EXPLAINABLE MODELING OF GENDER-TARGETING PRACTICES IN TOY ADVERTISING SOUND AND MUSIC

Luca Marinelli, Charalampos Saitis

Centre for Digital Music, Queen Mary University of London

ABSTRACT

This study examines gender coding in sound and music, in a context where music plays a supportive role to other modalities, such as in toy advertising. We trained a series of binary XGBoost classifiers on handcrafted features extracted from the soundtracks and then performed SAGE and SHAP analyses to identify key audio features in predicting the gender target of the ads. Our analysis reveals that timbral dimensions play a prominent role and that commercials aimed at girls tend to be more harmonious and rhythmical, with a broader and smoother spectrum, while those targeting boys are characterised by higher loudness, spectral entropy, and roughness. Mixed audience commercials instead appear to be as rhythmical as girls-only ads, although slower, but show intermediate characteristics in terms of harmonicity and roughness. This study highlights the importance of music in shaping societal norms and the need for greater accountability in its use in marketing and other industries. We provide a public repository containing all code and data used in this study.

Index Terms— xai, applications, music information retrieval, gender and media

1. INTRODUCTION

The literature on gender coding in sound and music is sparse and contradictory [1, 2, 3], as it lacks an encompassing theoretical framework. We argue that the focus should not be on the composer's gender in relation to gendered musical material, but rather on the observable gendering of musical messages, as proposed by Tagg and Clarida [1, p. 665]. This line of inquiry is more readily structured in contexts where music plays a supportive role to other modalities, such as in advertisements. The current research specifically targets toy commercials, which are notoriously gender-polarised [4], thereby highlighting how advertising music might be utilised to reinforce traditional gender roles and stereotypes.

By extracting handcrafted features from the soundtracks of these commercials, and by employing Shapley Additive Global importancE (SAGE) [5] and SHapley Additive exPlanations (SHAP) [6] on a series of XGBoost [7] classifiers, our analysis centers on identifying the most relevant audio features in predicting the gender target of a commercial. This study thus delves into the application of handcrafted audio descriptors in eXplainable Artificial Intelligence (XAI) for music analysis, and examines the descriptors' interpretability, particularly in contexts beyond their design scope.

By modeling the underlying data generating processes, our aim is to uncover differences in the audio signals of commercials targeted at different gender-based customer segments, thereby providing insights into how sound and music choices might reflect the decision-making processes or unconscious gender biases of marketers. Ultimately, our research aims to lay the groundwork for a comprehensive theory of message production, in relation to the selection and composition of gender-based sound and music in toy advertisements. To the best of our knowledge, this investigation is the first of its kind to leverage XAI for exploring gender-coding in music.

2. BACKGROUND

2.1. XAI in social studies and music

A growing body of research in computational social studies leverages approaches from XAI to gain insights about complex relationships in the data that would be otherwise impossible to gain via more traditional methods such as parametric modeling. Nesa et al. [8] used SHAP to identify the primary correlates of crime due to adolescent drug addiction by pinpointing behavioral disorders and predicting the tendency to engage in criminal activities. Colak et al. [9] identified students at risk of dropping out of school in low-income countries, and highlighted how socio-economic characteristics like involvement in household farming and the father's education level are pivotal in predicting and preventing school dropout. In a study exploring explainable machine learning for predicting cleared homicides in the U.S., Campedelli [10] utilised SHAP to determine key features influencing clearance patterns, revealing significant elements such as the nature of the homicide, the weapon used, and demographics of victims and offenders, offering insights for targeted strategies to enhance police effectiveness in solving homicides.

In a study investigating the link between music preferences and moral values [11], SHAP analysis indicated that elements in lyrics associated with hierarchy and tradition values had higher predictive power compared to those related to empathy and equality, underscoring the role of music listening behaviors in understanding individual moral values. Furthermore, Fink et al. [12] surveyed over 5000 individuals across three continents during the first COVID-19 lockdown to examine changes in music listening for socio-emotional coping. The study found that more than half of the respondents used mu-

Luca Marinelli is a research student at the UKRI Centre for Doctoral Training in Artificial Intelligence and Music, supported jointly by UK Research and Innovation [grant number EP/S022694/1] and Queen Mary University of London. Code and data are available at https://github.com/marinelliluca/explainable-modeling. We thank Emmanouil Benetos for the helpful comments.

sic to cope, with those experiencing negative emotions using music for emotional regulation, and those with increased positive emotions using it as a substitute for social interaction. Utilising light gradient-boosted regressor models and SHAP, the most significant predictor of the use of music for coping was found to be an individual's interest in "coronamusic," highlighting the role of music in addressing socio-emotional needs during societal crises.

