The following article appeared in Korvel G., Kostek B.: Discovering Rule-Based Learning Systems for the Purpose of Music Analysis. Journal of the Acoustical Society of America. Vol. 146, iss. 4, 2947 (2019) and may be found at (https://doi.org/10.1121/1.5137237).

Discovering Rule-Based Learning Systems for the Purpose of Music Analysis

Gražina Korvel[1], Bozena Kostek[2]
[1]Institute of Data Science and Digital Technologies, Vilnius University, Lithuania

[2]Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Audio Acoustics Laboratory, Narutowicza 11/12 80-233 Gdansk Poland
bokostek@audioacoustics.org

Music analysis and processing aim at understanding information retrieved from music (Music Information Retrieval). For the purpose of music data mining, machine learning (ML) methods or statistical approach are often employed. Their primary task is recognition of musical instrument sounds, music genre or emotion contained in music, identification of audio, assessment of audio content, etc. In terms of computational approach, music databases contain imprecise, vague and indiscernible data objects. Moreover, most of the machine learning algorithm outcomes are given as a black-box result. Also, underfitting or overfitting may occur, meaning that either the model description is not complex enough or the test set is too small or not sufficiently representative. Thus the goal is to generalize the model. To overcome some of these problems, rule-based systems may be used, e.g., based on rough set theory that shows the outcome in the form of rules interconnecting features retrieved from music. A potential of the rough set-based approach, a rule-based classifier was shown in the context of music genre recognition. The results were analyzed in terms of the recognition rate and computation time efficiency.

1. INTRODUCTION

Nowadays, when speaking about the quality evaluation of musical instrument sounds, the approach is different depending on the area of applicability and expert knowledge. Quality of sound is vital for a musician to hear its timbre, a musicologist to discern the main characteristics of music performance, a composer to create new sounds, an audio engineer to prepare a song mix, scientists to analyze the nature of sound, music or an audio signal. Quality evaluation is also an essential part of testing coding algorithms within standardization organizations (ITU, International Telecommunications Union [11][12][24][30]; ISO (International Organization for Standardization), EBU (the European Broadcasting Union (EBU), IEC (International Electrotechnical Commission), AES (Audio Engineering Society), MPEG (Moving Picture Experts Group) [9], etc. Thus, their approach is to test coding efficiency but, at the same time, also quality of an audio signal by means of subjective tests. To make testing easy to reproduce and compare, procedures should adhere to several guidelines concerning listening test procedure, listening environment, listening system, listening level, eligibility of subjects, subjective evaluation criteria and opinion scales, equivalence of opinion scales between languages, category judgment, instructions to subjects practical procedures for subjective testing [11][15][30][34]. Furthermore, this is only a part of the bigger picture as the way of recording or statistical analyses performed may influence the overall judgment as well. Since the subject is very broad, thus in this paper issues related to listening tests and their analysis are discussed.

Organizing a subjective listening test is also an extensive topic, as it depends on several factors. Quality may concern sound and its detailed perceptual characteristics or music, produced by a musician or recorded by an audio engineer. To be assessed, they both need an appropriate vocabulary, easy to understand, and be unequivocally interpreted by the subjects, such as brightness, intimacy, liveness, sharpness, fullness, density and definition. Moreover, this vocabulary may differ in relation to the task assigned to the testers. It may concern quality of an instrument sound or a synthesized sound, music [15][16][17], mood of music [19]. It is worth noting that this vocabulary is language-

Post-print of: Korvel G., Kostek B.: Discovering Rule-Based Learning Systems for the Purpose of Music Analysis. Journal of the Acoustical Society of America. Vol. 146, iss. 4, 2947 (2019). DOI: https://doi.org/10.1121/1.5137237

dependent [29]. For comparing sound/music quality the expressions used in tests may be adequate cross-language, but for other cases, they may not easily be translated to another language.

A simple translation of vocabulary from English into Polish, and then using it in subjective tests is insufficient. This also refers to other languages, e.g., Lithuanian. Many such words are inadequate to describe music and emotions that it creates. Although some of these words may easily be understood to the testers, they may not be commonly used in the context of music. An excellent example of such a situation is 'valence' form the Thayer's model (Valence/Arousal) [33]. Moreover, it was discovered in the study co-authored by her that Polish expressions are less diverse than English. Thus, a question arises, whether preparing a vocabulary for listening quality evaluation test, one should create it on the basis of one's language or use terms derived from models created in English.

Concerning judgment, also other problems appear. They are connected with the scale used [2][3][23], musical experience, the multidimensionality of the problem assessment [1][2][3][34][35], and most importantly, with the fact that correlation between perceptual assessment and objective measures as direct measurement of the perceived audio quality does not exist [2]. Objective quantification of the perceived sound is a critical issue. Typically, a mean opinion scale (MOS) is used in subjective assessment based on a five-point rating scale and the average of quality assessment, but several other scales are employed as well [5][13]. Examples are as follows: degradation category rating (DCR), absolute category rating (ACR), perceptual audio quality measure (PAQM) [14], MUSHRA (Multiple Stimulus with Hidden Reference and Anchors) [36] based on continuous 100-point ratings. The scale is from "Excellent" (100) to "Bad" (0) [36].

The issue related to the correlation between descriptors and parameters derived from the sound may be solved by creating a predictive model of audio quality by means of machine learning or rule-based decision systems. This will be discussed later on.

2. MUSIC ASSESSMENT BY SUBJECTIVE TESTING

A. MUSIC ASSESSMENT

In this Section, a short review of an approach to music assessment is presented. In Fig. 1 characteristics of a "bright" sound (flute G4 sound) are presented in the form of spectral analyses (spectrogram, Mel-cepstrogram and chromagram). Opposite to "bright", an example of "warm" sound is also included in Fig. 1 (bassoon A3 sound). Differences between all characteristics are easily discerned. Both terms are mastered by musicologists and audio engineers. They are also easy to use in listening tests as these terms are metaphorically synonymous with light and dark.

When we create a representation of "bright" and dark "sounds", we would like to correlate them with one or more descriptors that may be derived objectively, based on an analysis. For example, one may use the spectral centroid parameter as a measure of *brightness* of a sound. Spectral Centroid (SC) is defined (see Eqs. (1) and (2)) as a center of gravity of the spectrum, i.e., the weighted average frequency spectral power density ratios [26]. However, it is more difficult to judge an audio signal as bright, dark or warm – the opposite of bright music, however, these terms are still applicable to music (see Fig. 3).

$$\text{Brightness} = \frac{\sum_{i=f_c}^{M_{FT}/2} PS(i)}{\sum_{i=1}^{M_{FT}/2} PS(i)} \tag{1}$$

