Fraunhofer IDMT

Software Solutions for Automatic Music Analysis, Audio Matching, and Machine Listening

Semantic Music Technologies

Fraunhofer Institute for Digital Media Technology

- Round 200 employees, apprentices, PhD students, interns, research and student assistants
- Headquarters Ilmenau and Branch Oldenburg
- Business Units
  - Acoustics
  - Media Management and Delivery
  - Industrial Media Applications
  - Oldenburg Branch for Hearing, Speech and Audio Technology HSA

Head office Ilmenau

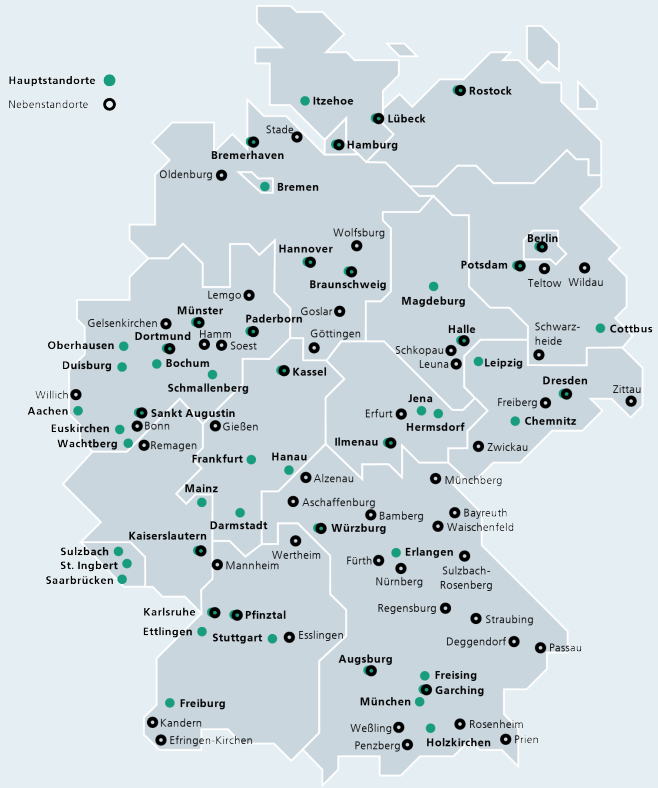
The Fraunhofer-Gesellschaft

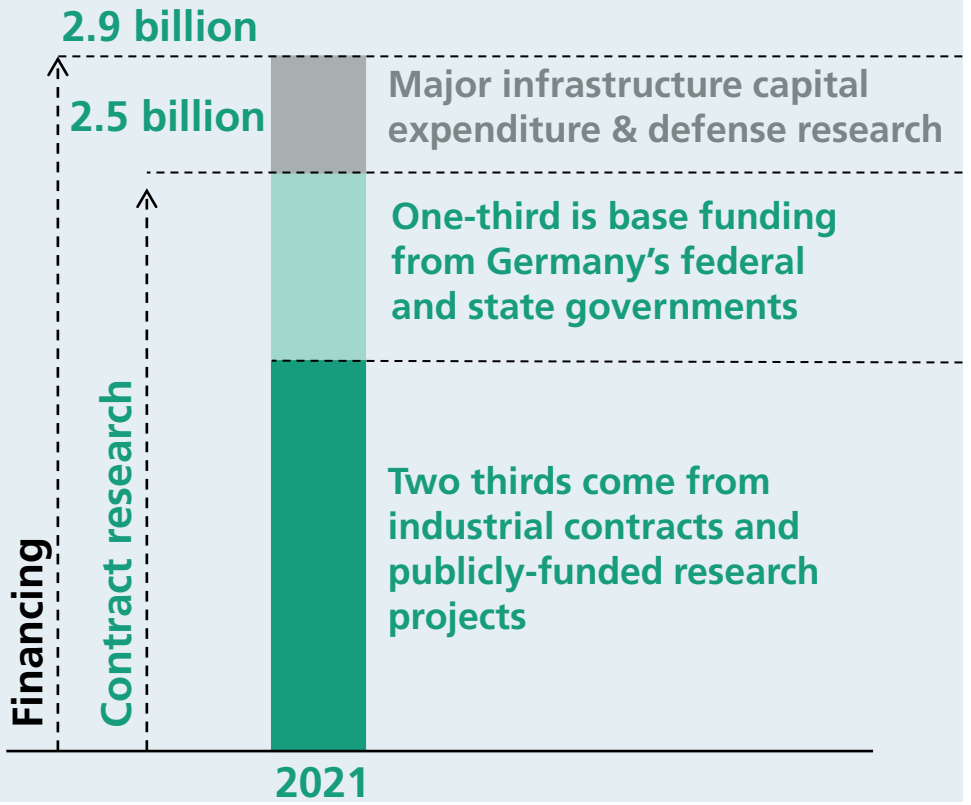
At a glance

Applied research with a focus on key future-relevant technologies and the commercialization of findings in business and industry. A trailblazer and trendsetter in innovative developments.

> 30,000 employees

76 institutes and research units





Fraunhofer IDMT

Semantic Music Technologies Group at Fraunhofer IDMT

Development of algorithms for signal processing and machine learning

- Audio Matching
  - Detection of media content via fingerprinting for broadcast monitoring, usage analysis, content tracking
- Automatic audio and music analysis
  - Search and fast discovery of content using metadata for music search and recommendation engines, music games and music learning programs as well as in music production
- Machine Listening
  - Sound detection and interpretation for environmental, noise, and traffic monitoring as well as acoustic monitoring for industrial quality control

Semantic Music Technologies Group at Fraunhofer IDMT

The Team

Hanna Lukashevich
Head of Semantic Music Technologies

Jakob Abeßer
Senior Scientist

Michael Tänzer
PhD student

Andrew McLeod
PostDoc

Holger Großmann
Software Developer

Patrick Kramer
Software Developer

Audio Matching

Fraunhofer IDMT

