Listening to Live Music: Life Beyond Music Recommendation Systems

Bożena Kostek
Gdansk University of Technology
Faculty of Electronics, Telecommunications and Informatics
Audio Acoustics Laboratory
Gdańsk, Poland
bokostek@audioacoustics.org

***Abstract*—This paper presents first a short review on music recommendation systems based on social collaborative filtering. A dictionary of terms related to music recommendation systems, such as music information retrieval (MIR), Query-by-Example (QBE), Query-by-Category (QBC), music content, music annotating, music tagging, bridging the semantic gap in music domain, etc. is introduced. Bases of music recommender systems are shortly presented, including the mechanisms underlying these systems. Also, usage of machine learning versus statistics is discussed with regard to the recommender systems working. Moreover, listening to music through players implemented on computers or mobile devices as opposed to listening to live music in the context of social and technology implications, i.e. live performance contrasting issues related to music quality. Finally, future directions in the music recommendation area and live music are discussed, including performance on virtual musical instruments.**

Keywords—Music Information Retrieval (MIR), music annotation, music recommendation systems, live music, virtual musical instruments

I. INTRODUCTION

Imagine that you listen to what you like, and no one is suggesting what this should be… It is said that a prerequisite for the Internet of Things (IoT) is tagging. In many known IoT scenarios this simply refers to Radio Frequency Identification (RFID) technology and sensors, however tagging means also digital watermarking, barcodes, QR (*Quick Response*) codes and ID3 tags. The last notion expands the IoT usage to music technology, then we can perceive the whole music ecosystem as an inventory controlled by computer-based technology, which includes not only music but also its users. One should be aware that our music taste, mood, what music we share, what repertoire we download or stream, these are all tagged, thus we constitute an integral part of this ecosystem.

Over the last few decades we have seen the music industry purposefully adapting to changes in technologies. It is said that in 2017, for the first time, streaming and downloads revenue outweighed physical music sales such as CDs and vinyl. In fact music streaming comprises streaming socio-musical platforms (e.g. Spotify, Apple Music, Tidal, Deezer), musical social systems (e.g. last.fm, Rate Your Music, etc.) and music distribution services (e.g. Soundcloud, Bandcamp), and above all, hundreds of millions of users generating income as account holders or profits from subscriptions and advertisements. There are also some interesting facts: streaming takes place at the expense of downloading music from the Internet, for example it is for the first time the leading driver of revenue for the U.S. recorded music business, whereas downloading of albums compared to 2016 decreased by approx. 15%. On the other hand, vinyl sales increased by 20% compared to 2016 and it already accounts for 10% of sales of all physical media [1].

An overall definition of the network services may be such as they are created by separate components (songs, albums, etc.,) in combinations demanded by users. Looking at the characteristics and functionality side of these services, one may easily discern differences between streaming services (high availability of music, personalized radio stations, music recommendation, the ability to add friends and watching them, creating and sharing playlists, subscription fee), socio-musical services (pages of artists, albums, songs, events, the ability to track music listened to by the user (i.e. scrobble [2]), the possibility of discussing music, recommendation based on the user's taste, comparing one's music taste with other users, providing statistical data, free of use) and distribution services intended mainly for debuting artists (the ability to buy music and gadgets of bands/artists, comments and reviews of listeners, music recommendation system, creating playlists, watching artists). It is also interesting to compare mechanisms used in some large contenders, to name a few: Spotify, Google Play Music, Apple Music, Deezer which are listed among over 105 streaming websites in the world [3-6].

