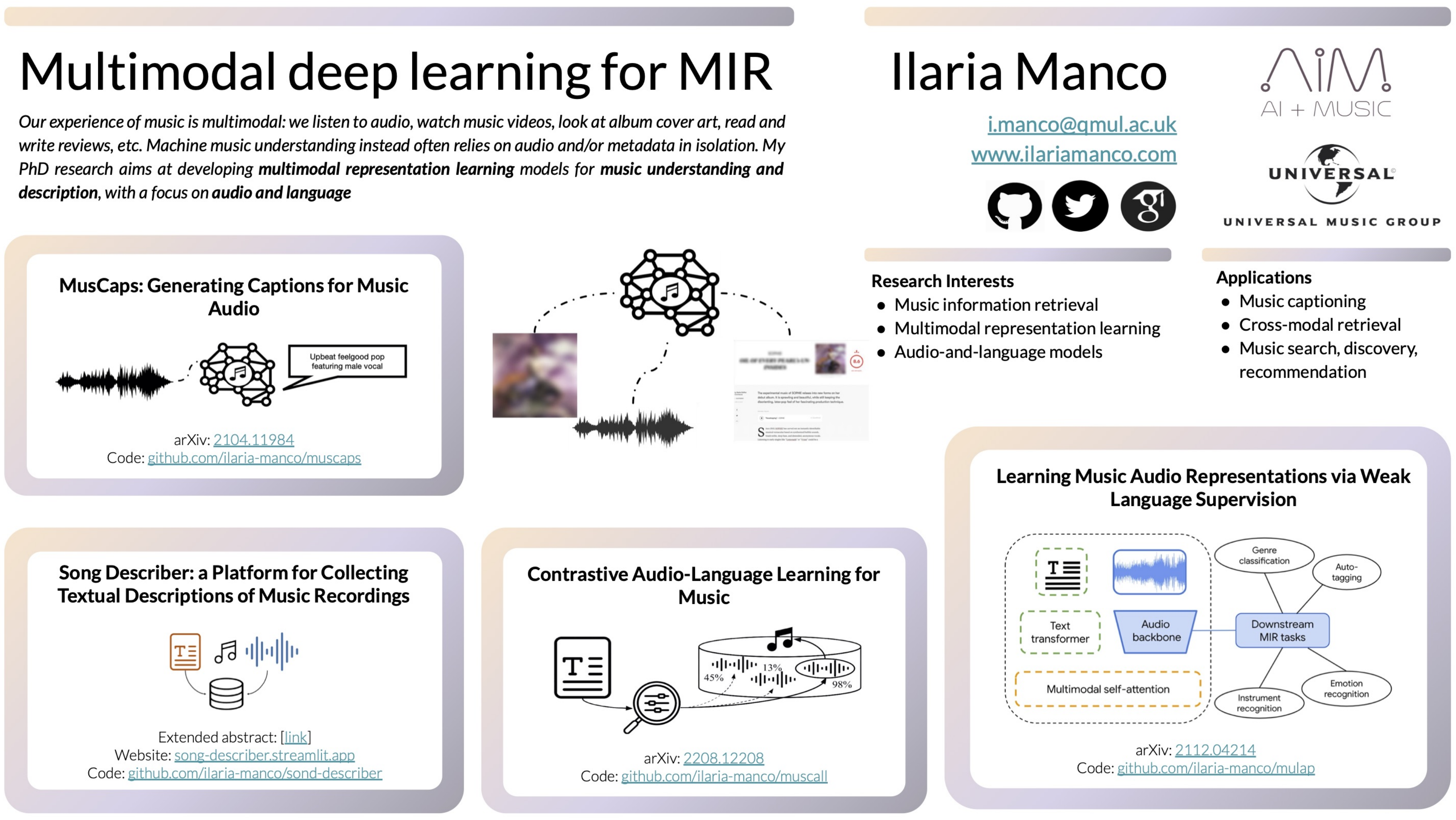
AIM Summer School 2023

Bridging audio and language for music understanding

Ilaria Manco

2019



Multimodal deep learning for MIR

*Our experience of music is multimodal: we listen to audio, watch music videos, look at album cover art, read and write reviews, etc. Machine music understanding instead often relies on audio and/or metadata in isolation. My PhD research aims at developing **multimodal representation learning** models for **music understanding and description**, with a focus on **audio and language***

Ilaria Manco

i.manco@qmul.ac.uk
www.ilariamanco.com

Research Interests

- Music information retrieval
- Multimodal representation learning
- Audio-and-language models

Applications

- Music captioning
- Cross-modal retrieval
- Music search, discovery, recommendation

MusCaps: Generating Captions for Music Audio

arXiv: 2104.11984
Code: github.com/ilaria-manco/muscaps

Song Describer: a Platform for Collecting Textual Descriptions of Music Recordings

Extended abstract: [link]
Website: song-describer.streamlit.app
Code: github.com/ilaria-manco/sond-describer

Contrastive Audio-Language Learning for Music

arXiv: 2208.12208
Code: github.com/ilaria-manco/muscall

Learning Music Audio Representations via Weak Language Supervision

arXiv: 2112.04214
Code: github.com/ilaria-manco/mulap

Finding

Joint audio-language learning leads to better music representations and enables new music-related tasks.

Question

Are we doing joint multimodal training right?

Supervisor(s): Emmanouil Benetos, George Fazekas, Elio Quinton