factor is defined as the ratio (expressed in dB) of the peak amplitude to the rms value of the waveform. It is calculated as:

$$Crest = 20 * log_{10}\left(\frac{|x|_{peak}}{x_{rms}}\right) \quad (2)$$

  in which x represents the waveform.

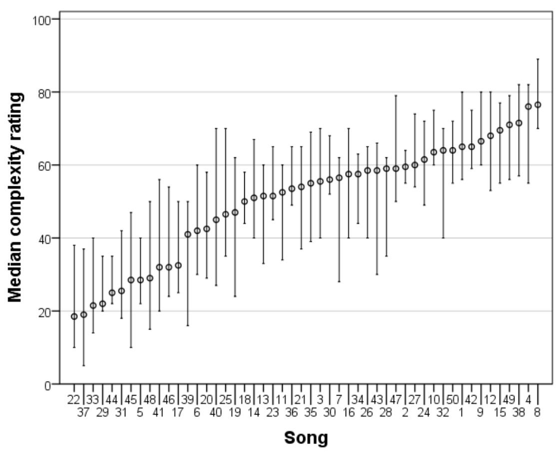

Fig. 1. Median complexity ratings for 50 pop/rock songs in the complexity rating experiment of [3] with 12 NH subjects. Songs are ranked from less complex to more complex. Error bars indicate 95% confidence interval.

- Tempo (*mirtempo*): The tempo is estimated by first computing an onset detection curve, showing the successive bursts of energy corresponding to the successive pulses. Next, the periodicities are detected by computing an autocorrelation function of the onset detection curve. Finally, a peak picking is applied to the autocorrelation function to determine the tempo (expressed in beats per minute).

- Pulseclarity (*mirpulseclarity*): The pulseclarity estimates the rhythmic clarity, indicating the strength of the beats estimated by the *mirtempo* function, and is based on the maximum correlation value in the autocorrelation curve computed for tempo estimation. This is described in more detail in [11].

Other features from the MIRtoolbox that were considered for the music complexity model were *mirbrightness*, *mirattackslope* and *mireventdensity*, but the inclusion of these did not increase the prediction quality. Moreover, these features were significantly correlated with respectively roughness ($r(50) = 0.59$, $p < 0.001$), pulseclarity ($r(50) = 0.64$, $p < 0.001$) and again roughness ($r(50) = 0.67$, $p < 0.001$). Consequently, these features were discarded from the music complexity model.

B. *Music complexity model*

By combining the music features from the song excerpts and the corresponding median complexity ratings, a model for music complexity was developed. This was accomplished by fitting a linear regression model. The regression coefficients are given in Table I. In the first column the different features are presented. The second and the third column show the coefficients in the linear regression model, respectively with and without normalization of the features. With normalization, all features are in the range from 0 to 1, and therefore the importance of a specific feature in the music complexity model is determined by the absolute value of its corresponding coefficient as listed in the second column. Positive coefficients contribute to a higher

ISBN 978-0-9928626-7-1

TABLE I. REGRESSION COEFFICIENTS FOR THE MUSIC FEATURES IN THE MUSIC COMPLEXITY MODEL

Features	Coefficients (normalized features)	Coefficients (original features)
Roughness	41.88	0.34
Crest factor	-23.86	-2.20
Pulseclarity	17.93	23.13
Tempo	11.33	0.092
const	30.80	13.20

complexity, whereas negative coefficients contribute to a lower complexity. Nevertheless, to facilitate the calculations in the model, the original features without normalization are used in the remainder of the paper.

C. *Results*

The predicted complexity with the linear regression model for the fifty pop/rock song excerpts correlated well with the complexity ratings from the complexity rating experiment with a correlation coefficient of $r = 0.83$ and a mean absolute error of 6.71. In order to check the generalization of the model, ten-fold cross-validation was performed. The correlation coefficient remained high at $r = 0.80$ with a mean absolute error of 7.26. In Fig. 2 the predicted complexity is plotted against the rated complexity for all song excerpts.