2.2. Sound and music in gendered toy marketing

Over four decades of research have consistently shown gender polarisation in children's TV advertising [13, 14, 15]. Such polarisation is in fact materialised in various elements of commercials targeted at girls, boys, and mixed audiences, including sound (such as voices and background music), language, camera techniques, setting, interactions, activities, and color schemes.

Regarding the auditory aspects, Welch et al. [13] observed that the gender of the voice-over typically aligns with the intended audience of the advertisement. However, they also noted a prevalence of male voices in commercials for mixed audiences, a trend later corroborated by others [15]. Additionally, it was found that advertisements aimed at boys tend to feature noisier backgrounds, louder music, and more sound effects [15]. A different research [14] discovered that music in advertisements targeting girls usually has a softer tone and often features sung narration. Whereas Johnson and Young [16] highlighted a phenomenon they termed "gender exaggeration," where male voice-overs are notably deep, growling, or aggressive, while female voice-overs possess a high-pitched, melodic quality.

2.3. Gender-coding in music

Gender-coding in sound and music ensues from the historical sedimentation, in musical practice, of multimodal associations between gendered meanings in language, visual images, and musical structures [17]. In particular, certain instruments have been consistently linked with masculine or feminine qualities, even when their sounds are heard separately from any visual association with the instrument itself [18]. Sergeant and Himonides [19, 2] explored if the sounds or arrangement in Western art music could indicate the sex or gender of the performer or composer. While they discovered no link between the gender of the composer or performer and the gendering of music, listeners consistently identified gendered aspects in music, related to features like minor or major keys and tonal weight or density. Wang and Horvát [3] analysed over 200k songs from more than 8k artists worldwide, spanning 40 years of pop music, using twelve computational descriptors of musical parameters and perceptual features. They discovered significant differences in eleven of these twelve parameters based on the composers' gender, indicating distinct, genre-transcending gendered music styles in the global music industry. Tagg [1, p. 665] researched the reception of gendered meanings in TV themes music and found a strong consensus among listeners. Factors like average tempo, consistency in rhythm and dynamics, and prominent bass lines were identified as key elements influencing the perception of gender in these tunes.

3. DATASET

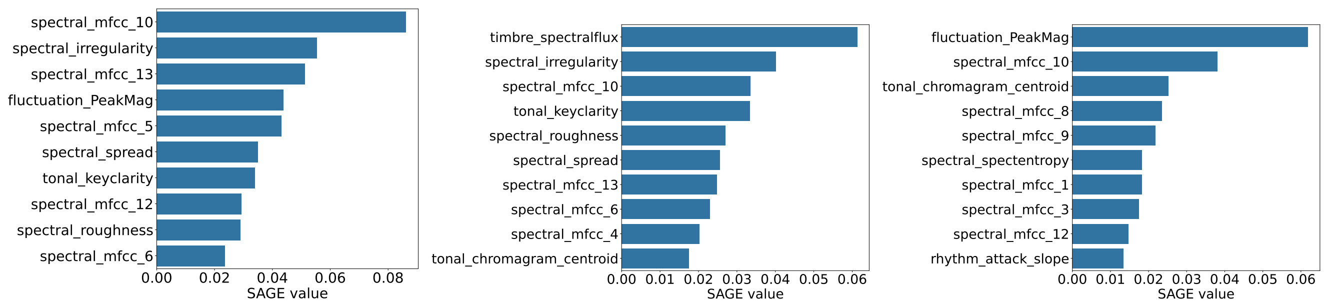
In March 2022, we collected a sample of 5614 videos from Smyths Toys Superstores' official YouTube channel, a leading toy retailer in the UK which features a wide range of prominent brands in the industry. To ensure comparability with previous studies [20, 21], we selected only high-quality videos intended for television. To minimise duplicates, we removed videos with the same title from our sample. Given that we are interested in understanding the gendering of sound and music in the toy industry at large, we needed to enforce some balance across gender targets. We thus performed a *preliminary* tagging of 1778 commercials based on their intended target audience (feminine, masculine or mixed audience) using simple heuristics regarding the gender of the majority of presenters featuring in the commercial, the colour coding of the video and ultimately the category of the product. This resulted in 780 'feminine', 509 'masculine', and 489 'mixed audience' commercials. A final sample of 606 commercials, spanning over 10 years from 2012 to 2022, was obtained by randomly sampling from each category 202 videos. This sample was then manually coded following a structured procedure.

