Gender-Coded Sound:
Analysing the Gendering of Music in Toy Commercials via Multi-Task Learning

Luca Marinelli
QMUL
l.marinelli@qmul.ac.uk

György Fazekas
QMUL

Charalampos Saitis
QMUL

Go to the demo:

Open in Colab

Highlights

- We outline a method to automatically analyse how music in TV advertising aimed at children is deliberately used to reinforce traditional gender roles.
- Our dataset of 606 commercials included music-focused mid-level perceptual features, multimodal aesthetic emotions, and content analytical items. Despite the limited size, we obtained noteworthy results by leveraging multi-task transfer learning.
- Models were trained to categorise commercials into masculine, feminine, and mixed audiences. To provide explainability for the classification, the models were jointly trained to perform regressions on emotion ratings and on mid-level perceptual attributes.

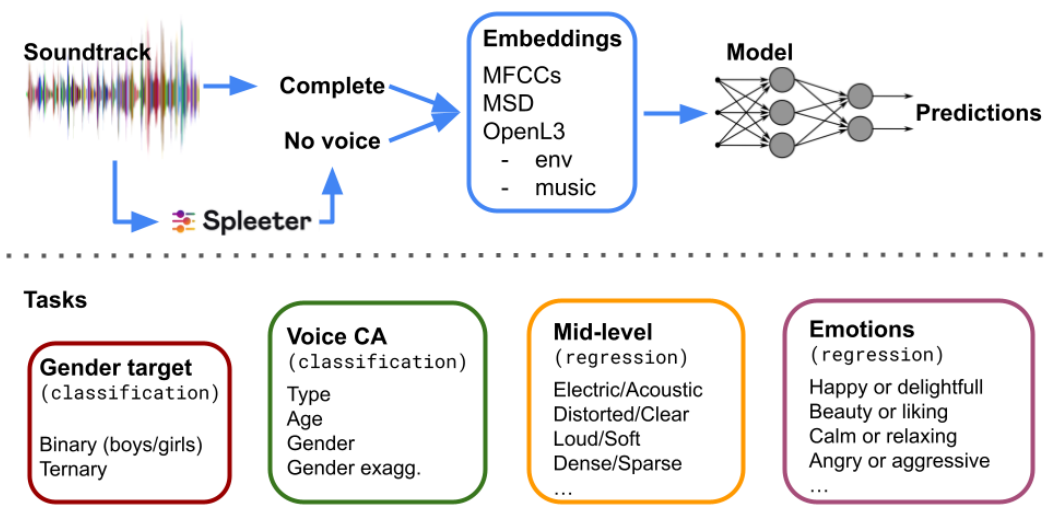

Figure 1. Brief overview of the experimental pipeline.

Embeddings	Voice	Target F1	Secon. F1	Avg. R^2 emo	Avg. r emo	Avg. R^2 mid	Avg. r mid
mfcc	no	.79 ± .08	.66 ± .07	.02 ± .17	.36 ± .11	.14 ± .16	.48 ± .09
mfcc	yes	.78 ± .10	.65 ± .07	.06 ± .16	.38 ± .11	.13 ± .15	.43 ± .10
msd	no	.87 ± .05	.66 ± .06	.25 ± .11	.54 ± .08	.35 ± .14	.62 ± .09
msd	yes	.95 ± .04	.79 ± .05	.26 ± .15	.56 ± .09	.30 ± .12	.58 ± .09
openl3_env	no	.91 ± .05	.72 ± .06	.34 ± .10	.61 ± .08	.41 ± .10	.66 ± .07
openl3_env	yes	.95 ± .04	.77 ± .05	.34 ± .13	.62 ± .08	.35 ± .12	.62 ± .08
openl3_music	no	.87 ± .09	.71 ± .06	.31 ± .16	.56 ± .19	.39 ± .16	.64 ± .15
openl3_music	yes	.91 ± .11	.76 ± .10	.29 ± .17	.56 ± .19	.31 ± .14	.59 ± .13

Table 1. Mean and std from 5x repeated 5-fold CV (binary gender Target task).

