Music information retrieval—The impact of technology, crowdsourcing, big data, and the cloud in art

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

The exponential growth of computer processing power, cloud data storage, and crowdsourcing model of gathering data bring new possibilities to music information retrieval (MIR) field. Mir is no longer music content retrieval only; the area also comprises the discovery of expressing feelings and emotions contained in music, incorporating other than hearing modalities for helping this issue, users' profiling, merging music with social media and qualitative recommendations in music services. Moreover, 5G telecommunications networks, characterized by "near-instant and everything in the vicinity talks with one another," with exponentially faster download and upload speeds, may change the existing models and create a new age of interconnectedness. This paper aims at showing some of the already highly exploited technologies and crowdsourcing models applied to music processing. Several studies are discussed in details, such as, e.g., deep learning applied to music, a way to generate an expanded training sets using 2-d data such spectrograms, mel-cepstrograms, chromagrams, and waveform-based representations of the signal instead of feature vectors in machine learning, allowing to retain all nuances related musical articulation in the signal. Also, a discussion is to be outlined, expanding the issue of the impact of these new technologies on the artistic and aesthetic values of music.

1. INTRODUCTION

This paper aims at showing some of the already well-exploited technologies applied to music processing. Several issues are discussed, such as, e.g., music services, deep learning applied to music, a way to generate an expanded training sets using 2D data such spectrograms, mel-cepstrograms, chromagrams, and waveform-based representations of the signal instead of feature vectors in machine learning, allowing to retain all nuances related to musical articulation in the signal. Also, a discussion is to be outlined, expanding the issue of the impact of these new technologies on the artistic and aesthetic values of music.

The starting point is to show, however, to what extent the latest technologies contribute to science, and in particular to musical acoustics and music processing. When addressing these technologies point by point, computer processing power, cloud data storage, and crowdsourcing model of gathering data, 5G telecommunications networks, one can see that they affect the music area enormously. First of all, computer processing brought instant search through data storage that nowadays is organized as a cloud. Moreover, this instant access is helped by faster communication means, such as 5G. Also, computing power or computational power available today is an important factor in music-related data processing and access. Definitions of computing power either refer to the number of operations that a computer can carry out in one second or to the speed that instructions are carried out, expressed in terms of kiloflops, megaflops, etc. Also, one can find that CPUs is based on their speed per Watt of power, because the cost of powering the CPU outweighs the cost of the CPU itself. The point is that, nowadays, one can perform tasks that were not available for an individual

computer user even recently. In musical acoustics, it should be regarded as a possibility to apply deep learning for music analysis, music retrieval, music synthesis, etc., based on a personal computer that has a sufficient number of graphics processing units (GPU). This is especially important as music is often processed nowadays not as an audio signal but as two-dimensional images. Secondly, one cannot ignore the resources that are behind social services. It encompasses both music and users. They can be defined as collaborative technology users, i.e., listeners who nowadays not only listen to music but deliver data for the music services to customize their services based on the feedback received. There are also persons willing to create new sets of music data in the form of crowdsourcing [2][3], which is voluntary taking part in on-line experiments and thus providing data needed [35][36]. The above-given issues will shortly be described in the following Sections.