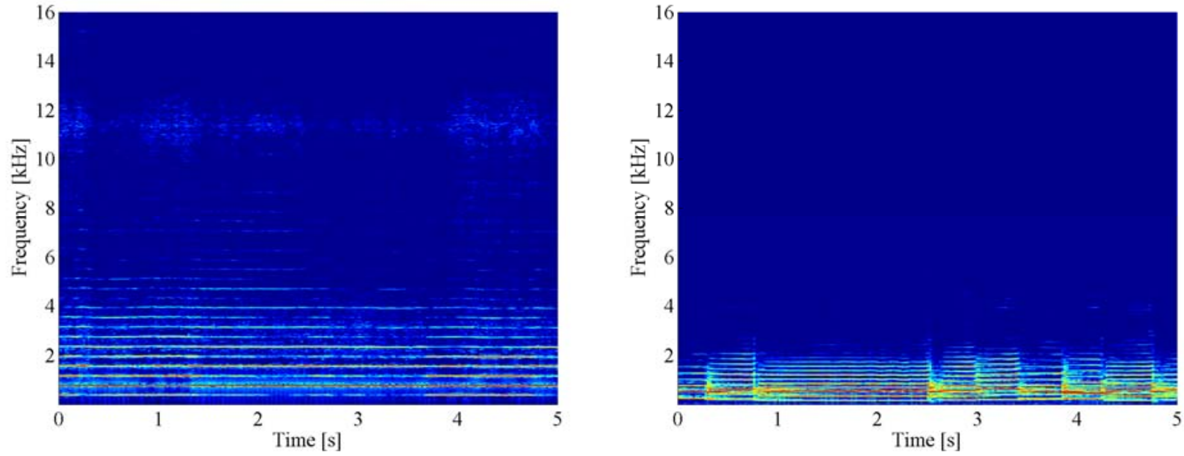
where f_c is cut-off frequency, PS denotes the power spectrum of sound signal, M_{FT} is the number of Fourier transform coefficients. A cut-off frequency (f_c) was set to 1500 Hz.
The mean values of Brightness calculated for short-time (2048 samples) segments (see Table 1).

Table 1. Brightness calculated for short-time (2048 samples) segments

Sound/Music	Bassoon (A3)	Flute (G4)	Hard rock	Classical
Brightness	0.023	0.096	0.426	0.101

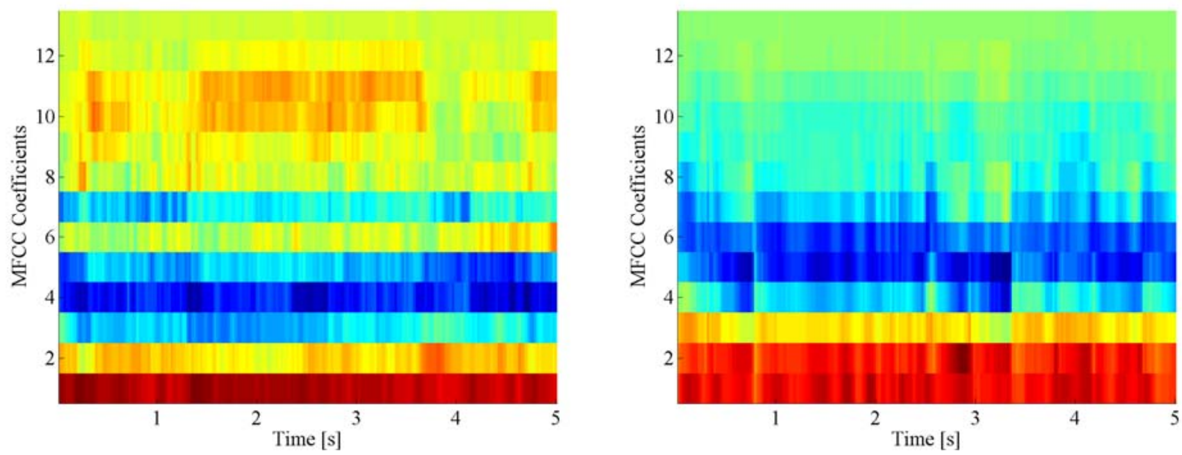
Post-print of: Korvel G., Kostek B.: Discovering Rule-Based Learning Systems for the Purpose of Music Analysis. Journal of the Acoustical Society of America. Vol. 146, iss. 4, 2947 (2019). DOI: https://doi.org/10.1121/1.5137237

Spectrograms



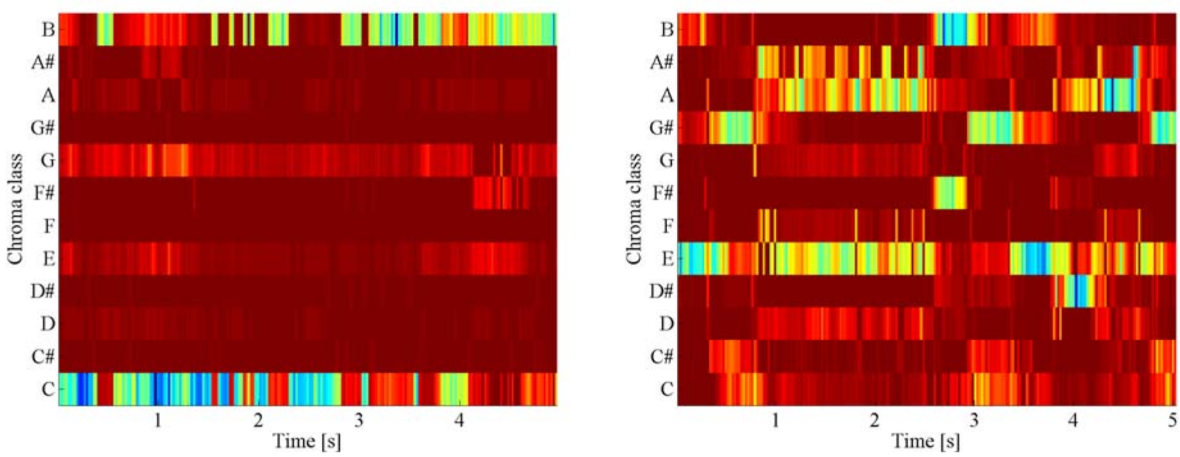
flute (G4) bassoon (A3)

Mel-cepstrograms



flute bassoon

Chromograms

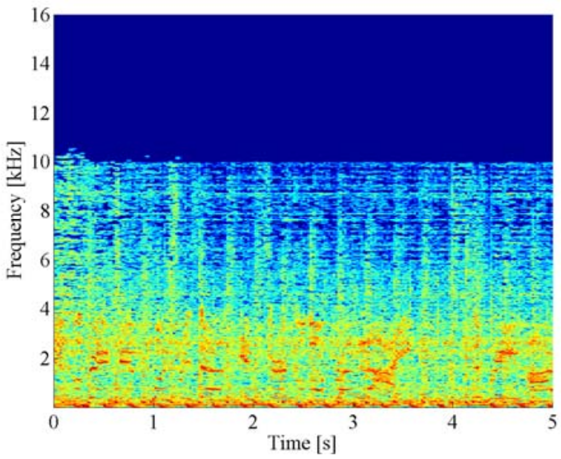


flute bassoon

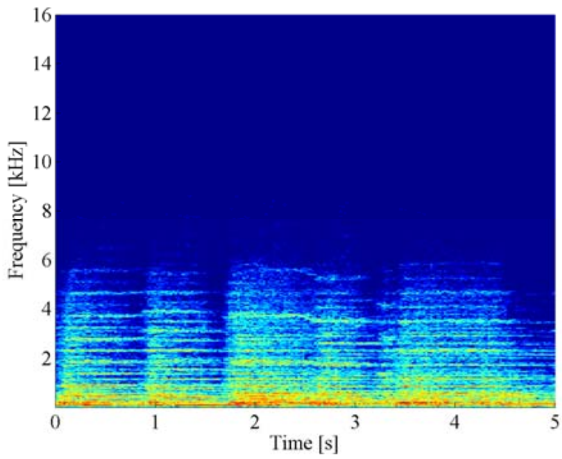
Figure 1. 2D spectral representation of flute and bassoon sounds.

