Siamese Neural Networks for Content-based Cold-Start Music Recommendation.

Michael Pulis
michael.pulis.18@um.edu.mt
University of Malta
Msida, Malta

Josef Bajada
josef.bajada@um.edu.mt
University of Malta
Msida, Malta

ABSTRACT

Music recommendation systems typically use collaborative filtering to determine which songs to recommend to their users. This mechanism matches a user with listeners that have similar tastes, and uses their listening history to find songs that the user will probably like. The fundamental issue with this approach is that artists already need to have a significant user following to get a fair chance of being recommended. This is known as the music cold-start problem. In this work, we investigate the possibility of making music recommendations based on audio content so that new artists still get a good chance of being recommended, even if they do not have a sufficient number of listeners yet.

We propose the use of Siamese Neural Networks (SNNs) to determine the similarity between two audio clips. Each clip is first pre-processed into a Mel-Spectrogram, which is then used as input to an SNN consisting of two identical Convolutional Neural Networks (CNNs). The output of each CNN is then compared together to determine whether two songs are similar or not. These were trained using audio from the Free Music Archive, with the genre used as a heuristic to determine the similarity between song pairs.

A query-by-multiple-example (QBME) music recommendation system was developed that makes use of the proposed content-based similarity metric to find songs that match the user's tastes. This was packaged inside an online blind-test survey, which first prompts participants to select a set of preferred songs, and then recommends a number of songs which the subject is expected to listen to and rate on a Likert scale. The recommendations from the proposed algorithm were stochastically interleaved with songs selected randomly from the preferred genres of the user, as a baseline for comparison. The participants were not aware that the recommendations came from two different algorithms.

Our findings show that 60.7% of the 150 participants gave higher ratings to the recommendations made by the proposed SNN-based algorithm. Findings also show that 55% of the recommended songs had less than 1,500 listens, demonstrating that the proposed content-based approach can provide a fairer exposure to all artists based on their music, independent of their fame and popularity.

CCS CONCEPTS

• **Computing methodologies → Supervised learning by classification**.

KEYWORDS

Music Recommendation, Deep Metric Learning, Siamese Networks

ACM Reference Format:
Michael Pulis and Josef Bajada. 2021. Siamese Neural Networks for Content-based Cold-Start Music Recommendation.. In *Fifteenth ACM Conference on Recommender Systems (RecSys '21), September 27-October 1, 2021, Amsterdam, Netherlands.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3460231.3478847

1 INTRODUCTION

Nowadays 89% of music consumption takes place through on-demand streaming services [IFPI 2019]. However, even though these services make use of sophisticated recommendation systems to suggest content to their listeners, there is a significant imbalance between the content that gets listened to and the full range of music available on the system. In the case of Spotify, roughly 43,000 artists account for 90% of streams [Spotify 2020], despite the platform offering over 3 million artists. One of the problems contributing to this phenomenon is the fact that recommendation techniques often rely on prior listening histories of other users to create links between related songs and artists. This technique is known as *collaborative filtering*, and it is the most commonly used technique in music streaming services [Feng et al. 2020; van den Oord et al. 2013; Zhang et al. 2018]. The main flaw of this technique is its inability to recommend new songs or artists, since listening history for this music does not exist yet. This issue is commonly referred to as the item cold-start problem [Cao et al. 2020; Magron and Févotte 2021; Oramas et al. 2017; Saveski and Mantrach 2014].

In this work we investigate mechanisms to analyse the actual music content and determine how similar a song is to the music that the user has already listened to. We propose a music similarity metric based on Siamese Neural Networks (SNNs) [Bromley et al. 1993]. Each song is converted to a Mel-Spectrogram, a bitmap representation of the audio signal converted to a *mel scale* [Stevens and Volkmann 1940; Stevens et al. 1937], which is then fed into a Convolutional Neural Network (CNN) [LeCun et al. 1995] to obtain a vector embedding. When the similarity between two songs needs to be calculated they are both processed by the same CNN, and the Euclidean distance between both vectors is computed. This is then fed into a fully connected layer of size 1 using a softmax activation function. This similarity metric is then used in a query-by-multiple-example (QBME) recommender system to find songs that are similar to the ones the user likes. This approach was evaluated through an

RecSys '21, September 27-October 1, 2021, Amsterdam, Netherlands

ACM ISBN 978-1-4503-8458-2/21/09.
https://doi.org/10.1145/3460231.3478847

online survey, where users were first asked to select the songs they like, and then rate the recommendations provided by the system.