2. MUSIC SOCIAL SERVICES

It should be noted that music services changed our way of listening to music. An overall definition of the **music network services** may be such as they are created by separate components (songs, albums, etc.,) in combinations demanded by users by the Internet. A person no longer buys a physical carrier containing music, but either upload a music item or – with streaming services – just listen to music online (see Fig. 1) [34]. Personal music tastes, mood, what music we share, what repertoire we download or stream, these are all tagged, thus we constitute an integral part of the music ecosystem that includes both music resources and its users. Characteristics and functionalities of the services contribute to differences between them, i.e.: **streaming services**, **socio-musical services** and **distribution services,** the last ones intended mainly for debuting artists [25]. A very short but at the same time, up-to-date review of music services is given by J. L. Wilson, PCMag editor in a review entitled: "The Best Online Music Streaming Services for 2020" [15]. Music services that are recalled and discussed are as follows [15]: SiriusXM Internet Radio [17] is given with characteristics such as follows: "Internet Radio's crisp audio, numerous live stations, and talk radio are a must-have for radio-streaming fans…". Then, Tidal [19] is "high-quality audio, music-related articles… is one of the best and most unique streaming audio services around". Amazon Music [5] with its large music library, a possibility to scroll lyrics, Alexa-assistance, etc., is also an important service. Then, goes Apple Music [7] with its more than 50 million tracks, Siri assistance and Apple watch integration. The main characteristics of Spotify [18], which has over 60 million active users and 15 million subscriptions, are as follows: "based on collaborative playlists, allowing to store audio files locally, and showing new music albums before they are released." This is one of the services that a user may follow the activities of other users associated with him/her. It operates on two layers: (1) **general map of relationships between songs** and (2) **personalization layer**. All data are tagged and interconnected and contained in the Spotify cloud. Processing these two sets of dependencies results in a weekly playlist. On this basis Release Radar (or Spotify's Discover Weekly) [18], a playlist of new releases recommended just for the particular user, is issued weekly. The statistical data that the user may see wrapped for the whole year may contain such numbers as 49,200 minutes, 3,381 different songs 1,548 different artists and along the way, 34 genres explored (this information was retrieved from a student's personal data).

Moreover, albums and artists listed are not a random selection of the algorithm, but they result from a careful analysis prepared by music journalists.

There are other players in music distribution. One of them is LiveXLive powered by Slacker [13], that combines "live music streams with Slacker Radio's DJs and deep music well to produce one of the most complete streaming music services on the market today." Then, Deezer [9], a streaming service, that contains a library of 53 million songs. A unique characteristic of Google Play Music [28] is that it integrates other services such as YouTube, thus linking music to videos. Moreover, last year, Google's Magenta [16] research department developed NSynth Super, an open-source synthesizer controlled by the NSynth algorithm, designed to create completely new sounds by learning about existing acoustic properties. Thus, computer-assisted music generation, known since 1990, appears as a new feature in music services. Machine learning is used in the art creation process.

Pandora [14], on the other hand, is characterized as „the once-pioneering streaming music platform finally has the feature set to compete with rival services". Also, this service is partly based on annotators' tags assigned to the music item. Finally, iHeartRadio [11] is seen as a service that "merges live radio and a curated catalog". Besides, there is also Last.fm [12], a very large social music recommending platform, which ranks album songs and artists. With this service it is possible to retrieve a chart of one's history of activities during a time; however, more detailed statistics are also given referring to the most popular tags or albums heard, new music releases

listened to, etc. This information is linked to the clock that shows when during a given day "listening activity" happens. Besides, it shows people having a similar music taste. This is an example of using feedback from the user's activity based on the so-called scrobbling, which is tracking information sent automatically to the system. All the above create **big data** in music, provided by tagging music and its users' behavior and actions.

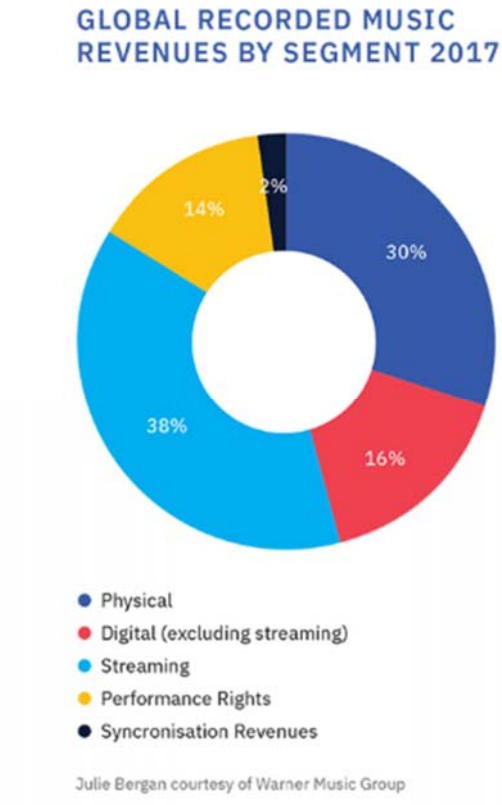

Figure 1. Music data divided by revenue segments **[34].**

Even though the above review shows that information retrieved from music and its listeners create a well-integrated ecosystem, there are still some basic problems in the music information retrieval (MIR) area, which are unsolved or only partially solved.

