

Publication Date: 10 August 2024

Archs Sci. (2024) Volume 74, Issue S1 Pages 20-25, Paper ID 2024s14.
<https://doi.org/10.62227/as/74s14>

Analysis of the Performance Forms and Techniques of Dynamic Performing Arts and Musical Performances in Musicals Based on Improved Neural Collaborative Filtering

Linlong Jiang^{1,*}

¹School of music, University of Sanya, Sanya 572000, Hainan, China.

Corresponding authors: (e-mail: 18907762163@163.com).

Abstract The performance of a musical is an artistic expression that takes into account both singing and drama, and it is a musical journey for the characters to achieve their goals under the prescribed situation given by the text of the musical. In-depth analysis of the situational performance of musical theatre singing, and discussing how musical theatre actors support the dramatic performance of musical theatre singing through situation establishment, character relationship processing and drama reinforcement are hot topics in the field of music research. Based on the improved neural collaborative filtering technology, this paper attempts to extract song content features from the macro attributes of music (genre, artist, etc.), and uses the Word2vec method to train from the original high-dimensional sparse feature vectors to obtain dense low-dimensional hidden feature vectors. The collaborative filtering, matrix decomposition and content recommendation methods based on the proposed method are integrated into the framework of deep learning technology, and an improved neural collaborative filtering music recommendation model is proposed to provide personalized song recommendations for each user. In terms of model structure, on the one hand, all features are grouped, and comprehensive group features are obtained after processing to enhance the expressive ability of the model; on the other hand, interactive features are added to the model through the outer product of vectors to learn more nonlinear information. Experiments are carried out on two music datasets. The experimental results show that the F1 and mAP of the model in this paper are much higher than other models, so the recommendation quality is significantly better than other models. In addition, due to the introduction of user information and song attributes, the cold start problem is alleviated, and the use of Word2vec method also effectively alleviates the problem of data sparsity, and the use of deep learning methods greatly improves the generalization ability and robustness of the model.

Index Terms neural collaborative filtering, optimization, dynamic performance, musical theatre performance

I. Introduction

Musical singing is an important form of expression for the characters in the play to express their emotions, pursue change and promote development. The singing under the prescribed situation is dramatic, and the actors strengthen the construction of the prescribed situation and the shaping of the characters in the singing. Singing and dramatic situations complement each other, which is one of the artistic characteristics of musicals [1], [2]. Musical theatre singing includes two aspects: vocal technique and situational performance. In my study and work, I found that many students or actors study music directly from musical scores when they study musical works, which often weakens the rules in scripts and scores. Situation to analyze the roles and events, find suitable situations to rely on their singing and performances, so as to complete the functions of shaping the role, explaining the relationship between the characters, and promoting the

development of the plot [3], [4]. Therefore, pay attention to situational performances-undertake important tasks of dramatic expression, performance continuation and outreach, make singing credible and watchable, so as to complete the integration of music and drama, and create artistic resonance with the audience. This thesis is born out of this thinking and artistic practice [5].

Constructivism is a theory based on postmodern cognitive paradigms. The constructivist curriculum view is formed on the basis of constructivist theory. Constructivism, as a kind of philosophical epistemology, has two cores: the constructive nature of knowledge and the constructive nature of cognition [6]. It condenses all the essence of postmodern cognitive paradigm and education, and forms a series of reasonable knowledge and cognition in postmodern education [7]. It profoundly and comprehensively affects the current education reform in the world. Constructivism pays great attention to

the learning subject experience and subject activities, and has an excellent intersection with the music curriculum reform, especially the affirmation of the foundation of experience and the subjectivity of students, which is in line with the general direction of the music curriculum reform. It provides new enlightenment for the reform of music curriculum [8]. The author believes that the most prominent expression of musicals in the process of singing and dissemination is its popular performance form, which extends from the initial discussion of social collective memory to the thinking of aesthetic modernity. In other words, this kind of close-to-nature, individualized performance and its dissemination and listening have created the creative form, narrative aesthetic characteristics, commercial consumerism, and cultural awareness of musicals [9]. This research involves many theoretical categories such as the creation, singing, aesthetic form, communication media and social culture of Chinese musicals, and combines interdisciplinary research methods and multi-dimensional research perspectives. Do research. Through the collection and arrangement of a large number of documents, the author found that the academic achievements of pop music research are seriously lacking compared with the results of other types of music research. There are some misunderstandings about the value judgment of musical theatre singing, and some scholars even believe that it belongs to the fringe of the discipline and has not been formed. scientific research system [10]. However, as an integral part of Chinese music culture, musical theatre singing has an important influence on its social function, production method, and communication media in the field of consumption aesthetics, showing its unique situation and modernity characteristics. From the perspective of popularization, this paper makes an in-depth exploration of musicals from the aspects of performance, creation, aesthetics, and social thought, which will have a great academic research space [11], [12].