Spectrograms

Post-print of: Korvel G., Kostek B.: Discovering Rule-Based Learning Systems for the Purpose of Music Analysis. Journal of the Acoustical Society of America. Vol. 146, iss. 4, 2947 (2019). DOI: https://doi.org/10.1121/1.5137237

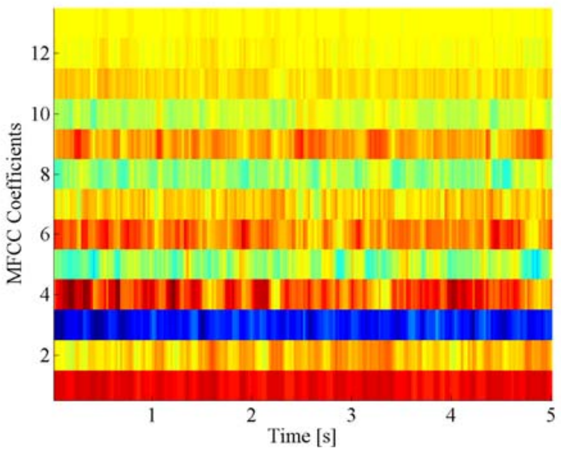


hard rock

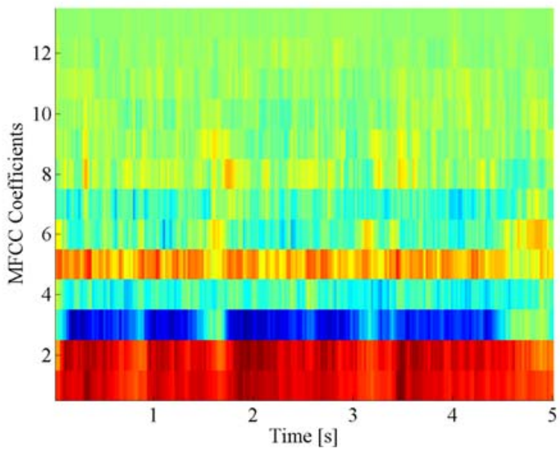


classical

Mel-cepstrograms

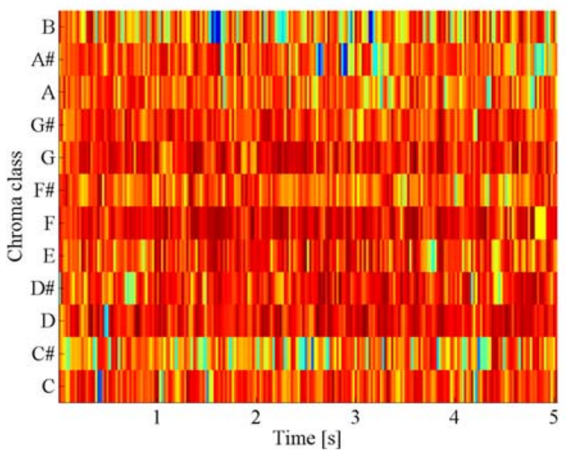


hard rock

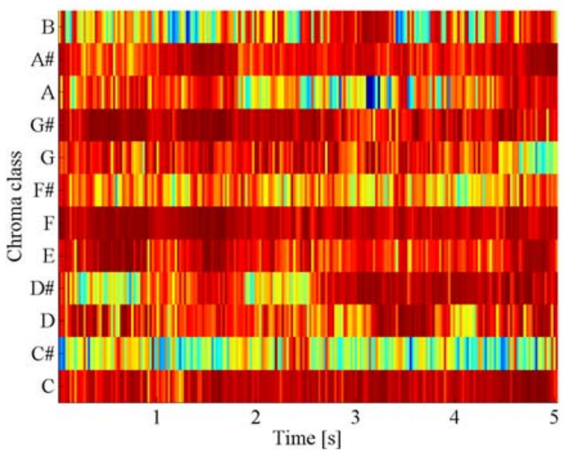


classical

Chromograms



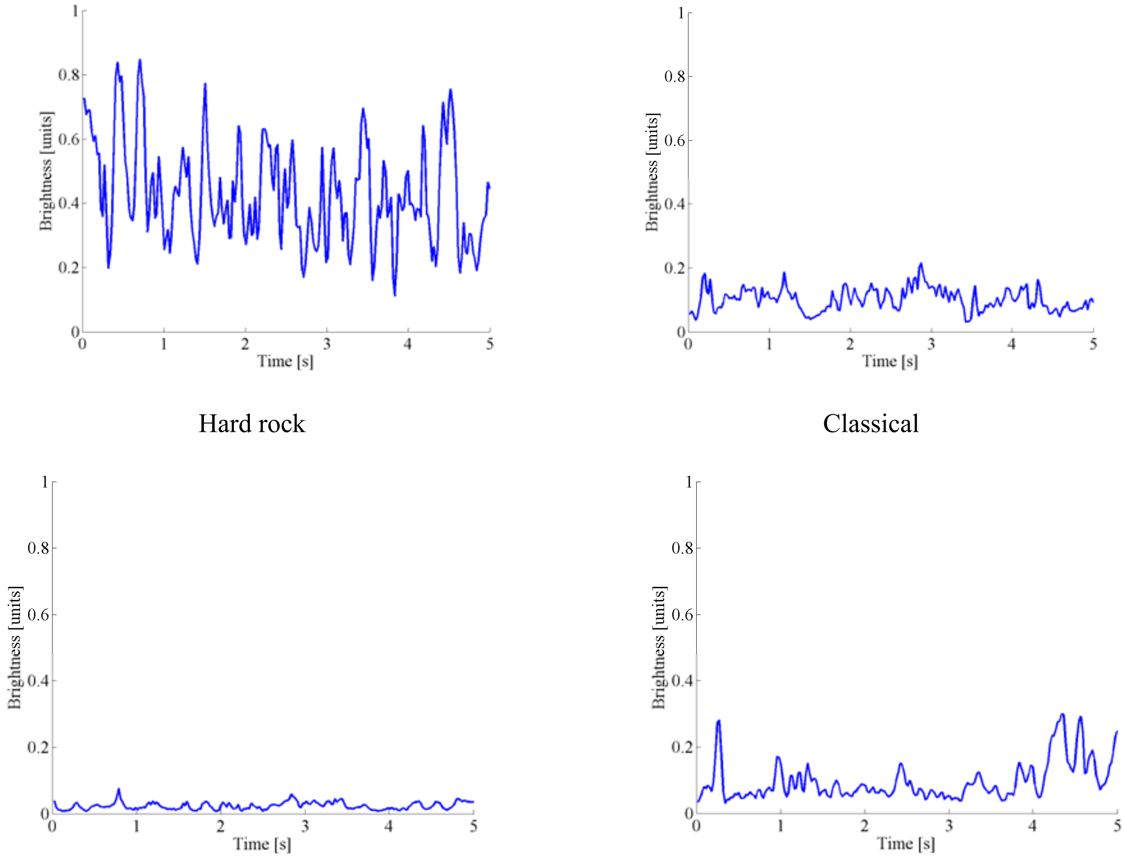
hard rock



classical

Figure 2. 2D spectral representation of hard rock and classical music genre excerpts.