2 RELATED WORK

In this section we will first analyse algorithms for computing similarity between songs, after which we will review the main frameworks used to make recommendations given a database of songs and a set of preferred ones.

2.1 Content-based Similarity Metrics

The music cold-start problem has been tackled in different ways in previous research. Oramas et al. [2017] use a combination of metadata, such as the artist biography, together with audio features to extract embeddings. This approach assumes that the publicly known information about the artist can help to provide a more accurate match to the user's tastes. Another approach is to focus solely on the audio content. Deep Metric Learning (DML) has been used in order to determine the similarity between two audio clips. One of the most common DML model is the SNN [Kaya and Bilge 2019]. This has been used successfully to determine similarity between audio clips [Manocha et al. 2017], where a dataset of 5,000 audio clips were converted to bitmap representations of log-scaled spectrograms. Pairs of the extracted images were then created, assigning a label of 1 to clips that were considered similar, and 0 to clips which were dissimilar. To decide which audio clips were similar or not, they were split into classes such as *'car passing by residential area'* or *'wind blowing'*. The SNN was then trained on this dataset to learn to encode each song into a vector of fixed length. The similarity between two songs was computed by calculating the Euclidean distance between the feature vector of any two songs. The system was evaluated by obtaining the top 25 most similar clips to each query clip. It was shown that for certain classes the system managed to obtain a mean precision of 0.78. Further analysis of the results also confirmed that the system was able to retrieve audio clips that were acoustically similar.

SNNs were also used to detect song covers [Stamenovic 2020], that is when the same song is played by a different artist or band. Since the SNN requires data to be fed in pairs, the Second-Hand-Songs dataset was used, which is a subset of the Million Song Dataset that only contains cover songs. Around 50K song pairs were made, split evenly between cover-song pairs, and non-cover-song pairs. For each song, the Constant Q Transform (CQT) was used to obtain a bitmap representation its audio content. The SNN model was then trained on the bitmap pairs to learn a similarity score, and to ultimately classify if a particular song pair includes a cover-song or not. The precision values showed that the system achieved comparable results to other methods while not reaching the state-of-the-art when attempting to detect cover song pairs. This highlighted the potential of SNNs to learn a similarity metric in the context of music, based on bitmap representations of song pairs. The effect of different bitmap representations used to represent audio content on CNN classification was investigated by Huzaifah [2017]. Several methods were outlined and compared such as Short-Time Fourier transforms (STFT) using both linear and Mel-spectrograms, as well as CQT. The findings showed that using a Mel-Spectrogram consistently gave the best median accuracy score between all transform operations for classification tasks, whilst also outperforming the baseline MFCC features. This shows that Mel-Spectrograms offer a viable representation of audio content for use within CNNs.

2.2 Content-based Music Information Retrieval

Content-Based Music Information Retrieval (CB-MIR) makes use of audio content to determine which songs to recommend. Query-based approaches were identified to be ideal for CB-MIR tasks based on similarity metrics [Murthy and Koolagudi 2018]. One such query-based approach was implemented to simulate a content based *query-by-example* (QBE) system which retrieves songs from a fixed database [McFee et al. 2012], with other works making similar use of the QBE paradigm for music recommendation [Harb and Chen 2003; McFee et al. 2012; Tsai et al. 2005].

Tavares and Manzolli [2014] implemented and compared several recommendation methodologies which made use of metrics defined on musical raw audio content, and QBE was compared to *query-by-multiple-example* (QBME). QBME is applicable where for all users, an entire *preference set* is known, which contains a number of songs which the user is known to enjoy, while in QBE, a single preference example is used to make the recommendation. The authors conclude that for systems that make use of content based similarity metrics and have access to a preference set for each user, QBME systems are able to provide better results than performing QBE on each song within the preference set individually and amalgamating the results. These results were also corroborated by Soleymani et al. [2015] where users were asked to create a preference set prior to making content based recommendations.