Spotify service of over 60 million active users and 15 million subscriptions, available on iOS, Windows, Android and BlackBerry platforms, operates on two layers [3]: (1) general map of relationships between songs and (2) personalization layer. The first one analyzes all playlists of all Spotify users, how they choose them, what songs are played and how they use other functions of the application. This enables a non-stop analysis in the Spotify cloud based on tags or other information to know what are interconnections between the users' and songs, who is completely unmatched, etc. In the personalization layer all the conclusions from the first analysis are confronted with the user's musical preferences, not only what the particular user listens to but what songs he/she likes to combine. Processing these two sets of dependencies results in a weekly playlist. On this basis Release Radar (or Spotify's Discover Weekly), a playlist of new releases recommended just for the particular user, is issued weekly. Albums and artists listed are not a random selection of the algorithm, but they result from a careful analysis prepared by music journalists, or people who know what new and noteworthy appears on the market [3]. The statistical data that one may see wrapped for the whole year may contain such numbers: 49,200 minutes, 3,381 different songs 1,548 different artists and along the way 34 genres explored.

978-1-5386-7114-6/18/$31.00 ©2018 IEEE

Another example of such services is Google Play Music [4]. It is said that the Android version of iTunes includes over 20 million available songs. The user can create an own music collection (up to 50,000 songs). Music recommendation system is based on playback history, social media activity and the user's preferences expressed in the service application. Lists of other users and recommendations for individual tracks are available for a particular user. A unique feature of the service is the possibility of adding own recordings and buying additional songs in the Google Play store [4]. Apple Music is an integrated system that combines streaming, Internet radio and a social platform that enables one to track the activity of the user's favorite artists [5], and the collaborative (create and share playlist) feature is not available. All features mentioned are especially of interest to the iTunes users [6]. It is reported that Deezer is the largest library of recordings, comprising over 35 million songs. The service differentiates between subscribers, the Premium version enables one to access higher-quality music files (320 kbps). It is available for downloading files and streaming recordings and can be used on SmartTV TVs and car-audio systems [7].

In this review paper, some notions related to music recommendation system based on social collaborative filtering are presented. Terms related to music recommendation systems, e.g. music information retrieval (MIR) [8], *Query-by-Example* (QBE), *Query-by-Category* (QBC), music content, music annotating, bridging semantic gap in music domain, etc. are introduced. Background of music recommender systems is shortly presented, including mechanisms underlying these systems. Also, usage of machine learning versus statistics is shortly discussed with regard to the recommender systems working. Moreover, listening to music through players implemented on computers or mobile devices is opposed to listening to live music in the context of social and technology implications, i.e. live performance contrasting issues related to music quality. Finally, future directions in the music recommendation area and live music are discussed, including performance on virtual musical instruments.

II. Dictionary of Music Ecosystem Terms

Annotation is generally understood as associating any element of musical content, such as note, lyric, title, music genre (metadata provided by ID3 format) with some additional information such as chords, comments, impressions. However, one may refer to the multidimensional music description, where the problem of fully annotating a music file is still far from being solved since music is very complex and exhibits a significant amount of variation. An example of such problems is assigning songs to a given genre which is often an arbitrary process and closely related genre/style overlap that cause difficulties with existing data extraction. MIR (*Music Information Retrieval*) methods, e.g. low-level (typically based on MPEG 7 standard) descriptors extracted from the audio signal [9] used in the process of music tagging coupled with machine learning return nearly perfect matches in music similarity search [10-14]. However, this is true for small databases (thousands of records), on which MIR studies are performed. Moreover there still exists the so-called semantic gap which refer to the discontinuity between different layers of music annotation, especially concerning low-level descriptors and high-level semantic representation [15]. Thus researches try to bridge that gap through their studies [16].

Tagging music data manually requires a person with a musical background, and the process takes a lot of time. It is very interesting that a new profession was created several years ago, candidates for which were sought by Pandora [17]. In the advertisement issued candidates were searched for with a strong background in music search and discovery, interest in natural language processing and signal processing as well as core machine learning and recommender systems to perform music annotation. But, very recently Spotify created Line-In, a music metadata editor, to all of its users for "suggest an edit". In this way a collaborative effort of social tagging will help music services to describe their cloud content.