Post-print of: Korvel G., Kostek B.: Discovering Rule-Based Learning Systems for the Purpose of Music Analysis. Journal of the Acoustical Society of America. Vol. 146, iss. 4, 2947 (2019). DOI: https://doi.org/10.1121/1.5137237



Hard rock Classical

Basson (A3) Flute (G4)

Figure 3. Brightness calculated for short-time (2048 samples) segments.

It may also be useful to observe what effect room characteristics may have on an audio signal [8]. This aspect is vital with regard to the listening conditions in subjective tests carried out. Figure 4 shows classical music reproduced in two rooms (room 1 - auditory room – larger, and room 2 - listening studio - smaller) with a professional high-end loudspeaker system (S1). The original signal has a higher level of high frequencies, which is not surprising, but at the same time, the reproduced signal has a higher level of low frequencies than the original one.

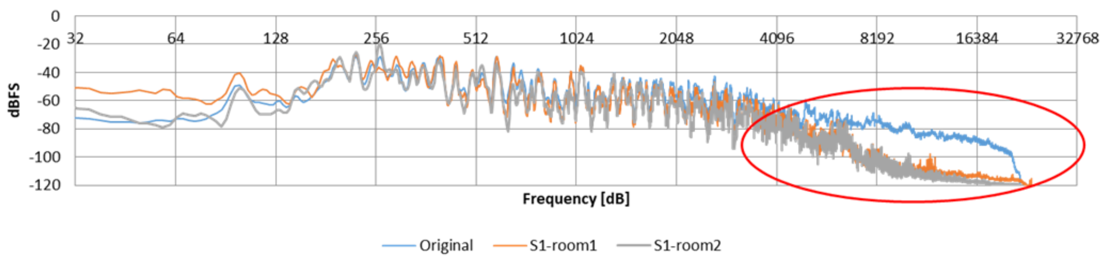


Figure 4. Classical music reproduced in two rooms (room 1 (larger) and room 2 (smaller)).

Analogously, Fig. 5 presents the same conditions but for rock music. Lack of low and high frequencies along with a lower level of the signal in the whole bandwidth is easily observed.

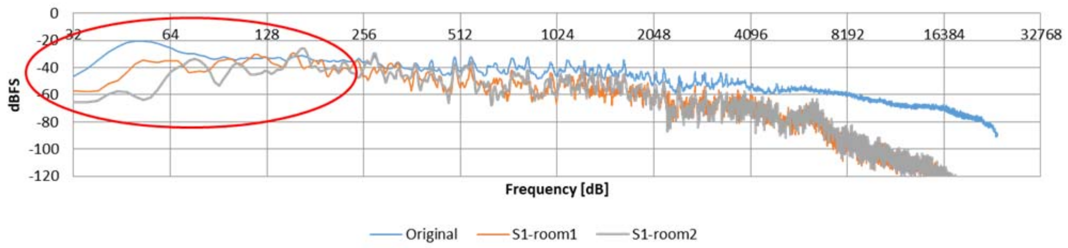


Figure 5. Rock music reproduced in two rooms (room 1 (larger) and room 2 (smaller)).

In Fig. 6 an interface of an application called "Subjective assessment of music recording", created for performing subjective tests is shown. Fig. 7 shows navigation to the folder where an audio file is located.

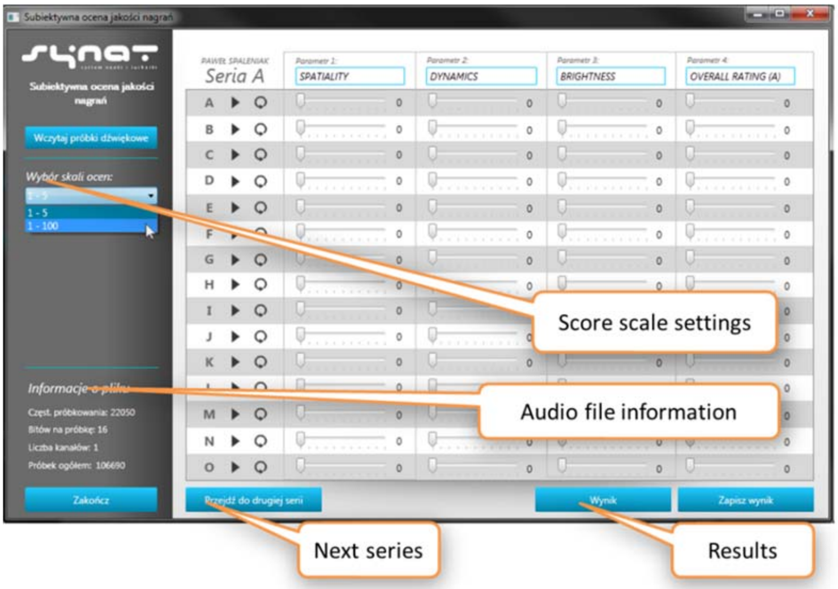


Figure 6. "Subjective assessment of music recording" application

Parameters that may be tested when listening to music are as follows: spatiality, dynamics, brightness (and warmth), transparency, coherence, dynamic balancing, base width and continuity, brightness, dynamics, dynamic balancing, power of sound, sound power, overall quality (some of them are listed in the interface shown in Fig. 6).

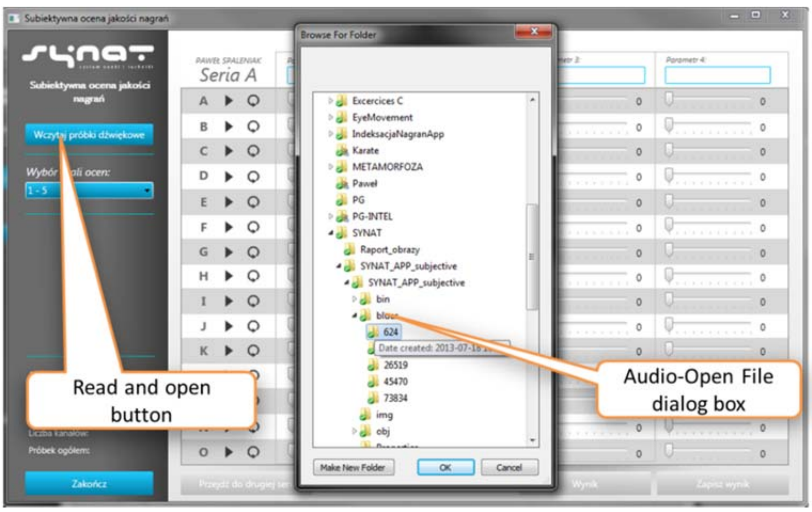


Figure 7. Navigating to the folder where an audio file is located.