The feature information of the music source signal can be automatically extracted by using the neural network, and then the extracted features of the sound source can be used for the subsequent separation step. Traditional music source separation algorithms often require researchers to select and design appropriate signal models to model the music sources to be separated, such as modeling the vocals and accompaniment in music separately, and then use signal processing method to achieve the separation of the music source signal. A common risk of these model-based approaches is that the core assumptions relied upon in modeling the source signal under study may be overturned. In addition, the modeling for a certain sound source is also susceptible to the interference of other sound source signals contained in the music.

II. Improved Neural Collaborative Filtering

If move the speaker diaphragm according to the recorded waveform, the sound is reproduced. A multi-channel signal simply consists of several waveforms captured by multiple microphones. Typically, a music signal is stereo, a combination of multiple channel signals. In general, an audio signal can be

represented by time as an independent variable. In the field of signal processing, this one-dimensional audio signal is usually analyzed by Fourier transform [13]. With the help of *FT* analysis, the distribution of different frequency components contained in the audio signal and their signal amplitudes. For stationary signals, there are two variables to focus on. They are time and frequency. With the help of *FT*, we can analyze the time domain characteristics of the signal in the time domain, and we can also convert it to the frequency domain to analyze the frequency domain characteristics. In particular, by performing *FT* analysis on the music signal, we can not only obtain the representation of the music signal in the frequency domain, but also obtain the energy distribution of the music signal. However, the Fourier transform analyzes a whole music signal, which has limitations. When we want to analyze the relationship between the frequency domain characteristic information of the signal and the time variable, we can no longer use *FT* to analyze it, and we can only analyze it separately in the time domain. and research in the frequency domain. cannot analyze the time-dependent relationship of frequency information in the signal, but it can analyze the relationship between the frequency components and phases contained in the entire signal and can analyze the amplitude distribution of different frequency components. In order to analyze the relationship between the frequency characteristics of the signal and the variable time, the time-frequency analysis method can be used. Its calculation method is shown in Eq1:

$$STFT(f, t) = \sum_{-\infty}^{+\infty} [x(t) m(t - \tau)] e^{-j2\pi f t} dt. \quad (1)$$

Selecting a suitable window function when using analysis will have a greater impact on the analysis results. If the width of the window function is narrow, the frequency resolution of the signal will be low. Conversely, if the width of the window function is wide, the time resolution of the signal will be low. Figure 1 is an example of a human voice amplitude spectrogram of a music signal.

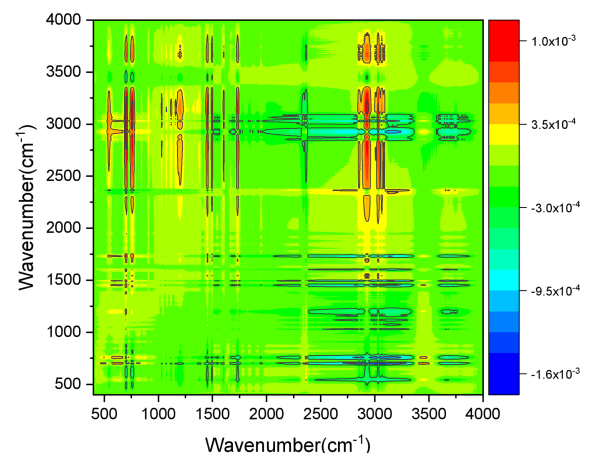


Figure 1: Example of vocal amplitude spectrogram of music signal

A. Deep Neural Network Collaboration and Convolution Operation

The fully connected neural network needs to calculate the weight information of the neurons in the whole network, which leads to the large amount of parameters of this kind of network, which also reduces the training speed of the entire model and brings about the problem of overfitting of the model. And it can effectively solve the problems encountered when the fully connected neural network processes image feature information. The process of convolution is actually the process of weighted summation. The number of convolution kernels used in this convolution operation corresponds to the number of output channels of the convolution operation. The schematic Figure 2 of its convolution operation is as follows:

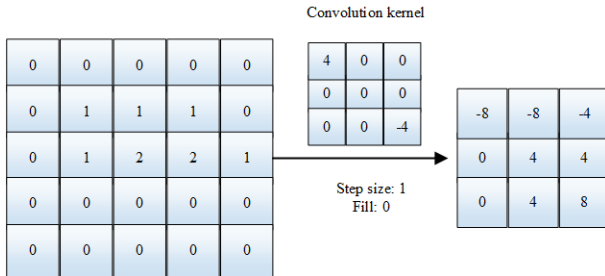


Figure 2: Schematic diagram of neural network convolution operation

Then move the window according to the preset fixed step size until all the input data are taken, and then the new eigenvalues will be spliced in order. The pooling layer directly reduces the resolution of the input features, reduces the calculation amount of the entire model, and also allows the network to obtain spatial invariance, which is also the extraction of features again. Figure 3 is a schematic diagram of the pooling operation.

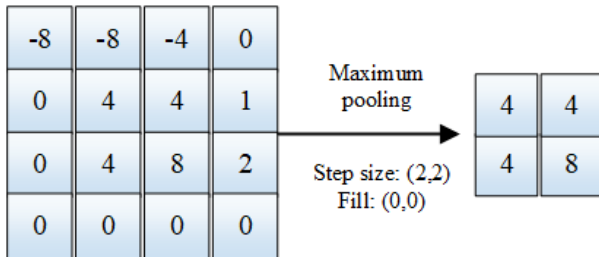


Figure 3: Schematic diagram of pooling operation

B. Normalization

Each time the data passes through a network layer and then is output, the parameters of the entire network layer will be updated. As the data is transmitted in each layer of the network, the distribution characteristics of the data must also be different from when it was first input to the network. If it is not corrected, The network needs to adapt to the input data of different distributions, and the learning cost of the

entire network will be greatly increased, resulting in a slow gradient descent of the network and a long training time [14]. This is the "Internal Covariate Shift" problem that has plagued researchers for a long time. In order to solve the problem, researchers have proposed many effective normalization processing algorithms to adjust the data distribution of each layer of the network. Therefore, the most important thing for user-based collaborative filtering recommendation is to find the K users that are most "similar" to the target user. The similarity calculation formula is as follows:

$$\begin{aligned} \text{sim}(x, y) &= \cos(x, y) = \frac{x \cdot y}{||x|| \cdot ||y||} \\ &= \frac{\sum_{k \in I_{xy}} r_{xk} r_{yk}}{\sqrt{\sum_{k \in I_x} r_{xk}^2} \sqrt{\sum_{k \in I_y} r_{yk}^2}}. \end{aligned} \quad (2)$$

Taking into account the difference in evaluation scales among different users, a modified cosine similarity is proposed, and its formula is as follows:

$$\text{sim}(x, y) = \frac{\sum_{k \in I_{xy}} (r_{xy} - \bar{r}_x) (r_{xy} - \bar{r}_y)}{\sqrt{\sum_{k \in I_x} (r_{xy} - \bar{r}_x)^2} \sqrt{\sum_{k \in I_y} (r_{xy} - \bar{r}_y)^2}}. \quad (3)$$

Therefore, the current recommendation algorithms are all hybrid recommendation models, but it is still necessary to know the applicable scenarios, advantages and disadvantages of each recommendation model, so as to distinguish the priority and improve the model in the future recommendation tasks. Table 1 is a summary of common problems of several commonly used recommendation algorithms.

The Skip-Gram model predicts the probability of occurrence of surrounding context words through several central words in a word sequence, and then smoothly predicts the occurrence probability of all words in the word sequence, as shown in Figure 4:

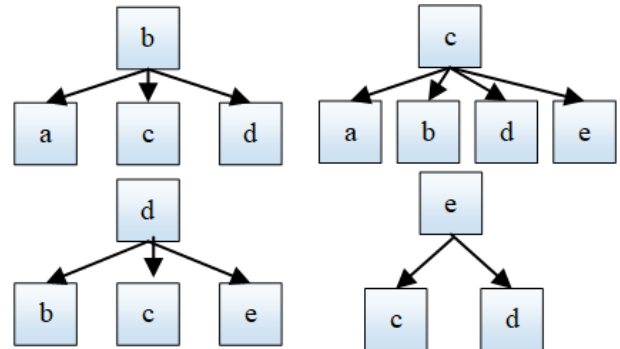


Figure 4: Schematic diagram of skip gram model Methods

In musicals, if the music signal is to be transmitted without distortion over long distances, it is necessary to encode the signal to be transmitted, convert the original signal into another encoded signal that can be easily transmitted for transmission, and then re-encode the signal at the receiving end of the

Recommended algorithm	Whether to alleviate data sparsity	Whether to alleviate cold start	Generate personalized recommendation
User-based CF	NO	NO	NO
Item-based CF	NO	NO	YES
LFM	YES	NO	YES
Content -based	YES	YES	YES
Paper model	YES	YES	YES

Table 1: Summary of common problems of several commonly used recommendation algorithms

signal. The signal is decoded to obtain the original signal. The encoder-decoder network structure is a network structure that applies encoding and decoding technology to the field of deep learning [15]. The encoder in this network structure corresponds to the encoder, which can effectively convert an input data into Other types of feature data are output, but the main feature information in the original data is not lost, so as to facilitate the operations we need on the feature data. The decoder part, as the decoder, is responsible for restoring the data features extracted by the former encoding to the scale and dimension of the original input data. In deep learning, you can use all convolutional layers to build an encoder-decoder network, you can also use RNN to build the encoder and decoder in the network, or you can mix and match, use CNN to build the encoder of the network, and use RNN or LSTM to build The decoder of the network to meet the processing needs of the entire network for feature informationbeg. The schematic diagram of its convolution matrix is shown in Figure 5:

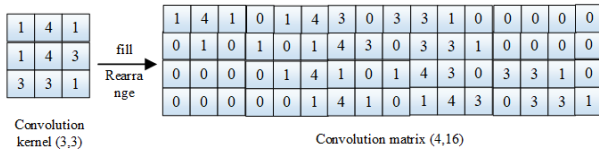


Figure 5: Experimental simulation diagram of convolution matrix

For audio signals, each harmonic component contained in the reconstructed signal is inseparable from the amplitude spectrum and phase spectrum corresponding to each harmonic component, and once the phase spectrum is disturbed, the reconstructed time domain signal will have a very large impact, so the phase spectrum is not manipulated in this chapter. The frame diagram of its music source feature extraction and separation algorithm is shown in Figure 6.

The structure of the deep convolutional encoder-decoder network model based on SA attention mechanism proposed in this paper is shown in Figure 7.

Among them, each up sampling block consists of 5 network layers, which are bilinear interpolation layer (BI), transposed convolution layer of size (3, 3), BN layer, dropout layer and Rely activation layer. The use of transposed convolution to construct up sampling blocks is abandoned, and bilinear interpolation is used to achieve up sampling, which reduces the amount of parameters and achieves the purpose of up sampling the feature map. The experimental environment is shown in Table 2.

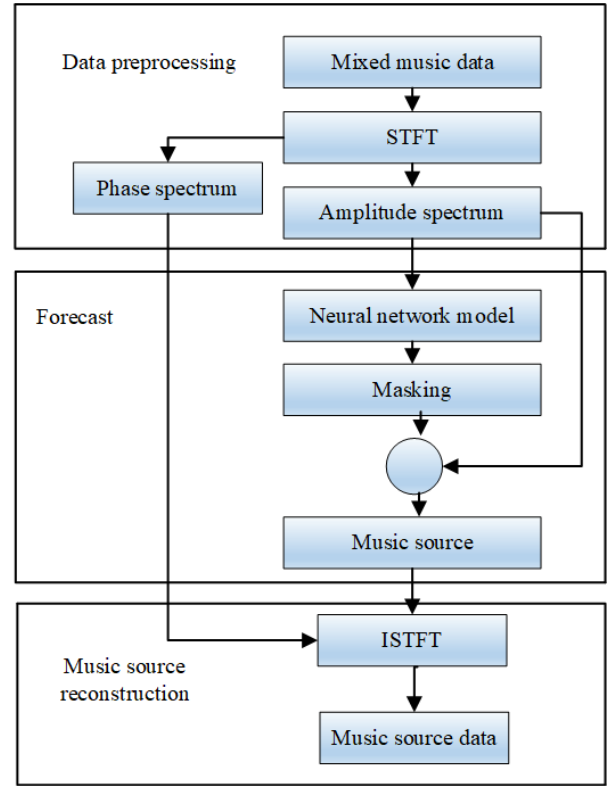


Figure 6: Framework diagram of music source feature extraction and separation algorithm