Bogdanov et al. [2010] compared the different possible implementations of QBME. They implemented three similarity metric based approaches for producing a set of recommendations given a preference set, and a candidate set of songs from which recommendations are made. In the first approach, *SEM-MEAN*, a mean representation of all the songs within it was created. In the second approach, *SEM-ALL*, the full preference set is retained, and the similarity metric is computed between all preference and candidate song pairs. In the third approach, *SEM-GMM*, a clustering based Gaussian Mixture Model was used. The most successful approach, in terms of lowest *fail* rate and the highest *hit* rate amongst the respondents, was found to be the *SEM-ALL*.

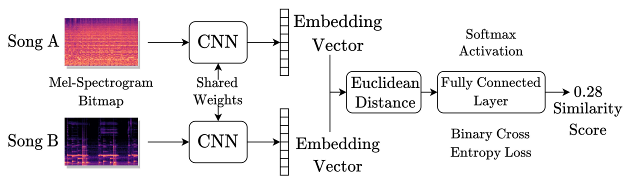
3 A CONTENT-BASED RECOMMENDATION SYSTEM

We propose a QBME song recommendation system that uses an SNN content-based similarity metric to find songs that are similar to a user's preference set, and clusters songs based on this similarity for efficient retrieval.

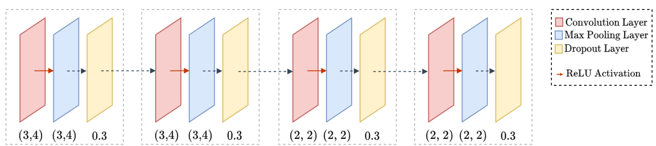
3.1 SNN Song Similarity Metric

A song similarity metric based on an SNN model, which uses a Mel-Spectrogram representation as its input was used as the content based similarity metric. A Mel-Spectrogram is a bitmap representation of an audio signal where the time-frequency relationship is retained, but the frequencies are translated to representations that reflect how humans perceive musical notes [Romani et al. 1982].

Libraries to perform this conversion are freely available, such as Librosa [1], which was used for this work. In order to perform this operation, two parameters are needed, the highest frequency (in Hz) to be represented, f_{max}, and the number of frequency bands to group the frequencies, n_{mels}. The commonly used values for these two parameters are $f_{max} = 8000$ and $n_{mels} = 128$ [Adebiyi 2020; Nodera et al. 2019].



(a) Siamese Neural Network Model



(b) CNN Structure

Figure 1: Structure of the SNN and CNN

The SNN then makes use of two identical CNNs which have shared weights, as shown in Figure 1. Each CNN is made up of 4 groups of layers, with each group of layers containing a Convolutional Layer and a Max-Pooling Layer [Stamenovic 2020], as shown in Figure 1. Since SNNs are known to be prone to overfitting [Manocha et al. 2017], a Dropout Layer was included within each batch. The convolutional layers within each group make use of a ReLU activation function, while the remaining layers make use of a linear activation function [Stamenovic 2020].

The output of each CNN is an embedding vector of length 48. The Euclidean distance is calculated between the two vectors, which is then fed into a fully connected layer of size 1 using a softmax activation function. The output of the similarity metric is thus a continuous value between 0 and 1 [Stamenovic 2020]. This final output layer makes use of a Binary Cross-Entropy loss function [Stamenovic 2020] to determine whether two songs are similar, 1, or not, 0.

In order to train the model, the similarity between two songs needs to be established. Throughout this work we chose to utilise the Free Music Archive (FMA) [2], which is a database of royalty-free music. As a similarity heuristic, the genre labels within the FMA dataset were used, where each song had one top-level genre label (for example Rock, or Pop). Positive (similar) song-pairs had the same top-level genres, while negative (dissimilar) song pairs had different top-level genres. Similar and dissimilar song pairs were given a label of 0 and 1 respectively. A total of 32,000 song pairs were used to train the model, with 80% used for training and 20% used for validation. Another separate set equivalent in size to the validation set was then used as a test set to measure the accuracy. The model was trained for 133 epochs with a batch size of 32, a dropout rate of 0.3 and optimised using an Adam Optimizer with a learning rate of 0.0008. This model was implemented using Keras and was trained on Google Colab. The source code is available online [3].