The user-based annotation is utilized in predicting the user's music preference, but human-based evaluation does not always give correct answers or is far from giving correct answers. However, **we expect better machine-learning than human-based performance…**

There is also a key problem related to the scalability of the proposed solutions, regardless of the type of application (research- or commercial-based). The last problem is addressed by deep learning, computing power, and music analyzed as 2D maps. This may solve the so-called sparsity of data related to collaborative filtering [21][25][27][31]. Because of the sparsity of the data, considering the co-occurrence statistics is unreliable. If there are millions of songs in a music service, then even very active users are not able to listen to 1% of the music sources. Thus this may result in unreliable recommendations [21][32].

Challenges that could be identified within the music technology area are related to the role of human factors such for example as the user's personality and experience, emotions, in the user's models and personalized services. A question arises whether the music creation process is still an art, especially as music created or rather – more adequately said - produced automatically is satisfactory to many users. There are two separate research trends to be discerned with the automatic process creation [10]. One has its roots in the area of musicology and focuses on algorithmic techniques of composing music. The second trend, in turn, focuses on applications, e.g., for automatic accompaniment. So some groups try to automate the creative process, while others create tools to support it.

Amper Music [6] brings this concept to the 21st century: it is a service that uses deep teaching to automatically compose music into multimedia materials based on the choice of "style" or "user mood" [6]. With

the support of artificial intelligence (AI) whole songs are created. Another technology example maybe Audionamix Xtrax Stems 2 [8], which is both machine-based and cloud-based software that enables us to fully separate mixed stereo recordings into three vocals, drums and melodic layer that can then later be used for remixes. Computer-based or including artificial intelligence in the process of creating music becomes a reality.

3. MACHINE-BASED MUSIC PROCESSING EXAMPLE

Machine learning is a term referring to specific techniques within the broader concept of artificial intelligence, allowing the system to find patterns in more massive data sets or make decisions in response to the occurrence of previously unseen data.

A. MUSIC GENRE RECOGNITION SYSTEM

The main component of musical genres recognition systems is the optimized parametrization block [20][22][23][26][29]. The prepared feature vector (FV) in this block should have a very good separability between parameters. Taking into account these assumptions, the feature vector containing 173 elements was conceived in earlier research studies carried out by the author and her collaborators [23][24][30]. A collection of 52532 music excerpts described with a set of descriptors obtained through the analysis of mp3 recordings was gathered in a database called SYNAT [24]. The SYNAT database was realized by the Gdansk University of Technology (GUT) [24]. For the recordings included in the database, the analysis band is limited to 8 kHz due to the music excerpts format, this means that the frequency band used for the parameterization is in the range from 63 to 8000 Hz. The prepared feature vector is used to describe parametrically each signal frame. The database stores 173-feature vectors, which in the majority are the MPEG-7 standard parameters [30]. The vector has additionally been supplemented with 20 Mel-Frequency Cepstral Coefficients (MFCC), 20 MFCC variances and 24 time-related 'dedicated' parameters. The vector includes parameters associated with the MPEG-7 standard, mel-cepstral (MFCC) parameters and is enlarged by the so-called dedicated parameters which refer to a temporal characteristic of the analyzed music excerpt, their names are included in Table 1. The list of parameters and their definitions were shown in the earlier study [4], however, it is worth noting that the proposed FV was used in the ISMIS 2011 contest in which there were over 120 participants [23]. The best contest result returned almost 88% of accuracy, and later in the authors' own study gained even better effectiveness [4]. Various configuration of the number of music tracks was used in experiments.

