Working Group Meeting – Thursday November 24th 2016 10-11am

In digital music, playlists are becoming the main way of consuming music (Song, Dixon, & Pearce, 2012) either on radios like BBC, on web streaming platforms such as the Obama's playlist on Spotify[1], on personal devices such as laptops, smartphones (Thalmann, Carillo, Fazekas, Wiggins, & Sandler, 2016) and MP3 Player (Nettamo, Nirhamo, & Häkkilä, 2006) or on connected speakers like Sonos[2]. Users can manually create their playlists but a growing number of them listen to automatically generated ones (Uitdenbogerd & Schyndel, 2002) created by music recommender systems (Yoshii, Goto, Komatani, Ogata, & Okuno, 2007) which suggests songs to fill each listeners' taste.

Such playlists creation requires that audio files are annotated with meaningful information called tag (Jaschke, Marinho, Hotho, & Schmidt-Thieme, 2007). A musical piece can gather one or multiple tags either understandable by common human listeners such as "happy" or not like "dynamic complexity" (Laurier & Herrera, 2007; Streich, 2006). Five methods are available to collect music tags (Turnbull, Barrington, & Lanckriet, 2008). Three of them implies humans, e.g. social tagging websites (Breese, Heckerman, & Kadie, 1998; Cohen & Fan, 2000; Levy & Sandler, 2007; Shardanand & Maes, 1995; Shepitsen, Gemmell, Mobasher, & Burke, 2008) also known as collaborative filtering and used by Last.fm[3] and Amazon[4] (Linden, Smith, & York, 2003), music annotation games (Law, Ahn, Dannenberg, & Crawford, 2007; M. I. Mandel & Ellis, 2008; Turnbull, Liu, Barrington, & Lanckriet, 2007) or surveys (Clifford, 2007; Turnbull & Barrington, 2008). The last two tagging methods are computer-based and include text mining web-documents (Knees, Pohle, Schedl, & Widmer, 2007; Whitman & Ellis, 2004) and auto-tagging audio content (Bertin-Mahieux, Eck, & Mandel, 2010; Prockup et al., 2015; Tzanetakis & Cook, 2002). Human labelling is time-consuming, prone to errors and abuse (Song et al., 2012) and like text mining web-documents it can only process a small subset of all available musical pieces. The remaining pieces in the long-tail (Hitt & Anderson, 2007) are thus affected by the cold start issue or data sparsity problem (Song et al., 2012). This issue is defined as the impossibility to recommend a song because of his novelty and so its lack of rating or tag (Eck, Lamere, Green, & Bertin-Mahieux, 2007; Li, Myaeng, & Kim, 2007; Schafer, Frankowski, Herlocker, & Sen, 2007; Schlüter & Grill, 2015). Besides, the cold start problem is often linked to popularity bias. Indeed, a known musical piece is more likely to be annotated on multiple criteria (Eck et al., 2007) whereas an unpopular one will remain poorly labelled. Obviously, the number of tags is highly correlated with the presence of a musical piece in multiple playlists[5] and emphasises the hubness effect (Flexer, 2015).

Auto-tagging or audio content-based tagging is relatively fast and solves the major problems of human-gathered tags. Auto-tagging analyses an audio signal and automatically generate tags from it (Celma, Ramírez, & Herrera, 2005; Eck et al., 2007; Hoashi, Matsumoto, & Inoue, 2003; Logan, 2002, 2004; M. Mandel & Ellis, 2008; Sordo, Laurier, & Celma, 2007; Tingle, Kim, & Turnbull, 2010; Turnbull, Barrington, Torres, & Lanckriet, 2007). A makeshift solution consists in combining the multiple tags generating methods (Bu, Tan, Chen, Wang, & Wu, 2010; Yoshii et al., 2007) to produce robust tags. However, automatic tags generation still contains precision issue and remains improvable.

Music is "Vocal or instrumental sounds (or both) combined in such a way as to produce beauty of form, harmony, and expression of emotion" as defined by the Concise Oxford Dictionary[6]. But until today it is still impossible to generate a faultless playlist with the tag "Instrumental" (Bayle, Hanna, & Robine, 2016; Ghosal, Chakraborty, Dhara, & Saha, 2013). Such instrumental music is not only needed for leisure but it is also used in therapy (Rosenblatt, 2015) and learning enhancement methods (Suárez, Elangovan, & Au, 2016; Zhao & Kuhl, 2016). To generate a perfect playlist is even harder when considering more complex tags such as genre (Hsu, Lin, & Chi, 2016; Jeong & Lee,