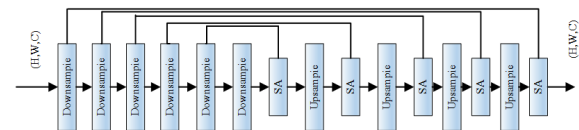


Figure 7: SA-based Encoder-decoder network

III. Case Study

A. Improve the construction accuracy of the musical sub-module

For musicals, the operation of aggregating the feature maps of the same level in the FEM structure is obtained by direct addition. The feature map of the input FEM is restored to the dimension before the input domain FEM module after the last layer of deconvolution upsampling module, and then the feature information after FEM feature extraction is obtained after passing through SAM. FEM does not change the size and number of channels of the input feature map. It uses a more efficient structure to extract the feature information of

Environment name	Specific configuration
Operating system	Ubuntu 18. 04
Development language	Python3
Deep learning framework	Pytorch1.8
Integrated development environment	Vscode
CPU	15-10400f
CPU	GTX1080
Memory	32G
Hard disk	500G

Table 2: Experimental environment and configuration table

the input data. At the same time, FEM can be easily embedded into the convolutional neural network, which will increase a certain amount of parameters. , but can effectively increase the feature extraction capability of the entire network. The schematic diagram of its accuracy is shown in Figure 8.

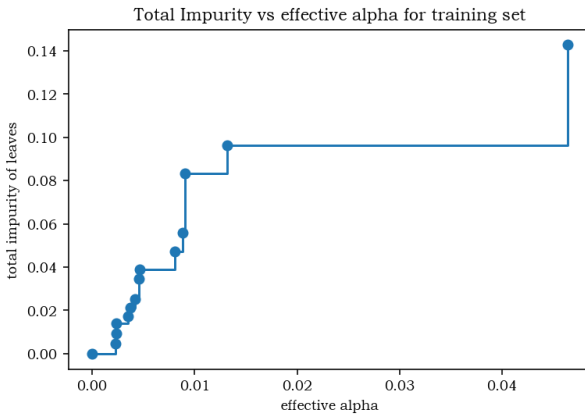


Figure 8: Schematic diagram of sub-module accuracy

Use two DNNs to extract features from the amplitude spectrum and use DNN to extract features from the phase spectrum of the mixed music signal. The amplitude spectrum feature is added to the network, and finally the two networks are used to predict the amplitude spectrum and phase spectrum of the music source signal, and then the separated music source signal can be reconstructed in the time domain with the help of ISTFT.

B. Improve the effect of music interpretation

After the partial derivative of the original phase spectrum is calculated in time and frequency, the phase compensation is performed, and then normalized to obtain two equal magnitudes. The modified phase spectrum is obtained by splicing the two phase spectrums in the channel direction and then inputting them into the FEM module proposed in Chapter 3, and then inputting the output of the FEM and the amplitude spectrum into the DNN. Perform feature information fusion and extraction to obtain the amplitude spectrum of the music source signal predicted by DNN, and then perform ISTFT transformation together with the original phase spectrum of the mixed music signal to obtain the separated and predicted time domain music source signal. As shown in Figure 9.