When reviewing the first decade of MIR studies, one can say that the starting point was Query-by-Humming/Singing/Whistling (QBH) as this is a very convenient way to find a given melody [18-19]. Later, more advanced MIR search appeared, i.e. Query-by-Example, which was at this stage more a research- than technology-oriented trend. Certainly, in the newest approach, based on mature technology, one may search for a given music genre, mood of the person listening to music or emotion contained in a musical piece. This refers to Query-by-Category (QBC): musical style, genre, mood/emotion (content-based) [13][16][20-21]. QBC-based retrieval is exploited by the scientific community as well as collaborative filtering and mining [22], differing in their approaches, i.e. scale of music data utilized and algorithms behind the search. Overall, it may be said that the research community employs feature extraction and machine learning approach, whereas statistics underlie searches in music services.

III. Mechanisms Underlying Music Reccomending

A challenge in search solutions lies in their scalability, i.e. they are very expensive, both in time and space when operating in large feature spaces. Big data needs reducing the complexity and high dimensionality of data. The algorithmic approach to that is Locality Sensitive Hashing (LSH), a randomized algorithm for solving Near Neighbor Search [23]. Its role is not to return an exact answer but to guarantee a high probability of bringing an answer close to the correct one. Overall, a triple consisted of (User, Item and Rating) is the basis of the collaborative filtering [24-28]. The user's preference for a music item is called a rating. The (user-item) matrix X with dimensions $K \times M$, is composed of K users, and M songs. A single matrix element is described by $x_{k,m} = r$, which means that the k-th user assigns the r rate for the m-th song. In the case of the user-based collaborative filtering each raw of the matrix denoted above is sorted by its similarity towards the k-th user's profile (see Fig. 1 for graphical explanation) [22]. The set of similar users can be identified by employing a threshold or selecting a group of top-N similar users. More detailed mathematical description of this method may be found in the work by Wang *et al.* [28]. Several approaches are employed in collaborative filtering, they may belong to the memory-based such as *Pearson correlation*, *cosine* *similarity* measures, model-based methods or algorithms belonging to ranking or probabilistic model are employed [24-28]. Collaborative filtering is to some extent problematic, especially for niche artists or artists not belonging to the mainstream of recommended songs or music pieces. They may never be recommended as no connections would be found between users and such songs. In fact, it is a good question whether this concerns unreliable recommendation or just reflects reality, i.e. sparse user-item matrix [22].

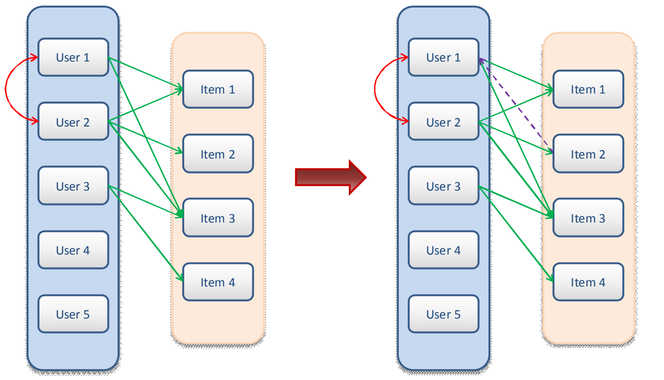

Fig. 1. An example of the user-based preference prediction for music items; a green line refers to the user's choice of songs, a red arc shows whether two users are interconnected by a given song, then in a consequence the user no. 2 may want to listen to another music item (violet dashed line) already listened to by the user no. 1.

IV. Live Music today

A very fundamental question remains open whether quality of music still is important. One may have an impression that this is no longer an issue as millions of users download and stream music of low quality. On the other hand live music events, both concert performances and reinforced events gather thousands of people. Moreover, it brings new technology to live reinforced music.

Present-day spatial audio formats (e.g. MPEG-H Audio Coding, HOA (*Higher Order Ambisonics*, etc.), emerged recently, focus on 3D audio reproduction [29]. The typical surround 5.1 loudspeaker layout (home theater format) refer to the horizontal plane, while examples of "3D" loudspeaker setups include 7.1 with two height channels, 9.1 and 22.2 [29]. However, especially interesting is ambisonics, a production format for 3D audio that is based on the representation of the sound field excitation as decomposition into orthonormal basis functions, the so-called spherical harmonics. It allows for a production process that is independent of the target playback system [29].