***Table 1 The list of parameters within the SYNAT music database* [23][30].**

No.	Parameter
1	Temporal Centroid
2	Spectral Centroid
3	Spectral Centroid variance
4-32	Audio Spectrum Envelope for particular bands
33	ASE average for all bands
34-62	ASE variance values for particular bands
63	averaged ASE variance
64	average Audio Spectrum Centroid
65	variance of Audio Spectrum Centroid
66	average Audio Spectrum Spread
67	variance Audio Spectrum Spread
68-87	Spectral Flatness Measure for particular bands
88	SFM average value
89-108	Spectral Flatness Measure variance for particular bands
109	averaged SFM variance
110-129	Mel-Frequency Cepstral Coefficients for particular bands
130-149	MFCC variance for particular bands
150	number of samples exceeding RMS
151	number of samples exceeding 2×RMS
152	number of samples exceeding 3×RMS

153	mean value of samples exceeding RMS, averaged for 10 frames
154	variance value of samples exceeding RMS, averaged for 10 frames
155	mean value of samples exceeding 2×RMS, averaged for 10 frames
156	variance value of samples exceeding 2×RMS, averaged for 10 frames
157	mean value of samples exceeding 3×RMS, averaged for 10 frames
158	variance value of samples exceeding 3×RMS, averaged for 10 frames
159	peak to RMS ratio
160	mean value of the peak to RMS ratio calculated in 10 subframes
161	variance of the peak to RMS ratio calculated in 10 subframes
162	Zero Crossing Rate
163	RMS Threshold Crossing Rate
164	2×RMS Threshold Crossing Rate
165	3×RMS Threshold Crossing Rate
166	Zero Crossing Rate averaged for 10 frames
167	Zero Crossing Rate variance for 10 frames
168	RMS Threshold Crossing Rate averaged for 10 frames
169	RMS Threshold Crossing Rate variance for 10 frames
170	2×RMS Threshold Crossing Rate averaged for 10 frames
171	2×RMS Threshold Crossing Rate variance for 10 frames
172	3×RMS Threshold Crossing Rate averaged for 10 frames
173	3×RMS Threshold Crossing Rate variance for 10 frames

173-element vector generates a very large amount of information describing a given track. This is why Principal Component Analysis (PCA) was applied to reduce the data redundancy as it transforms a number of possibly correlated variables into a smaller number of variables called principal components. To quantify the redundancy between data is to use the variance of the data to prepare a new set of parameters. The new components are a linear combination of parameters that carry most information about the test set., thus they are no longer refer to descriptors contained in the original feature vector. The PCA method can shorten the feature vector of 173 elements to a dozen parameters, or even to 19 components, which significantly reduces the computation time and efficiency. The *k*-NN algorithm effectiveness was compared with the Bayesian networks algorithm and Support Vector Machines using Sequential Minimal Optimization (SMO) to obtain the highest efficiency of classification [4]. In Fig. 2, the results of the classification of genres obtained on 32110 tracks, selected from the SYNAT database [4], are shown.

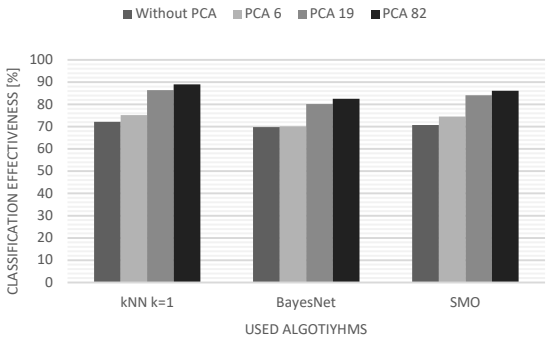

Figure 2. The average effectiveness of k-NN, Bayesian networks, SMO classification algorithms using the PCA ***method on 32110 tracks database.***

Besides the previously outlined average accuracy results Precision, Recall and F-Measure metrics were analyzed and they are shown in Fig. 3 [4].