Embeddings	Voice	Target F1	Secon. F1	Avg. R^2 emo	Avg. r emo	Avg. R^2 mid	Avg. r mid
mfcc	no	.52 ± .05	.67 ± .06	.04 ± .16	.37 ± .11	.14 ± .14	.48 ± .09
mfcc	yes	.48 ± .04	.67 ± .05	.05 ± .15	.38 ± .11	.15 ± .14	.46 ± .09
msd	no	.62 ± .05	.67 ± .06	.23 ± .15	.54 ± .10	.36 ± .12	.64 ± .07
msd	yes	.67 ± .06	.80 ± .05	.29 ± .12	.57 ± .08	.33 ± .10	.60 ± .07
openl3_env	no	.59 ± .06	.72 ± .06	.30 ± .12	.59 ± .08	.42 ± .10	.66 ± .07
openl3_env	yes	.66 ± .07	.77 ± .06	.34 ± .11	.61 ± .07	.35 ± .10	.62 ± .07
openl3_music	no	.64 ± .07	.73 ± .06	.32 ± .12	.60 ± .08	.43 ± .11	.68 ± .07
openl3_music	yes	.67 ± .04	.78 ± .05	.35 ± .12	.61 ± .08	.37 ± .10	.63 ± .07

Table 2. Same as Table 2, but results refer to ternary Target classification.

Background

- Evidence from musicology and psychology research shows that gender-loaded messages can be reliably encoded and decoded via musical sounds.
- Differences in commercials targeted at girls, boys, and mixed audiences have been found in terms of sound (voices, background music and sound effects), language, transitions and camera work, setting, interactions and activities, and colours.
- Works that try to address the rhetoric of multi-media content are still in their infancy. To the best of our knowledge, none of such studies has yet leveraged approaches and tools from music information retrieval.

Methodology

- Hierarchical data collection (Fig. 1 >> Tasks)
  - Content analysis (CA): double coded 15%, Cronbach's α > .8
    - Gender target (binary and ternary)
    - Voice(over) (type, age, gender, gender exaggeration)
  - Mid-level perceptual features (soundtrack only)
    - 152 musically trained paid part., UK, 77 F, 75 M, 40 ± 14 yo
    - 6 ratings per stimulus (minimum 5)
    - 15 music-focused bipolar scales
  - Aesthetic emotions (video and soundtrack)
    - 151 paid part., UK, 75 F, 76 M, 39 ± 13 yo
    - 6 ratings per stimulus (minimum 5)
    - 7 aesthetic emotions scales from AESTHEMOS [9]
- Multi-task learning models trained on various embeddings (Fig. 1) to simultaneously predict mid-level features regression, emotion regression, and all the CA classes.
  - Shared initial hidden layer with 128 units and then branch out
  - Each sub-task has its own hidden layer with 128 units

Results

- Tables 1 and 2 reveal stark differences between the soundtracks of commercials designed for feminine and masculine audiences.
- Binary classification: with voice F1 = .95 ± .04, without F1 = 91 ± .11.
- Ternary classification: with voice F1 = .67 ± .04, without F1 = .64 ± .07
  - ➔ mixed-audience commercials don't have a distinct style
- Mid-level perceptual descriptors were best modeled without human voice
  - ➔ participants were indeed able to focus on the background

Conclusions

- By examining the performance of different musical embeddings in classifying commercials targeted at different audiences, and by providing explainable inference, this study sheds light on the role of music in gendered marketing strategies.
- Such approach has significant implications for advertisers, policymakers, and broadcasters, who recently faced a public backlash against gendered toy marketing.
- This study highlights the importance of considering the role of music when regulating marketing strategies and developing more inclusive and diverse advertising campaigns.

This work was supported by the UKRI and EPSRC
[grant number EP/ S022694/1]

UK Research and Innovation
Communication Acoustics Laboratory
AI + MUSIC
centre for digital music
Queen Mary University of London