The *gender orientation* (also *target* audience) of the commercials was determined by the gender of the presenters. Following [16], in order to account for tokenism, whenever a presenter of the other gender was included in the background or for just a few seconds, these were considered token gender representations and not explicit market orientations. All fictional characters, even when realistic (e.g. from a video game), were not considered as actors/presenters and the corresponding commercials were coded as having no actors. Following previous studies [21, 22], in order to determine the reliability of the variable, 15% of the commercials was double-coded by two coders independently. We obtained a Krippendorff's alpha level of .91 and therefore met the standards of reliability required for this type of analysis [22]. Out of 606 analysed commercials, 163 were targeted at a feminine audience (girls only), 149 at a masculine audience (boys only), 200 at a mixed audience and 94 featured no actors.

4. EXPERIMENTAL SETUP

As part of data pre-processing, to avoid the jingle of the retailer in the last 5 seconds of most soundtracks, we trimmed them accordingly. Audio features—such as tonality, rhythm, structures, etc—were extracted with the MIRtoolbox (ver. 1.8.1) [23] using the function *mirfeatures*. The initial set was reduced to the features mean, and delta Mel-Frequency Cepstral Coefficients (MFCCs) were also removed to avoid MFCC-related descriptors to dominate the feature space.

We then trained a series of XGBoost [7] classifiers to predict the gender orientation of the commercials in the following settings: girls vs boys-only, girls-only vs mixed audience, boys-only vs mixed audience. Each model was trained on a further reduced set (via recursive feature elimination with cross-validation, RFECV) and hyperparameters were tuned in an Optuna [24] study with 2000 trials. We compare three different binary models instead of training a single model on all three classes because it is imperative that the model under analysis reflects at best the underlying data generating processes. Whereas, in a previous study [25] we showed that when compared to those from feminine and masculine-targeted commer-



Boys-only vs girls-only | Girls-only vs mixed | Boys-only vs mixed

Fig. 1. Importance of the top ten features for each binary classification setting, from SAGE analysis.

cials, the soundtracks of mixed-audience commercials have much less pronounced differences, which results in a significant drop in the performance of models trained to discriminate between soundtracks from all three target audiences at the same time.

5. RESULTS

Model performance was assessed using leave-one-out cross-validation. The F1 score achieved for the boys-only vs girls-only categorisation was 0.91, utilising 32 features. Comparatively, the girls-only vs mixed and boys-only vs mixed classifications yielded F1 scores of 0.75 and 0.73, with 32 and 38 features, respectively. For the sake of interpretability, in the following we will first focus our analysis on the top ten most important features (SAGE analysis), and secondly on the top ten features excluding MFCCs (SHAP analysis).

5.1. Global feature importance

SAGE delves into global interpretability (as opposed to local explanation), aiming to understand the overall dependence of a model on each feature by assessing the influence of each feature on the model performance, that is, across the entire dataset. In Fig. 1 we can see how different features hold varying degrees of importance across the three classification tasks. The presence of multiple MFCC bands in the top ten features across all tasks (on average five out of ten) suggests that timbre dimensions [26] have a prominent role in the gender coding of music and sound. In particular, ‘spectral_mfcc_10’ ranks within the top three for all tasks, emphasising its pivotal role in differentiating gender-targeted commercials.