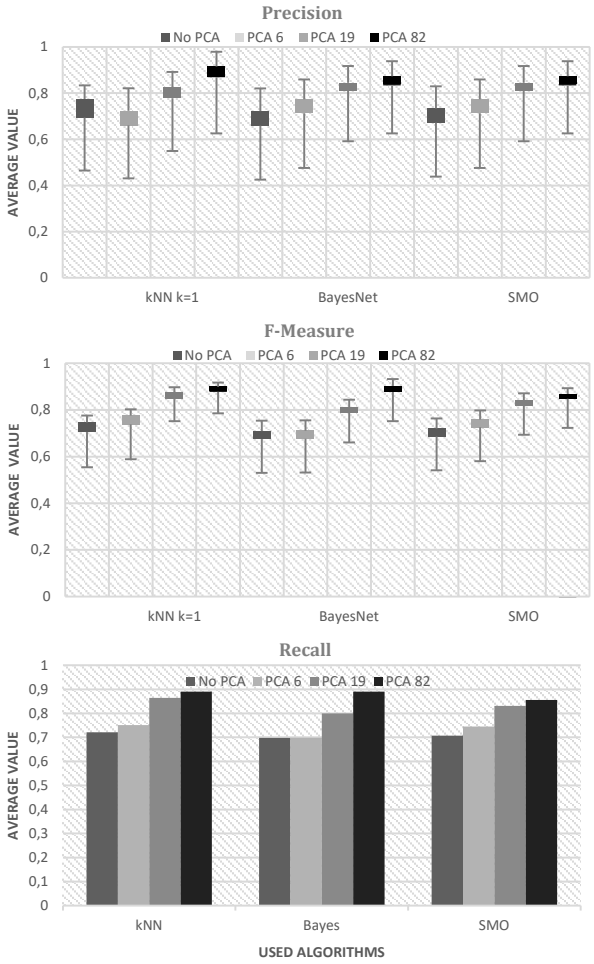

Figure 3. Precision, Recall and F-Measure values for music genre recognition based on a feature vector.

B. DEEP LEARNING-BASED MUSIC GENRE CLASSIFICATION EXAMPLE

Even though classification based on machine learning techniques when using a one-dimensional feature vector is not directly comparable to the employing deep learning algorithm and two-dimensional data, still it is interesting to see the outcome of both approaches. To carry out the experiment with the deep learning algorithm, an implementation of the Convolutional Neural Network (CNN) configured to recognize musical genres was used [1][33]. The classifier implementation was developed in Python using the popular deep learning library TensorFlow [33]. The structure of the network was designed in a trial-and-error manner, by introducing new network layers to the point in which there was no improvement of the network performance.

At the input of the classifier, 29-second excerpts of songs are given in the form of spectrograms on a mel scale. A 2D representation of an audio file may be obtained by calculating a distribution in time as a horizontal axis and MFCC (Mel Frequency Cepstral Coefficients) as a vertical axis. MFCC was chosen as a pre-processing method of recordings because they reflect a logarithmic way of sound perception, which is the case under consideration. The neural network used was prepared using five layers of networks. The algorithm input layer has 96 mel bands and 1366 signal frames organized in two network layers. At the output of the neural network, a one-dimensional vector is generated consisting of 10 categories identical to the tested musical genres. The values assigned to each category indicate the degree of belonging of the analyzed fragment of the work to the musical genre.

The process of recognizing music genres based on deep learning begins with data training. Training and test data in this experiment came from the GZTAN collection and were divided in a ratio of 90:10. The experiment was carried out twice, the first time the collection included 10 genres, the second time six. The network training has been started for 3000 iterations. However, the stability of the results obtained could already be observed at

the 500 epochs. The best efficiency of genre classification was achieved for the 2010 epochs and amounted to 87% for 10 music genres and for 6 music genres 89.17%.

The recognition efficiency of individual genres is presented in Figs. 4 and 5. The average value of the classification is presented as the orange line. For a set with 10 tested genres, blues, country, hip-hop, metal, pop, and rock have been recognized without error. The lowest effectiveness was recorded for disco and reggae genres. For a set with 6 genres, the efficiency increased slightly. The disco and classical genres were recognized more effectively. In the case of jazz and pop, there was a decrease compared to the collection of 10 genres. An algorithm that uses deep learning has proved to be an effective solution. Its use allows for effective recognition of music genres at no worse level than the k-NN algorithm.

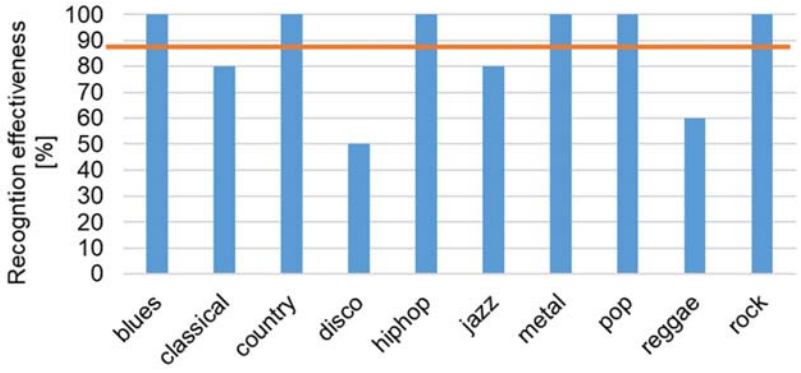

Figure 4. Classification efficiency for artificial neural networks for 10 music genres, based on CNN.