3.2 QBME Recommendation System

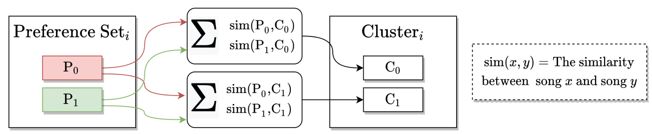


Figure 2: Score Calculation

To make recommendations, a QBME approach was used [Bogdanov et al. 2010; Tavares and Manzolli 2014] which utilises the aforementioned SNN model as its similarity metric. To return a list of recommendations from a dataset of 25,000 songs, given a preference set, P, a score is assigned to each song, with the top scoring songs being returned as the recommendation list. The score of a potential recommendation, c, is determined by summing the values of the similarity score between each song $p \in P$ and c, $sim(p, c)$. Similarly to the approach used in Bogdanov et al. [2010], to prevent song scores from being inflated due to the summing of hundreds of thousands of small values, the similarity score is only considered if it is greater than 0.5, as shown in Equation 1.

$$\text{score}(c) = \sum_{\forall p \in P} \begin{cases} 0 & \text{sim}(c, p) < 0.5 \\ \text{sim}(c, p) & \text{sim}(c, p) \geq 0.5 \end{cases} \tag{1}$$

To avoid a pairwise comparison between all entries, S, within the QBME system, the songs are clustered by the embedding output of the SNN beforehand using k-Means clustering [Burns and van Zyl 2020], with $k = 10$. The set of candidate recommendations, C, for a preference set, P, is then found by matching the cluster, $cluster(p)$, of each song in the preference set, $p \in P$, with that of the candidate songs, as follows:

$$C = \{c \mid c \in S \wedge \exists p \in P : cluster(c) = cluster(p)\} \tag{2}$$

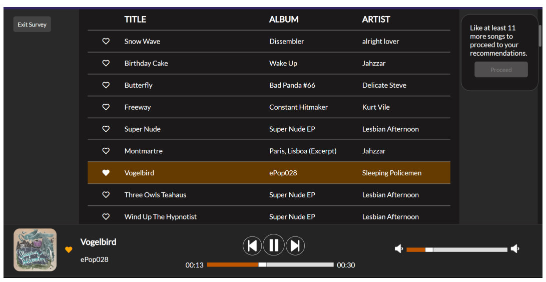
4 SURVEY METHODOLOGY

An online survey was implemented in the form of a responsive website to allow for easy distribution. The system, shown in Figure 3, presents up to 80 songs for each of the 16 genres present in the FMA dataset. These songs were selected according to the most listened-to songs within the dataset for that particular genre.

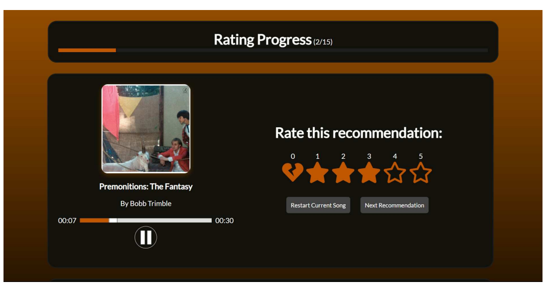
The participant can listen to songs from across the different genres, and press 'like' on the songs they enjoy, which adds the song to their preference set. After at least 12 songs are added, the participant is allowed to proceed to the second part of the survey, where recommendations are made based on the preference set. In

[1] https://librosa.org
[2] https://freemusicarchive.org/
[3] https://github.com/michaelpulis/SnnForCbColdStartMusicRecommendation

the second part of the survey, the participant rates recommendations made by the system, using a 0-5 star scale. Before proceeding to the next recommendation, the participant must listen to the song for at least 5 seconds. The survey makes use of two recommendation systems, where one is based on the SNN-based QBME recommendation system, while the other is a naive genre-based system, used as a baseline, which selects a song randomly from the same genres of the preference set. To select the genre to use for the next recommendation, the baseline system uses a probabilistic roulette wheel selection method, where the probability of each slice is linked with the amount of songs from that particular genre in the preference set. Songs from both recommendation systems were interleaved stochastically to eliminate bias [Dunsch et al. 2018; Morii et al. 2017]. A total of 150 participants were recruited for the survey, mostly through social media and email invitations, spanning different demographics. The survey was completely anonymous, and no details about the participant were collected.



