Live Soundtracks for Video Conferences
Sonic Thinking Winter Term 2021/22

Clemens Tiedt
clemens.tiedt@student.hpi.de

HPI-ITSE

December 12, 2021

Motivation

Motivation

- Empathy relies on many clues (body language, facial expression, voice cues)
- Online communication removes/degrades some cues, e.g.
  - Restricted camera angles
  - or no camera at all
  - compression applied to voice data

Approach

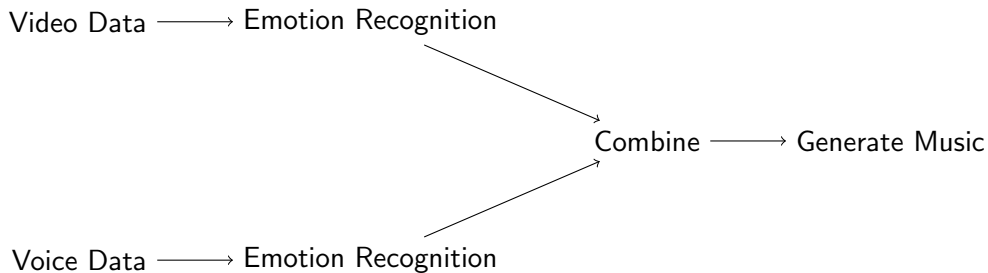# Approach

Approach

Approach

Approach

Figure: Overview of a possible pipeline

Tasks in this project

1. Recognize emotion from video and voice data
2. Generate music matching the emotional state
3. Do both of these *live*

Emotion Recognition

- Much previous work (Google Scholar: about 4.150.000 results)
- Some existing solutions are already useable live
- Maybe interesting: Machine Learning VS "manual" analysis
- Remaining task: Find appropriate solutions and possibly combine them

Music Generation

- Previous work exists, but more on recognizing emotion in music
- Many Machine Learning solutions
- Generating basic music is not that difficult
- Convey emotion e.g. through
    - speed and pauses
    - changes in pitch
    - different instruments

Live Music Generation

- Challenge: Played music should fit current state of emotion recognition

Live Music Generation

- ▶ Challenge: Played music should fit current state of emotion recognition
- ▶ Approach: Generate new music fragments in regular intervals
- ▶ Choose interval length so that
  - ▶ a melody fragment fits into it
  - ▶ the fragment is generated from recent data
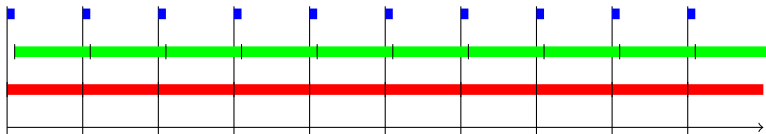
Figure: Schematic overview of music generation. The red line represents running emotion recognition, the blue one music generation, and the green one music playback.

Overview of Related Work

- Soundtracks and Empathy: Ansani et al. (2020)
- Speech Emotion Recognition: Huang et al. (2014)
- Facial Emotion Recognition: Mehendale (2020)
- Music Generation: Miyamoto et al. (2020)

Thank you for your attention!

References I

Alessandro Ansani, Marco Marini, Francesca D'Errico, and Isabella Poggi. How Soundtracks Shape What We See: Analyzing the Influence of Music on Visual Scenes Through Self-Assessment, Eye Tracking, and Pupillometry. *Frontiers in Psychology*, 11:2242, 2020. ISSN 1664-1078. doi: 10.3389/fpsyg.2020.02242. URL `https://www.frontiersin.org/article/10.3389/fpsyg.2020.02242`.

Zhengwei Huang, Ming Dong, Qirong Mao, and Yongzhao Zhan. Speech Emotion Recognition Using CNN. In *Proceedings of the 22nd ACM international conference on Multimedia*, MM '14, pages 801–804, New York, NY, USA, November 2014. Association for Computing Machinery. ISBN 978-1-4503-3063-3. doi: 10.1145/2647868.2654984. URL `https://doi.org/10.1145/2647868.2654984`.

Ninad Mehendale. Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*, 2(3):446, February 2020. ISSN 2523-3971. doi: 10.1007/s42452-020-2234-1. URL `https://doi.org/10.1007/s42452-020-2234-1`.

Kana Miyamoto, Hiroki Tanaka, and Satoshi Nakamura. Music Generation and Emotion Estimation from EEG Signals for Inducing Affective States. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*, ICMI '20 Companion, pages 487–491, New York, NY, USA, October 2020. Association for Computing Machinery. ISBN 978-1-4503-8002-7. doi: 10.1145/3395035.3425225. URL `https://doi.org/10.1145/3395035.3425225`.