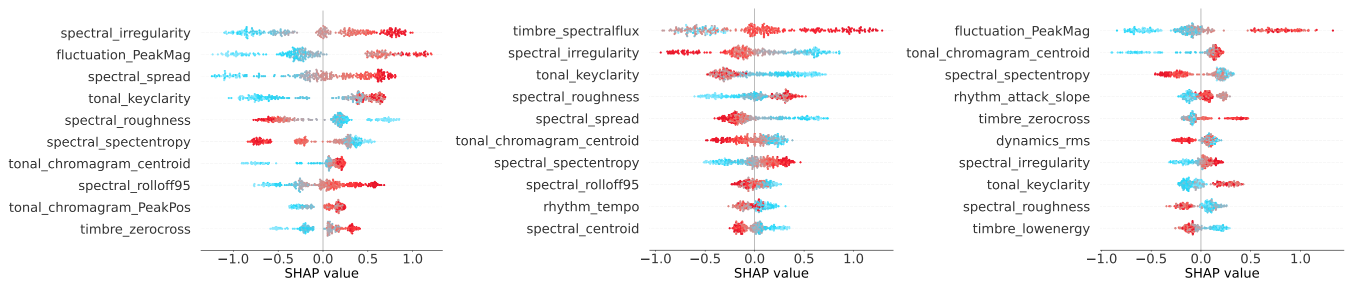
However, the remaining influential features show variability across tasks. In the comparison of ”boys-only vs girls-only” and ”girls-only vs mixed,” features such as ‘spectral_irregularity’, ‘spectral_spread’, ‘tonal_keyclarity’, and ‘spectral_roughness’ are consistently important. In contrast, the ”boys-only vs mixed” classification shares only ‘fluctuation_PeakMag’ with the ”boys-only vs girls-only” task and ‘tonal_chromagram_centroid’ with the ”girls-only vs mixed” task. This observation suggests that the models utilise a core set of features to distinguish between commercials aimed at boys-only and girls-only, as well as girls-only and mixed audiences. The models instead depend on a substantially different set of core features to differentiate between boys-only and mixed audience ads. Mixed-audience commercials may thus have more auditorily in common with boys-only than with girls-only ads.

5.2. Local feature contributions

SHAP focuses on the impact of individual features on individual predictions as it assigns each feature a value that represents whether it pushes the prediction—of a single data point—higher or lower. The decision, to focus here on the top ten features excluding MFCCs, is based on the understanding that while MFCCs are important correlates of timbre [26], they may not provide as much interpretive value for local contributions as other features would. This is aimed at deriving more meaningful insights into how clearly-interpretable audio characteristics contribute to gender coding in toy commercials.

Figure 2 shows the main differences in the audio signals of toy commercials targeted at different gender-based market segments. Let us proceed in order of appearance, by analysing the boys-only vs girls-only task. Spectral irregularity, as defined in Jensen’s work [27], is a measure of regularity of energy across a sound’s partials. It is the normalised sum of squared differences of the amplitude of adjacent partials in pseudo-harmonic sounds. Sometimes referred to as spectral deviation, it has been found to correlate with spectral fluctuations over time, one of the fundamental dimensions of timbre perception [28]. Though it might seem counter-intuitive at first glance, with higher values observed in commercials targeting the feminine audience compared to those aimed at masculine or mixed audiences, its behavior is likely a consequence of its original design for pseudo-harmonic sounds. In our context, this means that it might effectively serve as proxy for harmonicity. The fluctuation peak’s magnitude instead functions as a measure of rhythm regularity, with higher values correlating with girls-only commercials. Also the spectral spread (variance of the spectrum), the tonal key clarity (strength of the best estimated keys), the tonal chromagram centroid and position of peak, the spectral rollof at 95%, as well as the zero-crossing rate, all show higher values correlating with girls-only commercials. Whereas, spectral roughness and entropy both display higher values correlating with boys-only commercials.

Moving onto the girls-only vs mixed task, we first observe the notable absence of the fluctuation peak magnitude as an important predictor, indicating that girls-only and mixed audience commercials are both more rhythmical than boys-only ones (cf., next task), while no significant difference is emerging in the current task, al-



Boys-only vs girls-only   Girls-only vs mixed   Boys-only vs mixed

Fig. 2. SHAP analysis of the top ten non-MFCCs features for each binary classification setting, ordered in descending SAGE values. Low features values are depicted in cyan, middle values in grey, and high values in red. The positive class is always the second one in alphabetic order (e.g., girls, in boys vs girls).

though the tempo appears to be slower in mixed audience ads. The spectral flux instead indicates that girls-only commercials have on average a smoother spectrum compared to mixed audience ones. We also observe that the spectral centroid and the tempo are both typically higher for girls-only commercials. The rest of the descriptors in this task instead all display the same behaviour as the previously analysed task, where this time mixed audience commercials take the place of boys-only ones, though with less pronounced differences.