The Ambisonic technology enables one to musically compose and shape the contents of the presented sound field as a "Virtual Auditory Environment". To fulfill this goal, a concert venue employing the Ambisonics format, probably the first on that scale, consisted of a 25-channel loudspeaker system, an outside broadcasting van with a 23-channel loudspeaker system, as well as a 5.1 mixdown and a binaural headphone mixdown for national terrestrial and satellite radio broadcasting. It was organized and transmitted during the International Conference on Spatial Audio 2015 in September 2015 [30]. This was a highlight of this conference.

Moreover, ambisonics coupled with a personalized HRTF (*Head Relate Transfer Function*) enables fully perceived externalized sound while using headphones. This is a subject of intensive studies [31-33].

Finally, this is not a place to discuss basics and background of audio compression, but it should be reminded that only lossless music retains all the audio quality of the original source.

V. Conclusions

In conclusion it is noteworthy to refer to one of the Internet leitmotifs that says that "Future of the Music: Innovation Will Not Come from Within the Industry Itself". Five major trends that may change our way of listening to music and even how we perceive it are listed. While the first one seems already available, namely: personalized music driven machine learning application to select the most suitable songs for each listener, then the others are either new concepts or new applications of these concepts, i.e.: music used as emotional change vehicle understood as the change from spectate to participate (e.g. performances of Jurgis Didžiulis (aka Jurgis Did)); music written by robots employing artificial intelligence (AI) used nowadays by gaming industry, ad agencies and to a lesser extent by film makers; music as medicine matching one's natural day-night rhythms; binaural 3D sound, an emerging technology that aims at not only reproducing sound in a natural way as we perceive it, but also recording it the way we hear it. Also, a new relationship between music and the user and its environment driven by AI, IoT (*Internet of Things*), biometric data, multi-sensory experience, social media and other tools [34][36-37] may be envisioned. Talking of which, one should not forget about the 3D hologram world tour of a singer or a fictional singer being a pop star in Japan.

Of importance is also personal communication with music services, thus at least some of the further (or rather present) efforts is to be aimed at consolidating already existing ways of communication via avatars such as e.g. Alexa or Siri, alienating listeners from reality even more than today. This concerns especially the millennial generation which is the first that has been able to experience music almost entirely out of chronology [35].

Finally, contrasting this somber view of treating music services and users as a part of the IoT with very interesting alternatives that technology brought to music domain, may be desirable. As an example of such technology virtual instruments [36], analyzing and creating new music [37] or virtual air instruments may serve (see Fig. 2) [38]. Among these instruments one may find virtual drum, air guitar, xylophone, and membrane. The experience of playing on such instruments is described as inspiring, awe-inspiring, just stunning or "nothing short of phenomenal", especially as a live music event performance.

Lastly, *musification* and music morphing notions may be mentioned here [39], the first one understood as a musical representation of data or musical interpretation of a course of music events, the latter one as the gradual transition between musical gestalts or as the musical interpretation of processes [39]. They are based on mathematical analysis and music data interpretation. Engaging computers in music storytelling will certainly appear as one of the future directions of music creation.

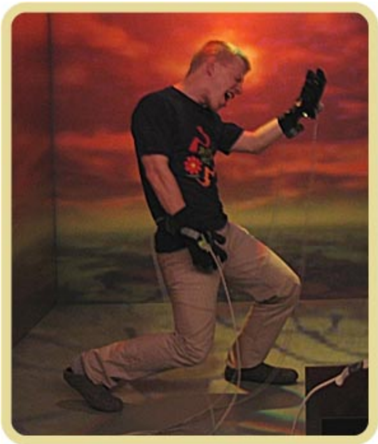

Fig. 2. Virtual air guitar created by Mäki-Patola *et al*. http://airguitar.tml.hut.fi/whatis.html [38][40].