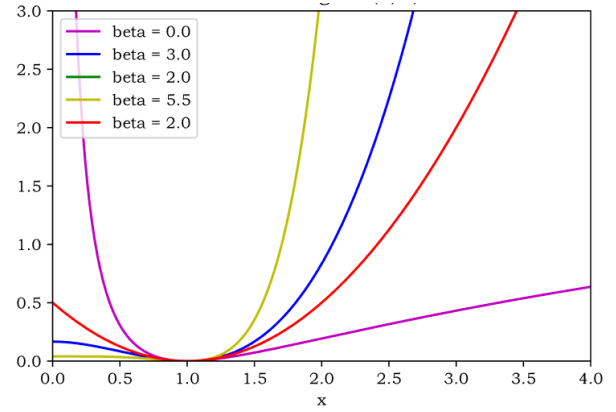


Figure 9: Amplitude spectrum and phase spectrum of music source signal

IV. Conclusion

This paper adopts the research method of combining historical theory and multi-discipline to explore the interactive relationship between the text and social objects of musical theatre singing, pay attention to the unity of musical theatre singing style and technical characteristics, and find its laws and characteristics. Taking the technology and market as the foothold, placing musicals in the context of aesthetic modernity, examining the dynamics of musicals, a product of mass life, in the entire historical trajectory from a multi-dimensional perspective, and studying its cultural production and consumption. The relationship between societies, and presents a cultural paradigm of modernity, entertainment, and artistry. By investigating traditional and deep neural network-based music source separation algorithms, it is found that the data-driven deep neural network-based music source separation algorithm has more advantages than traditional music source separation algorithms. In addition to better separation performance, it also has better generalization.

Funding

There is no funding support for this study.

References

- [1] Nassar, N., Jafar, A., & Rahhal, Y. (2020). A novel deep multi-criteria collaborative filtering model for recommendation system. *Knowledge-Based Systems*, 187, 104811.
- [2] Al Jawarneh, I. M., Bellavista, P., Corradi, A., Foschini, L., Montanari, R., Berrocal, J., & Murillo, J. M. (2020). A pre-filtering approach for incorporating contextual information into deep learning based recommender systems. *IEEE Access*, 8, 40485-40498.
- [3] Unger, M., Tuzhilin, A., & Livne, A. (2020). Context-aware recommendations based on deep learning frameworks. *ACM Transactions on Management Information Systems (TMIS)*, 11(2), 1-15.
- [4] Logesh, R., Subramaniaswamy, V., Malathi, D., Sivaramakrishnan, N., & Vijayakumar, V. (2020). Enhancing recommendation stability of collaborative filtering recommender system through bio-inspired clustering ensemble method. *Neural Computing and Applications*, 32(7), 2141-2164.
- [5] Yu, M., Quan, T., Peng, Q., Yu, X., & Liu, L. (2022). A model-based collaborative filtering algorithm based on stacked AutoEncoder. *Neural Computing and Applications*, 34(4), 2503-2511.
- [6] Liu, H., Wang, Y., Peng, Q., Wu, F., Gan, L., Pan, L., & Jiao, P. (2020). Hybrid neural recommendation with joint deep representation learning of ratings and reviews. *Neurocomputing*, 374, 77-85.

- [7] Chen, C., Zhang, M., Zhang, Y., Liu, Y., & Ma, S. (2020). Efficient neural matrix factorization without sampling for recommendation. *ACM Transactions on Information Systems (TOIS)*, 38(2), 1-28.
- [8] Ferrari Dacrema, M., Boglio, S., Cremonesi, P., & Jannach, D. (2021). A troubling analysis of reproducibility and progress in recommender systems research. *ACM Transactions on Information Systems (TOIS)*, 39(2), 1-49.
- [9] Zhou, J., Sun, J., Zhang, W., & Lin, Z. (2023). Multi-view underwater image enhancement method via embedded fusion mechanism. *Engineering Applications of Artificial Intelligence*, 121, 105946.
- [10] Zhou, J., Pang, L., Zhang, D., & Zhang, W. (2023). Underwater image enhancement method via multi-interval subhistogram perspective equalization. *IEEE Journal of Oceanic Engineering*, 48(2), 474-488.
- [11] Li, C., Kou, Y., Shen, D., Nie, T., & Li, D. (2024). Cross-Grained Neural Collaborative Filtering for Recommendation. *IEEE Access*, 12, 48853-48864.
- [12] Wu, K. (2023). Cultural Confluence: The Impact of Traditional and Modern Synergies in Chinese Juvenile Musical Theater. *International Journal of Education and Humanities*, 11(2), 192-199.
- [13] Hadar, T., & Rabinowitch, T. C. (2023). The varying social dynamics in orally transmitted and notated vs. improvised musical performance. *Frontiers in Psychology*, 14, 1106092.
- [14] Pino, M. C., Giancola, M., & D'Amico, S. (2023). The association between music and language in children: A state-of-the-art review. *Children*, 10(5), 801.
- [15] Chambers, S. (2023). The curation of music discovery: The presentation of unfamiliar classical music on radio, digital playlists and concert programmes. *Empirical Studies of the Arts*, 41(1), 304-326.

• • •