Post-print of: Korvel G., Kostek B.: Discovering Rule-Based Learning Systems for the Purpose of Music Analysis. Journal of the Acoustical Society of America. Vol. 146, iss. 4, 2947 (2019). DOI: https://doi.org/10.1121/1.5137237

There are also many studies related to mood classification with different findings and conclusions [10][25][29]. To evaluate several algorithms within the same system, MIREX (Music Information Retrieval Evaluation eXchange) organized a set of clusters within MIREX Audio Mood Classification task, i.e., five mutually exclusive categories [25]:

Cluster 1: passionate, rousing, confident, boisterous, rowdy;

Cluster 2: rollicking, cheerful, fun, sweet, amiable/good-natured;

Cluster 3: literate, poignant, wistful, bittersweet, autumnal, brooding;

Cluster 4: humorous, silly, campy, quirky, whimsical, witty, wry;

Cluster 5: aggressive, fiery, tense/anxious, intense, volatile, visceral.

Apart from the 'classical' mood models, such as Hevner, Thayer, Russel, etc. [7][32][33], this is one of the ways to deal with describing the emotional content of music and then search for correlation with parameters derived from music [19][20][21][22].

B. MUSIC ASSESSMENT BY MACHINE LEARNING

It is interesting to see whether listeners recognize songs correctly belonging to different music genres. Moreover, there is a question of how well a machine learning algorithm performs having the same task. In the experiment, several subjects participated in music genre recognition while listening to particular songs. The same songs were introduced to the classifiers, namely Bayesian Network [6] and *Sequential Minimal Optimization* Algorithm (SMO) [8].

75 high-quality fragments of songs with a length of 10 seconds were selected for the sample database, belonging to the following musical genres, i.e., pop, rock, rap/hip-hop, classical, jazz, electronic, hard rock/metal, blues, country, R&B, New Age and folk. These samples were prepared as being unequivocally or ambiguously belonging to a given genre. Different fragments of songs were also used, which could be classified differently, as well as the same songs recorded in different acoustic conditions (studio, concert, acoustic versions) or by performers representing various genres. Due to the large dataset of the test material, it was necessary to divide the test into five smaller surveys, each containing fifteen samples.

The findings were reported thoroughly in a paper co-authored by the authors [4], but the most interesting outcome was a comparison of the classifier results and listening tests. In the case of the Bayesian Network classifier and SMO classifier, the number of correctly and incorrectly evaluated samples was the same. In both cases, the number of incorrectly classified instances was three. Comparing the results obtained in the subjective tests and using learning algorithms, it may be noticed that among the samples selected for comparison with the subjective tests, only three were classified not in accordance with the opinion of the listeners. In the Bayesian Networks classifier, Bach's Largo was rated as electronic music, while Jarre's Equinoxe as classical. AC / DC's "Child Child" was classified as pop. Besides, listeners rated two samples of "Jasey Rae" as different genres (rock and pop), the same rating was given by the classifier. In contrast, the SMO classifier correctly recognized the rock sample "Problem Child" AC/DC and misjudged Henry Mancini's "Unchained Melody", representing jazz, assigning the song the genre of blues (see Fig. 8).

Post-print of: Korvel G., Kostek B.: Discovering Rule-Based Learning Systems for the Purpose of Music Analysis. Journal of the Acoustical Society of America. Vol. 146, iss. 4, 2947 (2019). DOI: https://doi.org/10.1121/1.5137237

Led Zeppelin - Since I've Been Lovina You - Blues	Led Zeppelin - Since I've Been Lovina You - Blues
Bach - Largo – Classic	Bach - Largo – Classic
Black Veil Brides - Overture – Classic	Black Veil Brides - Overture – Classic
Grieg - Poranek – Classic	Grieg - Poranek – Classic
Mozart - Eine Kleine Nacht Musik – Classic	Mozart - Eine Kleine Nacht Musik – Classic
Telemann - Trumpet Concert – Classic	Telemann - Trumpet Concert – Classic
Blues Brothers - Theme from Rawhide - Country	Blues Brothers - Theme from Rawhide - Country
Hannes Wader - Heute Hier, Morden Dort - Country	Hannes Wader - Heute Hier, Morden Dort - Country
Sheamus Fitzpatrick and the Mcnallv Boys - Whisky In The Jar - Country	Sheamus Fitzpatrick and the Mcnallv Boys - Whisky In The Jar - Country
Jarre - Equinoxe – Electronic	Jarre - Equinoxe – Electronic
Linkin Park ft Jay Z - Numb - Rap Hip Hop	Linkin Park ft Jay Z - Numb - Rap Hip Hop
Glenn Miller - In The Mood - Jazz	Glenn Miller - In The Mood - Jazz
Glenn Miller - Over The Rainbow - Jazz	Glenn Miller - Over The Rainbow - Jazz
Henry Mancini - Unchained Melody - Jazz	Henry Mancini - Unchained Melody - Jazz
Metallica - Whiskey In The Jar - Hard Rock Metal	Metallica - Whiskey In The Jar - Hard Rock Metal
Abba - Waterloo - Pop	Abba - Waterloo - Pop
Adele - Hello - Pop	Adele - Hello - Pop
Agnetha - I Should've Followed You Home - Pop	Agnetha - I Should've Followed You Home - Pop
All Time Low - Jasey Rae - Pop	All Time Low - Jasey Rae - Pop
Miley Cyrus - Wreckling Ball - Pop	Miley Cyrus - Wreckling Ball - Pop
ACDC - Problem Child - Rock	ACDC - Problem Child - Rock
All Time Low - Jasey Rae - Rock	All Time Low - Jasey Rae - Rock
Boston - More Than A Feeling - Rock	Boston - More Than A Feeling - Rock
Journey - Don't Stop Believing - Rock	Journey - Don't Stop Believing - Rock
Oueen - Mustapha - Rock	Oueen - Mustapha - Rock
The Calling - Wherever You Will Go - Rock	The Calling - Wherever You Will Go - Rock
The Kinks - You Really Got Me - Rock	The Kinks - You Really Got Me - Rock

Figure 8. Classification by listeners and machine learning algorithms (in red - incorrectly classified music excerpts) [4].

3. EXAMPLES OF RULE-BASED ANALYSES

A. RULE-BASED DECISION ALGORITHM

The rough sets theory was founded in the early 1980s by Z. Pawlak [27]. Its main application is the synthesis and effective analysis of data sets. Methods using rough set theory have found application, among others in data mining and knowledge discovery in complex classification tasks and in computer decision support systems [28]. The rough set theory rejects the requirement for well-defined boundaries of the set. The range of rough sets is defined as the lower and upper approximation, and the difference between the upper and lower approximation is defined by the boundary area, which includes all cases that cannot be classified without conflicts based on current knowledge. The lower approximation of the set is, therefore, the value to which all objects belong, which there is no doubt that they are representatives of this set in the light of knowledge. The upper approximation includes objects that cannot be ruled out that they are unambiguously representatives of this set. The edge of the set are all those objects for which it is not known whether they are or not representatives of a given set. The larger the edge area of the set, the less precise the objects in it. The rough set theory enables to process both quantitative and qualitative tabular data, obtained experimentally [28]. The basic data structure in information systems using rough set theories is a table. All data are grouped tabularly according to the principle that table rows are objects and the attributes are columns. Decision tables contain parameters that act as conditional attributes and the decision-making part, which means that the conditional attributes specify the value of the decision parameter. The table itself, however, does not allow directly understanding the relationship between the conditional and decision attributes of the described objects. Therefore, further processing is required to extract dependencies. In rough sets, this function is performed by the operation of creating reducts and then the decision rules. The reduct of a given information system is a set of attributes that allows distinguishing between object pairs in the information system [27][28]. This means that the reduct is the minimum subset of attributes that can be used to reflect the characteristics of the entire set.