Features that positively contributed to music complexity were roughness, pulseclarity and tempo (positive coefficients in Table I), whereas the crest factor negatively contributed to music complexity (negative coefficient in Table I). Song excerpts with more sensory dissonance, with more rhythmic clarity and/or higher in tempo were thus estimated with higher complexity, in case all other features remained unaltered. On the contrary, songs with higher dynamics were estimated with lower complexity, in case all other features remained unaltered.

Besides the features from the MIRtoolbox, also the crest factor was included in the music features. In the evaluation of the music preprocessing scheme in [12] an effect of genre was found, and it was suggested to be related to the different audio mixing trends in older recordings (less compression). Therefore, the crest factor was included in the model as a measure for the compression in complex music. With the given sound material as training data, the inclusion of the crest factor increased the prediction quality of the model substantially. Moreover, there was a significant correlation between the crest factor and the rated complexity (Pearson's $r(50) = -0.36$, $p = 0.009$).

IV. MUSIC PREPROCESSING FOR COCHLEAR IMPLANTS

In [3], a music preprocessing scheme was described that attempts to improve music appreciation for CI users by attenuating certain instruments and thus reducing music complexity. By using the music complexity model from the previous section, the complexity reduction of the music preprocessing scheme can be measured. Since (subjective) music complexity was correlated with the preferred attenuation parameter in the evaluation of the music preprocessing scheme in [3], modeling and measuring the music complexity can also be used in the music preprocessing scheme to fine-tune the adjustable attenuation parameter. In the following subsections the music preprocessing and its attenuation parameter are briefly reviewed and the results of the achieved complexity reduction is described and discussed.

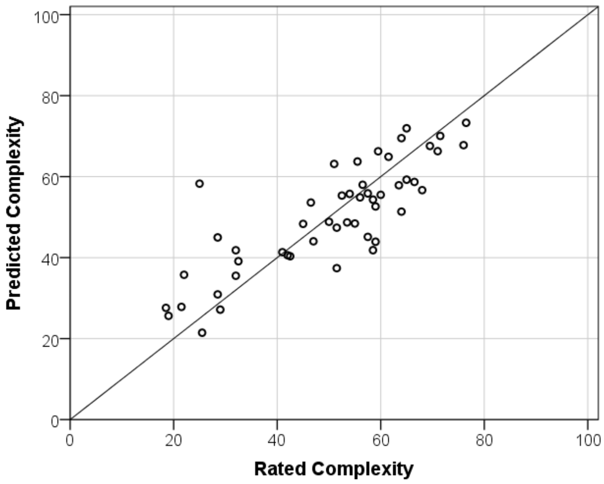

Fig. 2. Rated complexity for 50 pop/rock song excerpts obtained from the complexity rating experiment versus their predicted complexity obtained from the music complexity model based on music features.

A. *Music preprocessing for cochlear implants*

In the music preprocessing scheme in [3] vocals/bass/drums are extracted from the other instruments. The extraction is based on the representation of harmonic and percussive components in the spectrogram and on the panning settings in typical stereo recordings. With an adjustable attenuation parameter the balance between vocals/bass/drums and other instruments can be altered. In the take-home evaluation with CI subjects in [12] the attenuation parameter could range from -6 dB to 24 dB. An attenuation parameter of 0 dB corresponded to the original recording, whereas a negative and positive value represented respectively an amplification or attenuation of the other instruments in the final mix. The task for the subjects was to determine for each song the preferred attenuation parameter with which the music sounded most enjoyable. Although individual differences occurred, all subjects preferred an attenuation parameter setting to construct a mix with the other instruments attenuated, significantly different from the original mix. In [3] sound excerpts with high (subjective) music complexity were preferred with significantly higher attenuation parameter, compared to songs with lower complexity. The preferred average setting of the attenuation parameter for the 24 song excerpts used in [3] was positively correlated with the complexity of the song as rated by the NH subjects (Pearson's $r(24) = 0.67$, $p < 0.001$) with a linear fit (R^2=0.43):

$$A = 0.125 * C + 5.4 \quad (3)$$

in which A represents the preferred attenuation parameter (in dB) and C the rated music complexity. Consequently, by using (3) and the calculation of the (subjective) music complexity with the music complexity model from the previous section, an indication for the preferred attenuation parameter setting in the music preprocessing scheme can be determined.