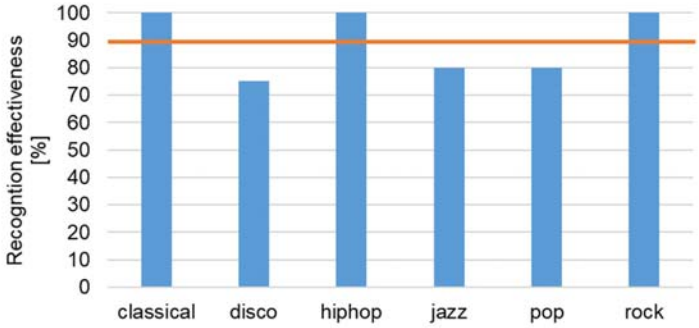

Figure 5. Classification efficiency for artificial neural networks for 6 music genres, based on CNN.

4. CONCLUSIONS

It is worth to notice that music understood very broadly, that encompasses art-related aspects, music analysis, music processing, music synthesizing, music classification, music services, etc., utilizes all state-of-the-art technology achievements. Some of the examples of the technology usage were recalled in this paper; however many other usage possibilities were not included in this review.

A session entitled Machine Learning in Musical Acoustics presented at the San Diego ASA Meeting, co-organized by the author and Scott Hawley, brought researches and audience for all over the world and was well-received by the audience. The session contained the following presentations: Automated Object Detection of Antinode Regions in Oscillating Steelpan Drum; Learning the nuance of musical instrument acoustics; Music Information Retrieval – the Impact of Technology, Crowdsourcing, Big Data, and the Cloud in Art; Convolutional Neural Network for Chordophones Recognition; PhonoNet: Multi-Stage Deep Learning for Raga Preservation in Hindustani Classical Music; Discovering Rule-Based Learning Systems for the Purpose of Music Analysis. It seems that the topics presented at this session will have a follow-up.

REFERENCES

[1] K. Choi, G. Fazekas, M. Sandler, K. Cho, "Convolutional recurrent neural networks for music classification," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, 2392–2396 (2017), DOI:10.1109/ICASSP.2017.7952585.

[2] A. Chamberlain, M. Bødker, and K. Papangelis, "Sounding Out Ethnography and Design: PAPERS Developing Metadata Frameworks for Designing Personal Heritage Soundscapes," J. Audio Eng. Soc., 66, 6, 468–477, (2018 June), DOI: https://doi.org/10.17743/jaes.2018.0025.

[3] C. Gomez, D. Schneider, K. Morales, J. Moreira De Souza, "Crowdsourcing for music: Survey and taxonomy," Conference: Systems, Man, and Cybernetics (SMC), (2012) IEEE International Conference on, DOI: 10.1109/ICSMC.2012.6377831

[4] P. Hoffmann, B. Kostek, "Bass Enhancement Settings in Portable Devices Based on Music Genre Recognition," J. Audio Eng. Soc., 63, 12, 980–989 (2016), DOI: 10.17743/jaes.2015.0087.

[5] https://music.amazon.com/home (accessed December '2019)

[6] https://www.ampermusic.com/music/ (accessed December '2019)

[7] https://www.apple.com/itunes/music/ (accessed December '2019)

[8] https://audionamix.com/technology/xtrax-stems/ (accessed December '2019)

[9] https://www.deezer.com/ (accessed December '2019)

[10] https://www.izotope.com/en/learn/what-the-machine-learning-in-rx-6-advanced-means-for-the-future-of-audio-repair-technology.html (accessed December '2019)

[11] https://www.iheart.com/ (accessed December '2019)

[12] http://www.last.fm/ (accessed December '2019)

[13] https://www.livexlive.com/ (accessed December '2019)

[14] http://www.pandora.com (accessed December '2019)

[15] https://www.pcmag.com/picks/the-best-online-music-streaming-services (accessed December '2019)

[16] https://research.google/teams/brain/magenta/ (accessed December '2019)

[17] https://www.siriusxm.com/streaming (accessed December '2019)

[18] https://www.spotify.com/ (accessed December '2019)

[19] https://tidal.com/ (accessed December '2019)

[20] K. Hyoung-Gook, N. Moreau, T. Sikora, "MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval," Wiley & Sons (2005).