The form of the derived rules by the rough set system is as follows:

(attribute_1)=(grade_1)and ...and (attribute_k)=(grade_k) => (sound_quality_i)=(grade_m)

Rules may not be equal - the rough set measure of the rule describing concept X is the ratio of the number of all examples from the concept X correctly described by the rule [27]:

$$\mu_{rs} = \frac{|X \cap Y|}{|Y|} \quad (3)$$

Post-print of: Korvel G., Kostek B.: Discovering Rule-Based Learning Systems for the Purpose of Music Analysis. Journal of the Acoustical Society of America. Vol. 146, iss. 4, 2947 (2019). DOI: https://doi.org/10.1121/1.5137237

where: X - is the concept, and Y - the set of examples described by the rule

The RSES system is a software tool with a visual interface to perform data explorations experiments (see Fig. 9) [31]. It comprises a decision table as well as rules derived from the data analyzed.

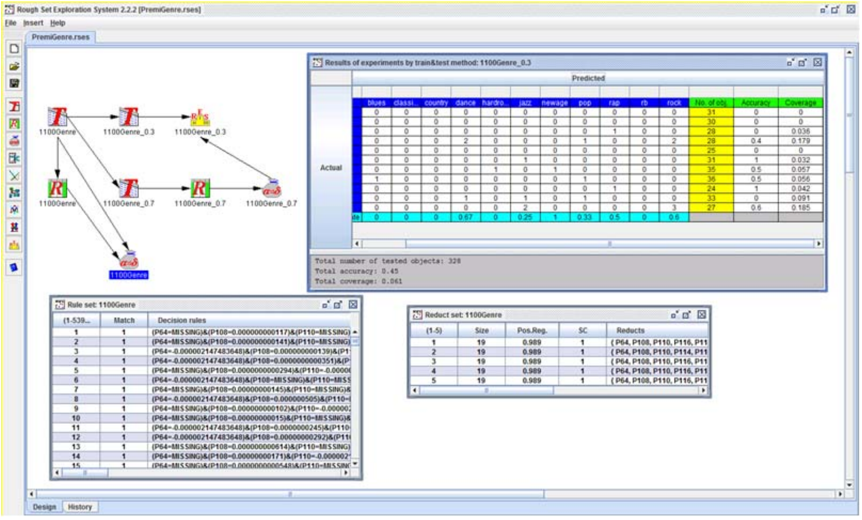


Figure 9. RSES – rule-based classifier based on the rough set method [30].

B. MUSIC GENRE RECOGNITION

Below, some results of music classification by means of two machine-learning algorithms (k-Nearest Neighbor (k-NN) and the Rough Set method) are shown. The Synat database containing approx. 50,000 music excerpts belonging to 22 music genres was employed [16][21]. However, only 1000 music excerpts were utilized in testing [18]. The analysis comprises 26-second long music excerpts. Each audio file was classified into one of the 11 music genres: Blues, Classical, Country, DanceDj, HardRock, Jazz, NewAge, Pop, R&B, Rap, and Rock.

Feature vectors contain 173 parameters based on MPEG7 standard [26] descriptors, MFCCs (Mel-Frequency Cepstral Coefficients), and the so-called dedicated descriptors [18] (see Table 2). Parameters were normalized to range <-1 ,1>. To reduce the number of parameters the PCA (Principal Component Analysis) method was used. For k-NN the following settings were used: k = 15, metric: City-SVD), for the rule-based classifier: exhaustive algorithm, rule generation settings: local rules.

Table 2. Feature vector utilized in music genre recognition [16][18].

No.	Parameter
p1	Temporal Centroid
p2	Spectral Centroid
p3	Spectral Centroid variance
p4-32	Audio Spectrum Envelope for particular bands
p33	ASE average for all bands
p34-62	ASE variance values for particular bands
p63	averaged ASE variance
p64	average Audio Spectrum Centroid
p65	variance of Audio Spectrum Centroid
p66	average Audio Spectrum Spread
p67	variance Audio Spectrum Spread
p68-87	Spectral Flatness Measure for particular bands
p88	SFM average value
p89-108	Spectral Flatness Measure variance for particular bands
p109	averaged SFM variance

Post-print of: Korvel G., Kostek B.: Discovering Rule-Based Learning Systems for the Purpose of Music Analysis. Journal of the Acoustical Society of America. Vol. 146, iss. 4, 2947 (2019). DOI: https://doi.org/10.1121/1.5137237

p110-129	Mel-Frequency Cepstral Coefficients for particular bands
p130-149	MFCC variance for particular bands
p150 - 173	RMS Parameters

RSES system-based classification resulted in reducts and rules. Most reducts consist of 2 or 3 parameters (see Fig. 10). The occurrence of parameters in reducts is shown in Fig. 11. The number of reducts and rules grows with the number of parameters (Fig. 12). Fig. 13 shows rules derived from the data. The overall classification effectiveness is shown in Table 3 and Figure 14.

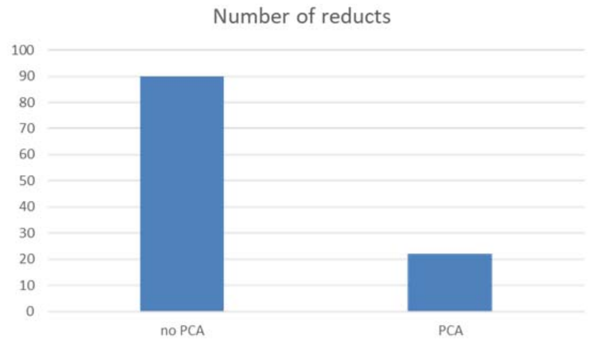


Figure 10. Number of reducts derived in the classification process

Figure 11. Number of rules generated

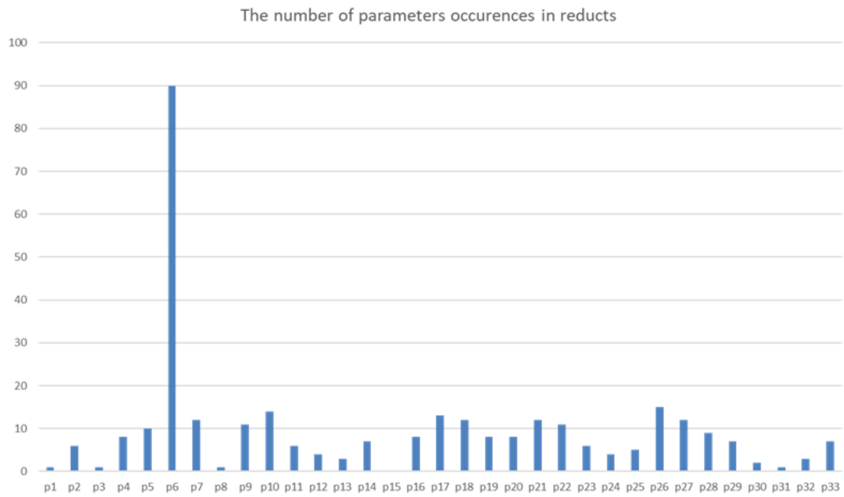


Figure 12. The occurrence of parameters in reducts.