ISBN 978-0-9928626-7-1

B. Complexity reduction

All song excerpts from the complexity rating experiment were processed with the music preprocessing scheme with attenuation parameter set to 24 dB, i.e. the level of the other instruments was reduced by 24 dB. The predicted music complexity in the original song excerpts was compared with the predicted music complexity of the preprocessed song excerpts by using a paired samples t-test. A significant reduction in complexity was found between the original song excerpts (M=50.3, SD=13.3) and the preprocessed song excerpts (M=41.8, SD=13.3) ($t(49) = 11.8$, $p < 0.001$). The average complexity reduction was 8.5 and ranged from 1.9 to 24.7. Expressed in percentage relative to the predicted music complexity of the original song excerpt, the average reduction was 18% and ranged from 4% to 71%. While the feature *tempo* was not changed with music preprocessing, the feature *roughness* was decreased by 14%, and the features *pulseclarity* and *crest* were increased by respectively 4% and 9%. Only for one song excerpt (Ex. 28) a small increase in complexity from 43.9 to 44.9 (2%) was found after preprocessing. In Fig. 3 the complexity reduction with the music preprocessing scheme is visualized as a boxplot for original and preprocessed excerpts.

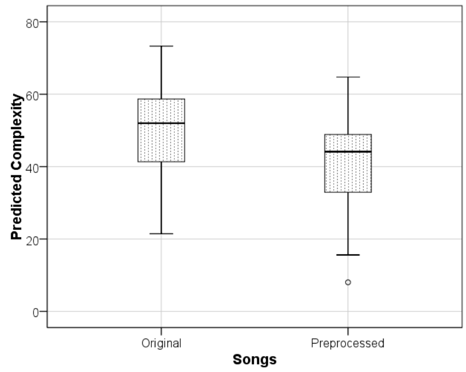

Fig. 3. Predicted complexity of the 50 original pop/rock song excerpts compared to the predicted complexity of the 50 preprocessed song excerpts (with attenuation parameter 24 dB) represented as boxplots.

In Table II, the complexity reduction of the two song excerpts with highest and lowest rated complexity (see Fig. 1) are presented. The second column represents the predicted complexity of the original song excerpt, the third column shows the predicted complexity of the same excerpt processed with the music preprocessing scheme with attenuation parameter 24 dB. Both complexity numbers were measured with the music complexity model from the previous section. The fourth column in Table II shows the complexity reduction achieved by the music preprocessing scheme.

TABLE II. COMPLEXITY REDUCTION OF TWO SONG EXCERPTS WITH HIGHEST AND LOWEST RATED COMPLEXITY (SEE FIG. 1)

Song	Predicted Complexity (original)	Predicted Complexity (att=24dB)	Complexity Reduction
1 Most complex (ex 8)	73.3	48.6	24.7 (33%)
2 Most complex (ex 4)	67.8	58.2	9.6 (14%)
1 Least complex (ex 22)	27.6	8.0	19.6 (71%)
2 Least complex (ex 37)	25.6	15.6	10.0 (39%)