From a quick analysis of the boys-only vs mixed task, we can observe mixed audience commercials this time taking on the place of girls-only commercials, for all previously discussed features. We then observe in terms of this task's specific differences, that the attack slope of rhythmic onsets is more pronounced for mixed audience commercials, that boys-only commercials are typically louder (dynamics RMS), and finally that the low energy rate (percentage of frames with less-than-average energy) is higher in boys-only commercials, which in turn indicates that subsequent frames are on average more contrastive than in mixed audience ads.

6. DISCUSSION

In our research, we learned that while global importance methods are effective for grasping overarching patterns, like the significance of MFCCs in our scenario, true interpretability depends on the specific features under scrutiny and their local impact. The combination of global and local methods resulted in greater intelligibility, where SAGE set the context and SHAP elucidated the detailed contributions. Although the most influential features differed across the tasks, a core set was consistently used by the models to distinguish between boys-only and girls-only commercials, and between girls-only and mixed audience ads. Notably, the core set for distinguishing boys-only commercials from those targeting mixed audiences differs substantially, suggesting that mixed-audience ads share more auditory characteristics with those aimed at boys-only than girls-only. Supporting this observation, the model for boys vs mixed required the most features (after RFECV) and also exhibited the lowest performance. Previous research highlighted this issue, as advertisements targeted at both boys and girls predominantly feature a masculinised style, which in turn seems to expose children to a broader narrative of dominant masculinity in media [15].

Analysing the SHAP results revealed distinct musical patterns in gender-targeted commercials. Commercials aimed at girls showcased greater harmony and rhythmic consistency, along with the broadest and smoothest spectrum. In contrast, boys-only advertisements were characterised by higher loudness, spectral entropy, and roughness, in accordance with previous studies [15]. Mixed audience commercials interestingly mirrored the rhythmic regularity of girls-only ads, yet their values for harmonicity and roughness were intermediate compared to both gender-specific categories. Notably, their spectral smoothness was significantly less pronounced than that observed in girls-only commercials. Gendered music styles in toy commercials appear thus to emerge as a result of deliberate marketing strategies, as such styles reflect gender stereotypes that are "ludicrously old-fashioned and offensively out of touch" [4] and still too largely represented in the industry.

7. CONCLUSION

Results highlighted the need for a careful understanding of the nature and intended use of features in XAI models. This is particularly relevant when features, such as the audio descriptors used here, are applied to contexts different from those they were originally designed for. A potential limitation of this study is that certain descriptors, like zero-crossing rate and spectral spread, which are primarily designed for isolated, monophonic sounds, might be less reliable in polyphonic settings where multiple sound sources overlap. Moreover, by compressing time series features into average values for each advertisement, our method potentially overlooks relevant differences in temporal variations of the audio signals. This research contributes to a deeper and unmediated understanding of how sound and music in toy advertisements reflect and reinforce gender biases, although having collected commercials intended for the UK, we can not claim cross-cultural generalisability of the results. Nonetheless, the results emphasise the significance of music's role in forming societal norms and calls for increased responsibility in its applications in marketing. Future studies could delve into how, and to which extent gendered music is leveraged in advertising tactics across various industries, its effects on consumer behavior and societial perception of gender, and ways to foster more inclusive marketing strategies.

8. REFERENCES

[1] Philip Tagg and Bob Clarida, *Ten little title tunes: Towards a musicology of the mass media*, Huddersfield: The Mass Media Music Scholars' Press, NY., 2003.

[2] Desmond C Sergeant and Evangelos Himonides, "Gender and music composition: A study of music, and the gendering of meanings," *Frontiers in Psychology*, vol. 7, pp. 411, 2016.

[3] Yixue Wang and Emőke-Ágnes Horvát, "Gender differences in the global music industry: Evidence from musicbrainz and the echo nest," in *Proceedings of the International AAAI Conference on Web and Social Media*, 2019, vol. 13, pp. 517–526.

[4] Cordelia Fine and Emma Rush, ""Why does all the girls have to buy pink stuff?" The ethics and science of the gendered toy marketing debate," *Journal of Business Ethics*, 2018.