---

[1] https://www.spotify.com
[2] https://www.sonos.com
[3] http://www.last.fm/
[4] https://www.amazon.fr/Applications-Amazon-Music/b?ie=UTF8&node=1949609031
[5] http://www.billboard.com/biz/articles/news/digital-and-mobile/5944950/the-echo-nest-cto-brian-whitman-on-spotify-deal-man-vs
[6] https://en.oxforddictionaries.com/definition/music

2016; Lu, Wu, Lu, & Lerch, 2016; Oramas, Espinosa-Anke, & Lawlor, 2016). Even state-of-the-art algorithms failed to propose good tag precision and accuracy when faced to big musical databases (Bayle et al., 2016; Bogdanov, Porter, Herrera, & Serra, 2016).

In the light of all these issues, a new paradigm is needed which rethinks the classification problem and focus the real question[7] which has to be solved (Sturm, 2016; Wiggins, 2009). If we consider a listener who would like to listen to an instrumental playlist with the most elaborated auto-tagging algorithms, he would encounter one song with singing-voice every five songs (Bayle et al., 2016). Indeed, those algorithms are made to improve precision, accuracy and recall as a whole, which is not suited for a perfect playlist generation. To do so it is preferable to tag songs only if the algorithm can ensure that the tag is 100% correct. So the new aim is to set 100% of precision while maximising the recall even if only a small number of correctly classified musical pieces are retrieved. This reduction cannot be a downside when considering the large amount of musical pieces available on popular audio streaming applications such like Spotify or Deezer[8]. For example, Deezer makes available more than 40 million musical pieces[9]. Moreover, the maximum playlist size authorised on streaming platform varies from 1000[10] for Deezer to 10000[11] for Spotify through 5000 for YouTube[12] and Google Play Music. The mean user's playlist size is certainly far below. Thus, recall maximisation defines the metric to be used for large-scale audio database tagging, playlist generation and song recommendation.

In this Working Group, I will tell you about my new paradigm and general progress of my thesis.

Bibliography

Bayle, Y., Hanna, P., & Robine, M. (2016). Classification à grande échelle de morceaux de musique en fonction de la présence de chant. In *Journées d'Informatiques Musicales* (pp. 144–152).

Bertin-Mahieux, T., Eck, D., & Mandel, M. (2010). Automatic Tagging of Audio: The State-of-the-Art. *Machine Audition: Principles, Algorithms and Systems*, 334–352. http://doi.org/10.4018/978-1-61520-919-4.ch014

Bogdanov, D., Porter, A., Herrera, P., & Serra, X. (2016). Cross-collection evaluation for music classification tasks. In *International Society for Music Information Retrieval* (pp. 379–385). New York. Retrieved from http://mtg.upf.edu/system/files/publications/bogdanov2016crosscollection.pdf

Breese, J., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. *Proceedings of the Fourteenth*, 43–52. Retrieved from http://dl.acm.org/citation.cfm?id=2074100

Bu, J., Tan, S., Chen, C., Wang, C., & Wu, H. (2010). Music recommendation by unified hypergraph: combining social media information and music content. In *Proceedings of the …* (pp. 391–400). New York, New York, USA: ACM Press. http://doi.org/10.1145/1873951.1874005

Celma, Ò., Ramírez, M., & Herrera, P. (2005). Foafing the music: A music recommendation system based on RSS feeds and user preferences. In *International Society for Music Information Retrieval* (pp. 457–464). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.74.1744

---

[7] http://ejhumphrey.com/?p=302

[8] http://www.deezer.com

[9] https://www.deezer.com/features

[10] http://support.deezer.com/hc/en-gb/articles/201193652-Is-there-a-limit-to-the-amount-of-tracks-in-a-playlist-

[11] https://community.spotify.com/t5/Desktop-Linux-Windows-Web-Player/Maximum-songs-on-playlists/td-p/108021

[12] https://developers.google.com/youtube/2.0/developers_guide_protocol_playlists?csw=1

Clifford, S. (2007). Pandora's Long Strange Trip. *Inc.* Retrieved from http://www.inc.com/magazine/20071001/pandoras-long-strange-trip_pagen_2.html

Cohen, W., & Fan, W. (2000). Web-collaborative filtering: Recommending music by crawling the web. *Computer Networks*, *33*(1), 685–698. Retrieved from http://www.sciencedirect.com/science/article/pii/S1389128600000578

Darwin, C. (1871). *The descent of man, and selection in relation to sex* (1st ed.). London: John Murray. Retrieved from http://darwin-online.org.uk/converted/pdf/1871_Descent_F937.1.pdf