ACKNOWLEDGMENT

This work has been supported by Statutory Funds of ETI Faculty, GUT.

REFERENCES

[1] 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

[2] http://www.last.fm/

[3] Spotify.com

[4] play.google.com/music

[5] https://www.apple.com/

[6] https://www.apple.com/itunes/music/

[7] https://www.deezer.com/

[8] ISMIR, The International Society for Music Information Retrieval website, http://www.ismir.net/

[9] A. Lindsay and J. Herre, "MPEG-7 and MPEG-7 Audio – An Overview," J. Audio Eng. Soc., vol. 49, no. 7/8, pp. 589-594, 2001.

[10] S. Quackenbush and A. Linsday, "Overview of MPEG-7 Audio," IEEE Trans. Circuits Syst. Video Technol., vol. 11, no. 6, pp. 725–729, 2001.

[11] B. Kostek, Perception-Based Data Processing in Acoustics. Applications to Music Information Retrieval and Psychophysiology of Hearing. Berlin, Heidelberg, New York: Springer Verlag, 2005.

[12] A. Rosner and B. Kostek, "Automatic music genre classification based on musical instrument track separation," J. Intell. Inf. Syst., vol. 5, 2017, DOI: 10.1007/s10844-017-0464-5.

[13] G. Tzanetakis, G. Essl, and P. Cook, "Automatic musical genre classification of audio signals," in Proc. Int.Symp. Music Information Retrieval (ISMIR), 2001.

[14] L. Vrysis, N. Tsipas, C. Dimoulas, and G. Papanikolaou, "Crowdsourcing Audio Semantics by Means of Hybrid Bimodal Segmentation with Hierarchical Classification," J. Audio Eng. Soc., vol. 64, no. 12 pp. 1042–1054, 2016, DOI: https://doi.org/10.17743/jaes.2016.0051.

[15] O. Celma, P. Herrera, and X. Serra, "Bridging the Music Semantic Gap, Workshop on Mastering the Gap: From Information Extraction to Semantic Representation," Budva, Montenegro, 2006.

[16] J. A. Burgoyne, I. Fujinaga, and J. S. Downie, "Music Information Retrieval," Chapter 15 in A New Companion to Digital Humanities, S. Schreibman, R. Siemens, J. Unsworth (Eds.), John Wiley & Sons, 2016, DOI: https://doi.org/10.1002/9781118680605.ch15.

[17] http://www.pandora.com

[18] A. Ghias, J. Logan, D. Chamberlin, and B.C. Smith, "Query by humming - musical information retrieval in an audio database," in Proceedings Multimedia'95, San Francisco, pp. 231–236, 1995.

[19] T. C. Nagavi1 and N. U. Bhajantri, "An Extensive Analysis of Query by Singing/Humming System through Query Proportion," Int. J. Multimed. Appl., vol. 4, pp. 73–86, 2012, DOI: 10.5121/ijma.2012.4606.

[20] B. Kostek, "Music Information Retrieval in Music Repositories," Chapter 17 in Rough Sets and Intelligent Systems Reference Library, in Suraj Z., Skowron A. (Eds.), Berlin-Heidelberg: Springer Verlag, pp. 463–489, 2013.

[21] B. Kostek, "Music Information Retrieval – Soft Computing versus Statistics," in Proc. CISIM 2015 Computer Information Systems and Industrial Management, pp. 36–47, 2015, DOI: https://doi.org/10.1007/978-3-319-24369-6_3.

[22] P. Grover, "Various Implementations of Collaborative Filtering. Comparison of different methods to build recommendation system using collaborative filtering," [Online] https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0.