[21] J.L. Konstan, L.G. Terveen, J.T. Riedl, "Evaluating Collaborative Filtering Recommender Systems Herlocker," ACM Transactions on Information Systems, 22, 1, Jan. (2004).

[22] B. Kostek, "Music Information Retrieval – Soft Computing Versus Statistics, Computer Information Systems and Industrial Management," 9339, K. Saeed, W. Homenda, Red. Cham: Springer International Publishing, 36–47 (2015), DOI: 10.1007/978-3-319-24369-6_3

[23] 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.

[24] B. Kostek, P. Hoffmann, P. Spaleniak, A. Kaczmarek, "Creating a Reliable Music Discovery and Recommendation System, Intelligent Tools for Building a Scientific Information Platform: From Research to Implementation," Springer Verlag (2014).

[25] B. Kostek, "Listening to Live Music: Life beyond Music Recommendation Systems," 2018 Joint Conference - Acoustics, Acoustics 2018, 134 - 139, Ustka, PL, 11.9.2018 - 14.9.(2018), DOI: 10.1109/ACOUSTICS.2018.8502385.

[26] T. Lindsay, J. Herre, "MPEG-7 and MPEG-7 Audio - An Overview," J Audio Eng Soc, 49, 7/8, 589–594 (2001).

[27] X. Mu, Y. Chen, T. Li, "User-Based Collaborative Filtering Based on Improved Similarity Algorithm," Proc. of the 3rd IEEE International Conference on Computer Science and Information Technology, 8, Chengdu, China, 76-80 (2010).

[28] play.google.com/music

[29] A. Rosner, B. Schuller, B. Kostek, "Classification of Music Genres Based on Music Separation into Harmonic and Drum Components," Archives of Acoustics, 39, 4, 629–638, 2015, DOI: 10.2478/aoa-2014-0068.

[30] A. Rosner, B. Kostek, "Automatic music genre classification based on musical instrument track separation," Journ. of Intelligent Information Systems, 5 (2017), DOI: doi:10.1007/s10844-017-0464-5.

[31] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, “Item-Based Collaborative Filtering Recommendation Algorithms,” Proc. 10th international conference on World Wide Web, New York, NY, USA, 285-295 (2001).
[32] M. Schedl, H. Zamani, CW. Chen, Y. Deldjoo, M. Elahi, Current challenges and visions in music recommender systems research, International Journal of Multimedia Information Retrieval, 7, 2, 95-116 (2018), https://doi.org/10.1007/s13735-018-0154-2.
[33] M. P. Shah, “Tensorflow Implementation of Convolutional Recurrent Neural Networks for Music Genre Classification,” meetshah1995/crnn-music-genre-classification (2018).
[34] “U.S. Music Industry Sees First Double Digit Growth in Almost 20 Years as Streaming Takes Over,” https://www.billboard.com/articles/business/7744268/riaa-us-music-industry-2016-revenue-double-digit-growth (accessed December ‘2019).
[35] L. Vrysis, N. Tsipas, C. Dimoulas, G. Papanikolaou, “Crowdsourcing Audio Semantics by Means of Hybrid Bimodal Segmentation with Hierarchical Classification,” J. Audio Eng. Soc., 64, 12, 1042-1054, December (2016), DOI: https://doi.org/10.17743/jaes.2016.0051.
[36] N. Vryzas, R. Kotsakis, A. Liatsou, C. Dimoulas and G. Kalliris, “Speech Emotion Recognition for Performance Interaction,” J. Audio Eng. Soc., 66, 6, 457–467, (2018 June), DOI: https://doi.org/10.17743/jaes.2018.0036.