Post-print of: Korvel G., Kostek B.: Discovering Rule-Based Learning Systems for the Purpose of Music Analysis. Journal of the Acoustical Society of America. Vol. 146, iss. 4, 2947 (2019). DOI: https://doi.org/10.1121/1.5137237

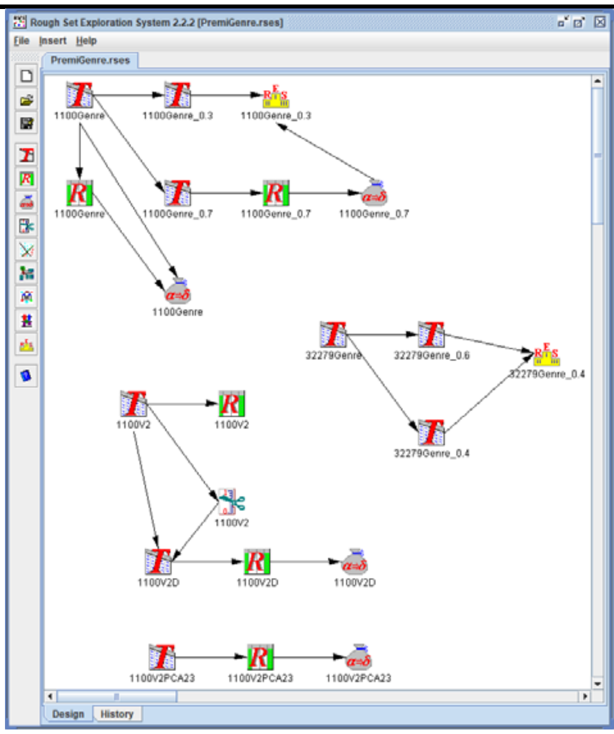


Figure 13. Rules derived from the RSES system in the case of music genre classification.

A rule derived from this analysis may have the form:

If *Temporal Centroid* is *high* and *averaged Spectral Flatness Measure variance* is high … and *RMS parameter* is high => *Rock genre.*

However, in Figs. 12 and 13, the occurrence of parameters and the reducts/rules are related to the PCA coefficients, thus they do not longer represent parameters from the feature vector shown in Table 2.

Table 3. Classification effectiveness of the rule-based and k-NN classifiers.

Genre [%]	Rule-based classifier			*k*-NN		
	no PCA	PCA	error	no PCA	PCA	error
Blues	0.844	0.917	0.073	0.744	0.818	0.074
Classical	0.909	1	0.091	0.889	1	0.111
Country	0.786	1	0.214	0.886	0.9	0.014
DanceDj	0.84	0.778	-0.062	0.8	0.727	-0.073
HardRock	0.65	1	0.35	0.75	0.905	0.155
Jazz	0.736	0.778	0.042	0.636	0.909	0.273
NewAge	0.867	1	0.133	0.767	0.909	0.142
Pop	0.788	0.87	0.082	0.688	0.861	0.173
R&B	0.739	0.767	0.028	0.639	0.757	0.118
Rap	0.585	0.783	0.198	0.485	0.871	0.386
Rock	0.749	1	0.251	0.649	0.958	0.309
Overall	0.772	0.899	0.127	0.721	0.874	0.153

Post-print of: Korvel G., Kostek B.: Discovering Rule-Based Learning Systems for the Purpose of Music Analysis. Journal of the Acoustical Society of America. Vol. 146, iss. 4, 2947 (2019). DOI: https://doi.org/10.1121/1.5137237

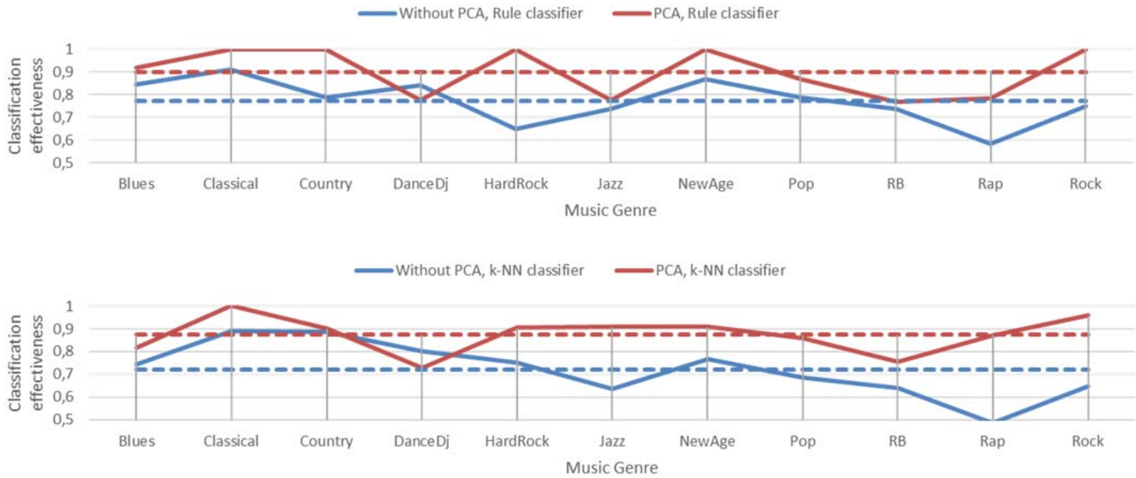


Figure 14. Classification effectiveness of the rule-based and k-NN classifiers.

4. CONCLUSIONS

The subjective test performed show that in the case of less known music genres, the subjects avoided assigning samples to these genres. This is probably due to the lack of knowledge of the definition of some genres or the ambiguity related to songs. The automatic classification of the songs confirmed the results of the subjective tests - the same samples of one song, which were classified by the listeners into two different genres, were assessed in the same way by both the Bayesian Network and SMO classifiers. This means that changing the instrumentation and the way the song is performed affects the parameters rated by listeners and classifiers.

In all machine-based test performed, the use of PCA improves the efficiency of the classification music genre by several percentage points. The rule0-based system occurred to be more effective than the baseline classifier (k-NN). „Clearly-defined” music genres have higher classification effectiveness. The results of the listening tests were confirmed by the automatic music genre classifications. It means that the instruments contained in music and the performance techniques affect similarly the way both listeners and the classifiers evaluate music. The rule-based algorithm classifies the music genre a few times longer than the k-NN algorithm, however, when the experiment goal is a thorough analysis, then it is a better solution than a black-box-type algorithm.