[23] M. Slaney and M. Casey, "Locality-Sensitive Hashing for Finding Nearest Neighbors," IEEE Signal Process. Mag., 2008, DOI: 10.1109/MSP.2007.914237

[24] I. Guy, N. Zwerdling, D. Carmel, I. Ronen, E. Uziel, S. Yogev, andS. Ofek-Koifman, "Personalized recommendation of social software items based on social relations," in Proc. 3rd ACM Conference on Recommender systems, New York, NY, USA, pp. 53–60, 2009, DOI: http://dx.doi.org/10.1145/1639714.1639725

[25] J. L. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating Collaborative Filtering Recommender Systems Herlocker," ACM T. Inform. Syst., vol. 22, no. 1, 2004.

[26] X. Mu, Y. Chen, and T. Li, "User-Based Collaborative Filtering Based on Improved Similarity Algorithm," in Proc. 3rd IEEE International Conference on Computer Science and Information Technology, Chengdu, pp. 76–80, 2010.

[27] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms," in Proc. 10th International Conference on World Wide Web, Hong Kong, pp. 285–295, 2001, DOI: ACM 1-58113-348-0/01/0005.

[28] J. Wang, A. P. de Vries, and M. J. T. Reinders, "Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion," in Proc. 29th Annual Intern. ACM SIGIR Conference on Research and Development in Information Retrieval, 2006.

[29] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H Audio – The New Standard for Universal Spatial/3D Audio Coding," J. Audio Eng. Soc., vol. 62, no. 12, pp. 821–830, 2014.

[30] M. Frank and A. Sontacchi, "Case Study on Ambisonics for Multi-Venue and Multi-Target Concerts and Broadcasts," J. Audio Eng. Soc., vol. 65, no. 9, pp. 749–756, 2017, DOI: https://doi.org/10.17743/jaes.2017.0026.

[31] F. Brinkmann, A. Lindau, S. Weinzierl, S. van de Par, M. Müller-Trapet, R. Opdam, and M. Vorländer, "A High Resolution and Full-Spherical Head-Related Transfer Function Database for Different Head-Above-Torso Orientations," J. Audio Eng. Soc., vol. 65, no. 10, pp. 841–848, 2017, DOI: https://doi.org/10.17743/jaes.2017.0033.

[32] E. Hendrickx, P. Stitt, J. C. Messonnier, J. M. Lyzwa, B. F.G. Katz, and C. de Boishéraud, "Improvement of Externalization by Listener and Source Movement Using a 'Binauralized' Microphone Array," J. Audio Eng. Soc., vol., 65, no. 7/8, pp. 589–599, 2017, DOI: https://doi.org/10.17743/jaes.2017.0018.

[33] F. Klein, S. Werner, and T. Mayenfels, "Influences of Training on Externalization of Binaural Synthesis in Situations of Room Divergence," J. Audio Eng. Soc., vol. 65, no. 3, 2017, DOI: https://doi.org/10.17743/jaes.2016.0072

[34] M. Lech and B. Kostek, "Evaluation of the influence of ergonomics and multimodal perception on sound mixing while employing a novel gesture-based mixing interface," J. Audio Eng. Soc., vol. 61, no. 5, pp. 301–313, 2013.

[35] Is technology changing our relationship with music? [Online] https://www.futurity.org/music-festivals-technology-1322452-2/

[36] L. Turchet, A. McPherson, and M. Barthet, “Co-design of a Smart Cajón,” J. Audio Eng. Soc., vol. 66, no. 4, pp. 220–230 2018, DOI: https://doi.org/10.17743/jaes.2018.0007.

[37] G. W. Don, K. Muir, G. B. Volk, and J. S. Walker, “Music: Broken Symmetry, Geometry, and Complexity”, Notices Am. Math. Soc., vol. 57, no. 1, 2010.

[38] T. Mäki-Patola, J. Laitinen, A. Kanerva, and T. Takala, “Experiments with virtual reality instruments,” in Proc. Conference on New Interfaces for Musical Expression, Vancouver, BC, CA, pp. 11–16, 2005.

[39] J. Edlund, “A Musification and View. The Virtues of the Musifier: A Matter of View,” [Online] http://www.musifier.com/index.php/tech/musification-and-view.

[40] Airguitar [Online] http://airguitar.tml.hut.fi/whatis.html