Eck, D., Lamere, P., Green, S., & Bertin-Mahieux, T. (2007). Automatic generation of social tags for music recommendation. *Advances in Neural Information Processing Systems*, 1–8. Retrieved from http://machinelearning.wustl.edu/mlpapers/paper_files/NIPS2007_892.pdf

Flexer, A. (2015). Improving visualization of high-dimensional music similarity spaces. In *International Society for Music Information Retrieval* (pp. 547–553). Malaga, Spain. Retrieved from https://www.researchgate.net/publication/294855594_IMPROVING_VISUALIZATION_OF_HIGH-DIMENSIONAL_MUSIC_SIMILARITY_SPACES

Ghosal, A., Chakraborty, R., Dhara, B., & Saha, S. (2013). A hierarchical approach for speech-instrumental-song classification. *SpringerPlus*. Retrieved from http://link.springer.com/article/10.1186/2193-1801-2-526

Hegel, G. W. F. (1998). *Hegel's Aesthetics: Lectures on Fine Art, Vol. I.* (O. U. Press, Ed.). Clarendon. Retrieved from http://www.amazon.com/Hegels-Aesthetics-Lectures-Fine-Vol/dp/0198238169

Higham, T., Basell, L., Jacobi, R., Wood, R., Ramsey, C. B., & Conard, N. J. (2012). Testing models for the beginnings of the Aurignacian and the advent of figurative art and music: The radiocarbon chronology of Geißenklösterle. *Journal of Human*, *62*(6), 664–676. Retrieved from http://www.sciencedirect.com/science/article/pii/S0047248412000425

Hitt, M., & Anderson, C. (2007). The Long Tail: Why the Future of Business Is Selling Less of More. *Academy of Management Perspectives*, *21*(2), 83–85. Retrieved from http://www.jstor.org/stable/27747380

Hoashi, K., Matsumoto, K., & Inoue, N. (2003). Personalization of user profiles for content-based music retrieval based on relevance feedback. In *Proceedings of the eleventh ACM international conference on Multimedia* (pp. 110–119). http://doi.org/10.1145/957013.957040

Hsu, K., Lin, C., & Chi, T. (2016). Sparse coding based music genre classification using spectro-temporal modulations. In *International Society for Music Information Retrieval* (pp. 744–750). New York. Retrieved from https://pdfs.semanticscholar.org/0129/b46c14d3dc926fa80f26f2ee66ebee8d81b8.pdf

Jaschke, R., Marinho, L., Hotho, A., & Schmidt-Thieme, L. (2007). *Knowledge Discovery in Databases: PKDD 2007*. (J. N. Kok, J. Koronacki, R. Lopez de Mantaras, S. Matwin, D. Mladenič, & A. Skowron, Eds.)*Lecture Notes In Computer Science* (Vol. 4702). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-74976-9

Jeong, I., & Lee, K. (2016). Learning temporal features using a deep neural network and its application to music genre classification. In *International Society for Music Information Retrieval* (pp. 434–440). New York. Retrieved from https://www.researchgate.net/profile/Il_Young_Jeong/publication/305683876_Learning_temporal_features_using_a_deep_neural_network_and_its_application_to_music_genre_classification/links/5799a27c08aec89db7bb9f92.pdf

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *The Behavioral and Brain Sciences*, *31*(5), 559–621. http://doi.org/10.1017/S0140525X08005293

Knees, P., Pohle, T., Schedl, M., & Widmer, G. (2007). A music search engine built upon audio-based and web-based similarity measures. *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR 07*, 447. http://doi.org/10.1145/1277741.1277818

Laurier, C., & Herrera, P. (2007). Audio music mood classification using support vector machine. *MIREX Task on Audio Mood Classification*, 2–4. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:AUDIO+MUSIC+MOOD+CLASSIFICATION+USING+SUPPORT+VECTOR+MACHINE#0

Law, E. L. M., Ahn, L. Von, Dannenberg, R. B., & Crawford, M. (2007). Tagatune : a Game for Music and Sound Annotation. *In International Conference on Music Information Retrieval (ISMIR'07)*, 361–364. Retrieved from http://www.cs.cmu.edu/~elaw/papers/ISMIR2007.pdf

Lehner, B., Widmer, G., & Sonnleitner, R. (2014). On the reduction of false positives in singing voice detection. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 7480–7484. http://doi.org/10.1109/ICASSP.2014.6855054

Levy, M., & Sandler, M. B. (2007). A semantic space for music derived from social tags. *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, *1*, 411–416. Retrieved from http://ismir2007.ismir.net/proceedings/ISMIR2007_p411_levy.pdf

Li, Q., Myaeng, S. H., & Kim, B. M. (2007). A probabilistic music recommender considering user opinions and audio features. *Information Processing and Management*, *43*(2), 473–487. http://doi.org/10.1016/j.ipm.2006.07.005

Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, *7*(1), 76–80. http://doi.org/10.1109/MIC.2003.1167344

Logan, B. (2002). Content-Based Playlist Generation: Exploratory Experiments. *Ismir*, 6–7.