(a) Creating Preference Set



(b) Rating Recommendations

Figure 3: Screenshots of the two parts of the survey

5 EVALUATION

After the SNN model was trained using the genre-based similarity heuristic, it was first evaluated for its accuracy. The test set consisted of 6,400 songs and achieved an accuracy of 81.64%. A further analysis on the performance of each model per genre was performed, where 1,000 song pairs were tested. Figure 4 shows the accuracy and loss of the learnt model for the 16 genres, together with the number of songs from each genre found in the dataset.

To evaluate the SNN-based recommendation system with real users, the online survey described in Section 4 was conducted, recruiting a total of 150 participants over 26 days. The online survey was accessible using either a desktop computer or a mobile phone.

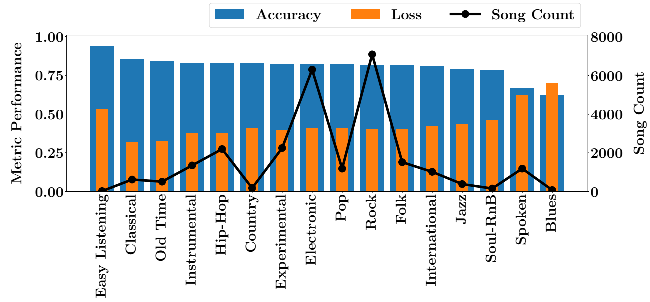
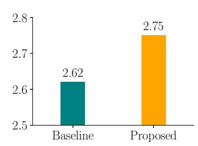
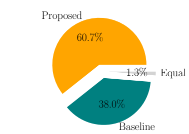


Figure 4: Accuracy and Loss of the SNN similarity model, together with the number of songs per genre.

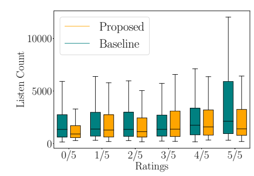
Figure 5 shows that the recommendations made by the SNN-based recommendation system achieved a higher rating than those made by the baseline naive genre-based system. The findings also show that 60.7% of the participants preferred the SNN-based recommendation system overall. These results were also analysed statistically through a Mann Whitney U-test, with the results being significant at the ($p < 0.05$) level.



(a) Average Rating Per System



(b) Preferred System



(c) Listen count distribution per rating category

Figure 5: Preference Statistics

The distribution of the listen count from the FMA dataset of each song when grouped by the ratings given by the participants can be seen in Figure 5. These show that the recommendations given by the proposed system, which were rated 5/5 by participants had a lower median listen count, and a smaller variance. This indicates that the proposed system achieved high ratings by making highly personalised recommendations, irrespective of the popularity of the songs and artists, based solely on the audio content of the songs within the participants' preference sets. In comparison, the baseline system achieved 5/5 ratings by recommending songs that coincidentally had a higher listen count and a noteworthy increase in variance, indicating that the recommendations were more generic and less personalised than those of our proposed system. The same effect can be observed in almost all of the other rating categories, confirming the fact that while the proposed system recommended less popular songs and artists, it still achieved significantly higher ratings by participants. The difference in the distribution of the listen counts between the baseline and proposed systems, illustrated in Figure 5, was found to be statistically significant using a Mann Whitney U-test at the ($p < 0.05$) level, where the proposed system had a significantly lower median listen count, thus giving an opportunity for lesser known artists to be discovered. 55% of the recommended songs had less than 1,500 listens, demonstrating the proposed system's ability to recommend novel songs and artists effectively. This shows that for the participants, the proposed system offered a solution to the issue of novel song and artist recommendation.

6 CONCLUSION AND FUTURE WORK

An SNN-based song similarity model was proposed to learn song similarity based on audio content. The results show that the developed model was able to compute similarity for different genres of music with a consistent level of accuracy and loss across most of the genres. An accuracy of 81.64% was achieved on the test set of 6,400 song pairs. The model was then embedded within a QBME music recommendation system that considers all of the songs within the listener's preference set to provide a list of recommendations that sound similar.