C. Discussion

In the evaluation of the music preprocessing scheme with CI subjects in [3] and [12] the test subjects were asked to determine their preferred setting for the attenuation parameter in order to make the music sound most enjoyable for them. By increasing the attenuation parameter the level of the "other" instruments was decreased (i.e. other than vocals/bass/drums). In [3] a significantly higher attenuation was preferred for songs with higher (subjective) music complexity. It was assumed that the music complexity was reduced by attenuating/reducing the other instruments with the music preprocessing scheme. This assumption was supported by [1] in which a strong negative correlation was found between music appreciation and (subjective) music complexity, and also in [13] CI subjects judged music that involved multiple instruments, on average, less pleasant than music played by a single instrument. By determining a model for music complexity, based on the subjective complexity ratings from the complexity rating experiment, this assumption was confirmed. In the sound excerpts discussed in this paper a music complexity reduction from 4% to 71% was achieved with the music preprocessing scheme, based on the music complexity model. The main features contributing to this music complexity reduction were the lower sensory dissonance (*roughness*) and the higher peak-to-rms ratio (*crest*). The feature *pulseclarity*, which also gave a positive contribution to music complexity, was in most songs slightly increased by the music preprocessing scheme, since besides vocal enhancement, the music preprocessing scheme also enhanced rhythm/beat. Although the feature *tempo* is also included in the music complexity model, this feature was not altered by the music preprocessing scheme. It should be noted that the model was developed based on the (subjective) music complexity ratings of 50 pop/rock song excerpts. Heavy rock songs were rated as the most complex songs, whereas calmer songs, such as ballads, were typically rated least complex. Consequently, the feature *tempo* was included in the music complexity model. Nevertheless, the intrinsic complexity which originated from the feature *tempo* could not be reduced by the music preprocessing scheme.

According to the music complexity model, the complexity was not reduced for song excerpt 28. Investigating the music features, it was found that on the one hand, *roughness* was only slightly reduced and on the other hand *pulseclarity* was increased and *crest* was slightly decreased. This particular song excerpt has the leading vocals panned off-center, which is not optimal for the music preprocessing scheme that is (partly) based on the typical panning of vocals in the center of the stereo image.

V. CONCLUSION

Based on the music features from pop/rock song excerpts and their corresponding subjective complexity ratings, a music complexity model was developed. This model was used to objectively evaluate the complexity reduction in the music preprocessing scheme of [3], and can also be used to give an indication for the preferred attenuation parameter setting in the music preprocessing scheme with complex music.

REFERENCES

[1] 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, 40(2), pp. 78-112, 2003.

[2] W. Buyens, B. van Dijk, M. Moonen, and J. Wouters, "Music mixing preferences of cochlear implant recipients: a pilot study," International Journal of Audiology, 53(5), pp 294-301, 2014.

[3] W. Buyens, B. van Dijk, J. Wouters, and M. Moonen, "A stereo music preprocessing scheme for cochlear implant users," IEEE Transactions on Biomedical Engineering, 62(10), pp. 2434-42, 2015.

[4] 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, 2015.

[5] 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, 140(6), pp. 4338-4349, 2016.

[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 (TASLP), 24(3), pp. 445-458, 2016.

[7] J.S. Nemer, G.D. Kohlberg, D.M. Mancuso, B.M. Griffin, M.V. Certo, et al., "Reduction of the Harmonic Series Influences Musical Enjoyment With Cochlear Implants," Otology & Neurotology, 38(1), pp. 31-37, 2017.

[8] Pleasurize Music Foundation: http://www.pleasurizemusic.com (date last viewed 30/01/13).

[9] O. Lartillot, and P. Toiviainen, "A Matlab Toolbox for Musical Feature Extraction From Audio," International Conference on Digital Audio Effects, Bordeaux, 2007.

[10] R. Plomp, and W.J.M. Levelt, "Tonal consonance and critical bandwidth," The Journal of the Acoustical Society of America, 38(4), pp. 548-560, 1965.

[11] O. Lartillot, T. Eerola, P. Toiviainen, and J. Fornari, "Multi-feature modeling of pulse clarity: Design, validation, and optimization," International Conference on Music Information Retrieval, Philadelphia, 2008.

[12] 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, in press

[13] V. Looi, H.J. McDermott, C. McKay, and L. Hickson, "Comparisons of quality ratings for music by cochlear implant and hearing aid users," Ear and Hearing, 28(2), pp. 59S-61S, 2007.

ISBN 978-0-9928626-7-1 © EURASIP 2017