Logan, B. (2004). Music Recommendation from Song Sets. In *International Society for Music Information Retrieval* (pp. 10–14). Retrieved from http://shiftleft.com/mirrors/www.hpl.hp.com/techreports/2004/HPL-2004-148.pdf

Lu, Y., Wu, C., Lu, C., & Lerch, A. (2016). Automatic outlier detection in music genre datasets. In *International Society for Music Information Retrieval* (pp. 101–107). New York. Retrieved from https://wp.nyu.edu/ismir2016/wp-content/uploads/sites/2294/2016/07/215_Paper.pdf

Mandel, M., & Ellis, D. (2008). Multiple-Instance Learning for Music Information Retrieval. In *International Society for Music Information Retrieval* (pp. 577–582). Retrieved from http://www.ee.columbia.edu/~dpwe/pubs/MandelE08-MImusic.pdf

Mandel, M. I., & Ellis, D. P. W. (2008). A Web-Based Game for Collecting Music Metadata. *Journal of New Music Research*, *37*(2), 151–165. http://doi.org/10.1080/09298210802479300

Nettamo, E., Nirhamo, M., & Häkkilä, J. (2006). A cross-cultural study of mobile music. *Proceedings of the 20th Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of Australia on Computer-Human Interaction: Design: Activities, Artefacts and Environments - OZCHI '06*, 87. http://doi.org/10.1145/1228175.1228193

Oramas, S., Espinosa-Anke, L., & Lawlor, A. (2016). Exploring Customer Reviews for Music Genre Classification and Evolutionary Studies. In *International Society for Music Information Retrieval* (pp. 150–156). New York. Retrieved from http://irserver.ucd.ie/handle/10197/7975

Prockup, M., Ehmann, A. F., Gouyon, F., Schmidt, E. M., Celma, O., & Kim, Y. E. (2015). Modeling Genre with the Music Genome Project: Comparing Human-Labeled Attributes and Audio Features. *Proceedings of the 16th International Society for Music Information Retrieval*, 31–37. Retrieved from http://music.ece.drexel.edu/files/Prockup2015b.pdf

Rosenblatt, D. (2015). *Music Listening as Therapy*. Loma Linda University. Retrieved from http://scholarsrepository.llu.edu/cgi/viewcontent.cgi?article=1301&context=etd

Schafer, J., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. *The Adaptive Web*, 291–234. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-72079-9_9

Schlüter, J., & Grill, T. (2015). Exploring Data Augmentation for Improved Singing Voice Detection With Neural Networks. In *Proceedings of the 16th International Society for Music Information Retrieval*. Retrieved from http://www.ofai.at/~jan.schlueter/pubs/2015_ismir.pdf

Shardanand, U., & Maes, P. (1995). Social information filtering: algorithms for automating "word of mouth." *Proceedings of the SIGCHI Conference on*, 210–217. Retrieved from http://dl.acm.org/citation.cfm?id=223931

Shepitsen, A., Gemmell, J., Mobasher, B., & Burke, R. (2008). Personalized recommendation in social tagging systems using hierarchical clustering. *Proceedings of the 2008 ACM Conference on Recommender Systems*, 259–266. http://doi.org/10.1145/1454008.1454048

Song, Y., Dixon, S., & Pearce, M. (2012). A survey of music recommendation systems and future perspectives. *9th International Symposium on Computer Music Modeling and Retrieval*, (June), 19–22. Retrieved from http://cmmr2012.eecs.qmul.ac.uk/sites/cmmr2012.eecs.qmul.ac.uk/files/pdf/papers/cmmr2012_submission_42.pdf

Sordo, M., Laurier, C., & Celma, O. (2007). Annotating Music Collections: How Content-Based Similarity Helps to Propagate Labels. In *International Society for Music Information Retrieval* (pp. 531–534). Retrieved from https://www.researchgate.net/profile/Oscar_Celma/publication/220723033_Annotating_Music_Collections_How_Content-Based_Similarity_Helps_to_Propagate_Labels/links/00b7d5191638a68591000000.pdf