A blind-use quantitative survey was developed and distributed to 150 participants. The recommendations made by the proposed system were randomly interleaved with recommendations made by a naive genre-based baseline system. The proposed SNN-based recommendation system was able to achieve a significantly higher average rating than the baseline system, whilst also recommending songs that are significantly less popular than those recommended by the baseline system. The research carried out in this study also opens up various opportunities for future research, such as using more popular and modern songs, and using a better song similarity heuristic based on human assessment.

REFERENCES

Sodiq Adebiyi. 2020. *An Emotion Based Music Recommender System Using Deep Learning.* Ph.D. Dissertation. Dublin, National College of Ireland.

Dmitry Bogdanov, Martín Haro Berois, Ferdinand Fuhrmann, Emilia Gómez Gutiérrez, Herrera Boyer, et al. 2010. Content-based music recommendation based on user preference examples. In *Proceedings of the Workshop on Music Recommendation and Discovery 2010 (WOMRAD 2010).* CEUR Workshop Proceedings, Barcelona, Spain, 33–8.

Jane Bromley, James W Bentz, Léon Bottou, Isabelle Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, and Roopak Shah. 1993. Signature verification using a "siamese" time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence* 7, 04 (1993), 669–688.

Jamie Burns and Terence L van Zyl. 2020. Automated Music Recommendations Using Similarity Learning. In *First Southern African Conference for Artificial Intelligence Research.* SOCAIR 2020, South Africa, 288–303.

Ke Yin Cao, Yu Liu, and Hua Xin Zhang. 2020. Improving the Cold Start Problem in Music Recommender Systems. *Journal of Physics: Conference Series* 1651 (nov 2020), 012067. https://doi.org/10.1088/1742-6596/1651/1/012067

Felipe Dunsch, David K Evans, Mario Macis, and Qiao Wang. 2018. Bias in patient satisfaction surveys: a threat to measuring healthcare quality. *BMJ global health* 3, 2 (2018), e000694.

Chenjiao Feng, Jiye Liang, Peng Song, and Zhiqiang Wang. 2020. A fusion collaborative filtering method for sparse data in recommender systems. *Information Sciences* 521 (2020), 365–379. https://doi.org/10.1016/j.ins.2020.02.052

Hadi Harb and Liming Chen. 2003. A query by example music retrieval algorithm. In *Digital Media Processing For Multimedia Interactive Services: Proceedings of the 4th European Workshop on Image Analysis for Multimedia Interactive Services.* World Scientific Publishing, United Kingdom, 122–128.

Muhammad Huzaifah. 2017. Comparison of Time-Frequency Representations for Environmental Sound Classification using Convolutional Neural Networks. arXiv:1706.07156 [cs.CV]

IFPI. 2019. *IFPI releases Music Listening 2019 - IFPI.* Technical Report. IFPI. https://www.ifpi.org/ifpi-releases-music-listening-2019/

Mahmut Kaya and Hasan Şakir Bilge. 2019. Deep metric learning: A survey. , 1066 pages. https://doi.org/10.3390/sym11091066

Yann LeCun, Yoshua Bengio, et al. 1995. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* 3361, 10 (1995), 1995.

Paul Magron and Cédric Févotte. 2021. Leveraging the Structure of Musical Preference in Content-Aware Music Recommendation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021).* IEEE, Toronto, Canada, 581–585. https://doi.org/10.1109/ICASSP39728.2021.9414194

Pranay Manocha, Rohan Badlani, Anurag Kumar, Ankit Shah, Benjamin Elizalde, and Bhiksha Raj. 2017. Content-based Representations of audio using Siamese neural networks. *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.* 2018-April (oct 2017), 3136–3140. arXiv:1710.10974

Brian McFee, Luke Barrington, and Gert Lanckriet. 2012. Learning content similarity for music recommendation. *IEEE Trans. Audio, Speech Lang. Process.* 20, 8 (2012), 2207–2218. https://doi.org/10.1109/TASL.2012.2199109 arXiv:1105.2344