REFERENCES

[1] S. McAdams, S. Winsberg, S. Donnadieu, G. de Soete, and J. Krimphoff, “Perceptual scaling of synthesized musical timbres: common dimensions, specificities and latent subject classes”, Psychological Research, 58: pp. 177–192 (1995).
[2] S. Bech. N. Zacharov, “Perceptual Audio Evaluation. Theory, method and application”, Wiley (2006).
[3] J. Berg, “How do we determine the attribute scales and questions that we should ask of subjects when evaluating spatial audio quality? In Proc. Int. Workshop on Spatial Audio and Sensory Evaluation Techniques (2006).
[4] Dorochowicz, A. Majdańczuk, P. Hoffmann, B. Kostek, “Classification of musical genres by means of listening tests and decision algorithms”, ISMIS 2017, 23rd International Symposium on Methodologies for Intelligent Systems, Warsaw, Poland, 26.6.2017 - 29.6.(2017).
[5] Q. Huynh-Thu, M. N. Garcia, F. Speranza, P. Corriveau, A. Raake, "Study of Rating Scales for Subjective Quality Assessment of High-Definition Video", IEEE Transactions on Broadcasting. 57 (1): pp. 1–14 (2011). doi:10.1109/TBC.2010.2086750. ISSN 0018-9316.
[6] N. Friedman, D. Geiger, M. Goldszmidt, “Bayesian network classifiers", Machine Learning 29, pp. 139-164 (1997).
[7] K. Hevner, “Experimental studies of the elements of expression in music”, American Journal of Psychology, Vol. 48, pp. 246-268 (1936).
[8] P. Hoffmann, B. Kostek, “Bass Enhancement Settings in Portable Devices Based on Music Genre Recognition”, JAES Vol. 63, 12, pp. 980-989, December (2015), http://dx.doi.org/10.17743/jaes.2015.0087.
[9] https://mpeg.chiariglione.org/
[10] X. Hu, S. J. Downie, C. Laurier, M. Bay, A. F. Ehmann, “The 2007 MIREX audio mood classification task: Lessons learned,” Proceedings of ISMIR, Philadelphia, PA, USA, pp. 462–467 (2008).
[11] ITU-T Rec. P.10 Vocabulary for performance and quality of service (2006).

[12] ITU, International Telecommunications Union [https://www.itu.int/en/Pages/default.aspx][https://www.itu.int/pub/T-HDB-QOS.02-2011.
[13] S. Jamieson, "Likert scales: how to (ab) use them," Medical education, 38.12, pp. 1217-1218 (2004).
[14] M. Kahrs, K. Brandenburg, "Applications of Digital Signal Processing to Audio and Acoustics", Springer Science & Business Media (2006).
[15] D. Ko, W. Woszczyk, "Virtual Acoustics for Musicians: Subjective Evaluation ENGINEERING REPORTS of a Virtual Acoustic System in Performance of String Quartets," J. Audio Eng. Soc., Vol. 66, 9, pp. 712–723, (2018 September). DOI: https://doi.org/10.17743/jaes.2018.0038
[16] Kostek, P. Hoffmann, A. Kaczmarek, P. Spaleniak, "Creating a Reliable Music Discovery and Recommendation System", Springer Verlag, 107-130, XIII (2013).
[17] B. Kostek, "Observing uncertainty in music tagging by automatic gaze tracking", 42nd International Audio Eng. Soc. Conference, Ilmenau, Germany, July 22-24 (2011).
[18] B. Kostek, A. Kuprjanow, P. Zwan, W. Jiang, Z.W. Raś, M. Wojnarowski, J. Swietlicka, "Report of the ISMIS 2011 Contest: Music Information Retrieval, Foundations of Intelligent Systems," Lecture Notes in Computer Science (LNCS, 6804), Berlin, Heidelberg: Springer Berlin Heidelberg, 715–725 (2011), DOI: 10.1007/978-3-642-21916-0_75.
[19] B. Kostek, M. Plewa, "Parametrization and Correlation Analysis Applied to Music Mood Classification", Int. J. Computational Intelligence Studies, Inderscience Publishers, pp. 4-25 (2013), https://doi.org/10.1504/IJCISTUDIES.2013.054734.
[20] B. Kostek, M. Plewa, "Testing a Variety of Features for Music Mood Recognition", 166th Meeting Acoustical Soc. of America, No. 5, vol. 134, pp. 3994, San Francisco, USA, 2.12.2013 - 6.12.(2013).
[21] B. Kostek, M. Plewa, "Rough Sets Applied to Mood of Music Recognition", Federated Conference on Computer Science and Information Systems, vol. ISBN 978-83-60810-90, pp. 71 - 78, Gdansk, Poland, 11.9.2016 - 14.9.(2016), DOI: 10.15439/2016F548.
[22] Laurier, M. Sordo, J. Serra, P. Herrera, "Music Mood Representations from Social Tags", Proc. 10th International Society for Music Information Conference, Kobe, Japan, pp. 381-386 (2009).
[23] T. Letowski, "Sound quality scales and systems" (1995).
[24] [https://www.itu.int/en/Pages/default.aspx]
[25] MIREX 2009 Mood Multi Tag Data Description, http://www.music-ir.org/archive/papers/Mood_Multi_Tag_Data_Description.pdf
[26] MPEG 7 standard, http://mpeg.chiariglione.org/standards/mpeg-7
[27] Z. Pawlak, "Rough Sets", International J. Computer and Information Sciences, 11, pp. 341–356 (1982).
[28] J. F. Peters, A. Skowron, A. (Eds.): Transactions on Rough Sets, Lecture Notes in Computer Science, vol. 4100, Springer (2004–2019).
[29] M. Plewa M., B. Kostek, "Creating Mood Dictionary Associated with Music", 132 Audio Engineering Society Convention, preprint 8607, Budapest, Hungary, 26.4.2012 - 29.4.(2012).
[30] "Practical procedures for subjective testing", Handbook (2012) https://www.itu.int/pub/T-HDB-QOS.02-2011]
[31] "RSES 2.1. Rough Set Exploration System", User's handbook. http://logic.mimuw.edu.pl/~rses/RSES_doc.pdf. Warsaw (2004).
[32] J. A. Russel, A circumplex model of affects, Journal of personality and Social Psychology, 39, pp. 1161-1178 (1980).
[33] R. E. Thayer, "The Biopsychology of Mood and Arousal", Oxford University Press (1989).
[34] Vincent, M. G. Jafari, M. D. Plumbley, "Preliminary guidelines for subjective evaluation of audio source separation algorithms", Proc. of ICA Research Network International Workshop, pp. 93-96 (2006).
[35] N. Zacharov and G. Lorho, "What are the requirements of a listening panel for evaluating spatial audio quality?" Proc. Int. Workshop on Spatial Audio and Sensory Evaluation Techniques (2006).
[36] S. Zielinski, F. Rumsey, S. Bech, "On some biases encountered in modern audio quality listening tests-a review." J. Audio Eng. Soc. 56.6: 427-451 (2008).