Streich, S. (2006). Music complexity: a multi-faceted description of audio content. *Department of Technology*, *49*, 124. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.76.3588&rep=rep1&type=pdf

Sturm, B. (2016). Revisiting Priorities: Improving MIR Evaluation Practices. In *International Society for Music Information Retrieval* (pp. 488–494). New York. Retrieved from https://qmro.qmul.ac.uk/xmlui/handle/123456789/13353

Suárez, L., Elangovan, S., & Au, A. (2016). Cross-sectional study on the relationship between music training and working memory in adults. *Australian Journal of*, *68*(1), 38–46. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/ajpy.12087/full

Thalmann, F., Carillo, A. P., Fazekas, G., Wiggins, G. A., & Sandler, M. B. (2016). The Semantic Music Player: A Smart Mobile Player Based on Ontological Structures and Analytical Feature Metadata. *Proceedings of the 2nd Web Audio Conference (WAC-2016)*. Retrieved from https://smartech.gatech.edu/handle/1853/54596

Tingle, D., Kim, Y., & Turnbull, D. (2010). Exploring automatic music annotation with acoustically-objective tags. In *Proceedings of the international* (pp. 55–62). Retrieved from

http://dl.acm.org/citation.cfm?id=1743400

Turnbull, D., & Barrington, L. (2008). Semantic annotation and retrieval of music and sound effects. *Audio, Speech, and ...*, *16*(2), 467–476. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4432652

Turnbull, D., Barrington, L., & Lanckriet, G. (2008). Five approaches to collecting tags for music. *Ismir 2008*, 225–230. Retrieved from http://books.google.com/books?hl=en&lr=&id=OHp3sRnZD-oC&oi=fnd&pg=PA225&dq=Five+approaches+to+collecting+tags+for+music&ots=oCRLwHhy80&sig=ylzZX9MhF5YFAtPbLx_sEZNAdf0

Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G. (2007). Towards musical query-by-semantic-description using the cal500 data set. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 439–446). http://doi.org/10.1145/1277741.1277817

Turnbull, D., Liu, R., Barrington, L., & Lanckriet, G. (2007). Using Games To Collect Semantic Annotations Of Music. *Ismir 2007*, 1–3. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.5649&rep=rep1&type=pdf

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, *10*(5), 293–302. http://doi.org/10.1109/TSA.2002.800560

Uitdenbogerd, A., & Schyndel, R. (2002). A review of factors affecting music recommender success. *ISMIR 2002, 3rd*, 204–208. Retrieved from http://researchbank.rmit.edu.au/view/rmit:30951

Whitman, B., & Ellis, D. (2004). Automatic record reviews. *Proceedings of the 5th International Society for Mu- Sic Information Retrieval Conference*, 470–477. Retrieved from http://academiccommons.columbia.edu/catalog/ac:148699

Wiggins, G. A. (2009). Semantic Gap?? Schemantic Schmap!! Methodological Considerations in the Scientific Study of Music. In *2009 11th IEEE International Symposium on Multimedia* (pp. 477–482). IEEE. http://doi.org/10.1109/ISM.2009.36

Yoshii, K., Goto, M., Komatani, K., Ogata, T., & Okuno, H. (2007). Improving Efficiency and Scalability of Model-Based Music Recommender System Based on Incremental Training. In *International Society for Music Information Retrieval* (pp. 88–94). Retrieved from https://www.researchgate.net/profile/Hiroshi_Okuno/publication/220723011_Improving_Efficiency_and_Scalability_of_Model-Based_Music_Recommender_System_Based_on_Incremental_Training/links/0fcfd50f0a409405e5000000.pdf

Young, T. (1845). *A Course of Lectures on Natural Philosophy and the Mechanical Arts: pt. I. Mechanics. pt. II. Hydrodynamics. pt. III. Physics*. Taylor and Walton. Retrieved from https://books.google.fr/books?hl=fr&lr=&id=mQhDAAAAIAAJ&oi=fnd&pg=PA1&dq=.+Young,+A+Course+Of+Lectures+On+Natural+Philosophy+And+The+Mechanical+Arts+,+Vol.+1,+Taylor+and+Walton,+2+edition,+1845&ots=-GyLZWa5gQ&sig=AKXmlhm652G26jXR1bQXIvic-4w

Zhao, T., & Kuhl, P. (2016). Musical intervention enhances infants' neural processing of temporal structure in music and speech. *Of the National Academy of Sciences*, *113*(19), 5212–5217. Retrieved from http://www.pnas.org/content/early/2016/04/20/1603984113.short