[5] Ian Covert, Scott M Lundberg, and Su-In Lee, "Understanding global feature contributions with additive importance measures," *Advances in Neural Information Processing Systems*, vol. 33, pp. 17212–17223, 2020.

[6] Scott M Lundberg and Su-In Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[7] Tianqi Chen and Carlos Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[8] Meherun Nesa, Tumpa Rani Shaha, and Young Yoon, "Prediction of juvenile crime in bangladesh due to drug addiction using machine learning and explainable ai techniques," *Journal of Computational Social Science*, vol. 5, no. 2, pp. 1467–1487, 2022.

[9] Hazal Colak Oz, Çiçek Güven, and Gonzalo Nápoles, "School dropout prediction and feature importance exploration in malawi using household panel data: machine learning approach," *Journal of Computational Social Science*, vol. 6, no. 1, pp. 245–287, 2023.

[10] Gian Maria Campedelli, "Explainable machine learning for predicting homicide clearance in the united states," *Journal of Criminal Justice*, vol. 79, pp. 101898, 2022.

[11] Vjosa Preniqi, Kyriaki Kalimeri, and Charalampos Saitis, "More than words": Linking music preferences and moral values through lyrics," in *ISMIR 2022 Hybrid Conference*, 2022.

[12] Lauren K Fink, Lindsay A Warrenburg, Claire Howlin, William M Randall, Niels Chr Hansen, and Melanie Wald-Fuhrmann, "Viral tunes: changes in musical behaviours and interest in coronamusic predict socio-emotional coping during covid-19 lockdown," *Humanities and Social Sciences Communications*, vol. 8, no. 1, 2021.

[13] Renate L Welch et al., "Subtle sex-role cues in children's commercials.," *Journal of Communication*, 1979.

[14] Jenny Lewin-Jones and Barbara Mitra, "Gender roles in television commercials and primary school children in the uk," *Journal of Children and Media*, 2009.

[15] Barbara Mitra and Jenny Lewin-Jones, "Colin won't drink out of a pink cup," in *The Handbook of Gender, Sex, and Media*. Wiley Online Library, 2012.

[16] Fern Johnson and Karren Young, "Gendered voices in children's television advertising," *Critical Studies in Media Communication*, 2002.

[17] Nicola Dibben, "Gender identity and music," in *Musical identities*. New York: Oxford University Press, 2002.

[18] Lisa M Stronsick, Samantha E Tuft, Sara Incera, and Conor T McLennan, "Masculine harps and feminine horns: Timbre and pitch level influence gender ratings of musical instruments," *Psychology of Music*, 2018.

[19] Desmond C Sergeant and Evangelos Himonides, "Gender and the performance of music," *Frontiers in Psychology*, 2014.

[20] Mary Strom Larson, "Interactions, activities and gender in children's television commercials: A content analysis," *Journal of Broadcasting & Electronic Media*, 2001.

[21] Susan G Kahlenberg and Michelle M Hein, "Progression on nickelodeon? gender-role stereotypes in toy commercials," *Sex Roles*, 2010.

[22] Kimberly A Neuendorf, "Content analysis — A methodological primer for gender research," *Sex Roles*, 2011.

[23] Olivier Lartillot and Petri Toiviainen, "A matlab toolbox for musical feature extraction from audio," in *10th Int. Conference on Digital Audio Effects (DAFx-07)*. Bordeaux, 2007, vol. 237, p. 244.

[24] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019, pp. 2623–2631.

[25] Luca Marinelli, George Fazekas, and Charalampos Saitis, "Gender-coded sound: Analysing the gendering of music in toy commercials via multi-task learning," in *ISMIR 2023 Hybrid Conference*, 2023.

[26] Hiroko Terasawa, Malcolm Slaney, and Jonathan Berger, "The thirteen colors of timbre," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005*. IEEE, 2005, pp. 323–326.

[27] Kristoffer Jensen, *Timbre models of musical sounds*, Ph.D. thesis, Department of Computer Science, University of Copenhagen Copenhagen, 1999.

[28] Kai Siedenburg, Ichiro Fujinaga, and Stephen McAdams, "A comparison of approaches to timbre descriptors in music information retrieval and music psychology," *Journal of New Music Research*, vol. 45, no. 1, pp. 27–41, 2016.