Masahiro Morii, Takayuki Sakagami, Shigetaka Shinya Masuda, Okubo, and Yuki Tamari. 2017. How does response bias emerge in lengthy sequential preference judgments? *Behaviormetrika* 44 (2017), 575–591. https://doi.org/10.1007/s41237-017-0036-6

Y. V. Srinivasa Murthy and Shashidhar G. Koolagudi. 2018. Content-Based Music Information Retrieval (CB-MIR) and Its Applications toward the Music Industry: A Review. *ACM Comput. Surv.* 51, 3, Article 45 (June 2018), 46 pages. https://doi.org/10.1145/3177849

Hiroyuki Nodera, Yusuke Osaki, Hiroki Yamazaki, Atsuko Mori, Yuishin Izumi, and Ryuji Kaji. 2019. Deep learning for waveform identification of resting needle electromyography signals. *Clin. Neurophysiol.* 130, 5 (may 2019), 617–623. https://doi.org/10.1016/j.clinph.2019.01.024

Sergio Oramas, Oriol Nieto, Mohamed Sordo, and Xavier Serra. 2017. A Deep Multimodal Approach for Cold-Start Music Recommendation. In *Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems* (Como, Italy) *(DLRS 2017).* Association for Computing Machinery, New York, NY, USA, 32–37. https://doi.org/10.1145/3125486.3125492

Gian Luca Romani, Samuel J. Williamson, and Lloyd Kaufman. 1982. Tonotopic organization of the human auditory cortex. *Science (80-.).* 216, 4552 (jun 1982), 1339–1340. https://doi.org/10.1126/science.7079770

Martin Saveski and Amin Mantrach. 2014. Item Cold-Start Recommendations: Learning Local Collective Embeddings. In *Proceedings of the 8th ACM Conference on Recommender Systems* (Foster City, Silicon Valley, California, USA) *(RecSys '14).* Association for Computing Machinery, New York, NY, USA, 89–96. https://doi.org/10.1145/2645710.2645751

Mohammad Soleymani, Anna Aljanaki, Frans Wiering, and Remco C. Veltkamp. 2015. Content-based music recommendation using underlying music preference structure. In *2015 IEEE International Conference on Multimedia and Expo (ICME).* IEEE, Turin, Italy, 1–6. https://doi.org/10.1109/ICME.2015.7177504

Spotify. 2020. Spotify 2020 Shareholder Report. https://s22.q4cdn.com/540910603/files/doc{_}financials/2020/q2/Shareholder-Letter-Q2-2020{_}FINAL.pdf

Marko Stamenovic. 2020. Towards Cover Song Detection with Siamese Convolutional Neural Networks. arXiv:2005.10294 [eess.AS]

Stanley S Stevens and John Volkmann. 1940. The relation of pitch to frequency: A revised scale. *The American Journal of Psychology* 53, 3 (1940), 329–353.

Stanley Smith Stevens, John Volkmann, and Edwin Broomell Newman. 1937. A scale for the measurement of the psychological magnitude pitch. *The journal of the acoustical society of america* 8, 3 (1937), 185–190.

Tiago Fernandes Tavares and Jônatas Manzolli. 2014. Query-by-Multiple-Examples: Content-Based Search in Computer-Assisted Sound-Based Musical Composition. In *Music Technology meets Philosophy - From Digital Echos to Virtual Ethos: Joint Proceedings of the 40th International Computer Music Conference, ICMC 2014, and the 11th Sound and Music Computing Conference, SMC 2014.* Michigan Publishing, Athens, Greece, 397–401.

Wei-Ho Tsai, Hung-Ming Yu, and Hsin-Min Wang. 2005. Query-By-Example Technique for Retrieving Cover Versions of Popular Songs with Similar Melodies.. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, Vol. 5. ISMIR 2005, London, United Kingdom, 183–190.

Aaron van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. In *Advances in Neural Information Processing Systems (NIPS 2013)*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.), Vol. 26. Curran Associates, Inc., Lake Tahoe, USA.

Chunxia Zhang, Ming Yang, Jing Lv, and Wanqi Yang. 2018. An improved hybrid collaborative filtering algorithm based on tags and time factor. *Big Data Mining and Analytics* 1, 2 (2018), 128–136.