What is Audio Matching?

Goals

- Detection of a specific “recording” in a database
  - This is not similarity search!

Solutions

- Fingerprinting
  - Creation of a tiny representation (fingerprint) with suitable features and matching with a known reference database
  - Watermarking
  - Embed perceptually neutral information as a digital watermark

What is Audio Matching?

Use Cases

- TV & radio monitoring: Recognition of an item (song / commercial) even under noisy conditions
- De-duplication and repository cleanup
- Content reuse & rights tracking, metadata propagation
- Analysis of repetition and content propagation in a radio / TV program
- Detection of sampling plagiarism

Audio Fingerprinting
TV Audience Measurement

Developed by Fraunhofer IDMT and GfK Telecontrol

- Live and time-shifted audience measurement
- Based on audio fingerprinting, 1,000 channels can be measured simultaneously

Result

- Tiny fingerprints: 8MB/day meter fingerprint and 32MB/day reference fingerprint
- Time resolution of 1 second in use in TC UMC measurement device at GfK SE
- Used by Arbeitsgemeinschaft Fernsehfoschung AGF since summer 2012
- Technology licensed to GfK Telecontrol worldwide

Audio Fingerprinting

Radio Audience Measurement

Developed by Fraunhofer IDMT and GfK Telecontrol

- Live and time-shifted audience measurement
- Based on audio fingerprinting
- Robust against background noise (up to -15 dBA SNR) and audio and transmission quality

Result

- Tiny fingerprints: 5MB/day meter FP and 30MB/day reference FP
- Robust detection: up from 3 seconds of audio for clean signals and up from 20 seconds of audio for noisy signals
- Used in GfK MediaWatch, GfK's global radio audience measurement system integrated in a wristwatch

Music And Speech Detection

Fraunhofer IDMT

Music and Speech Detection

Goals

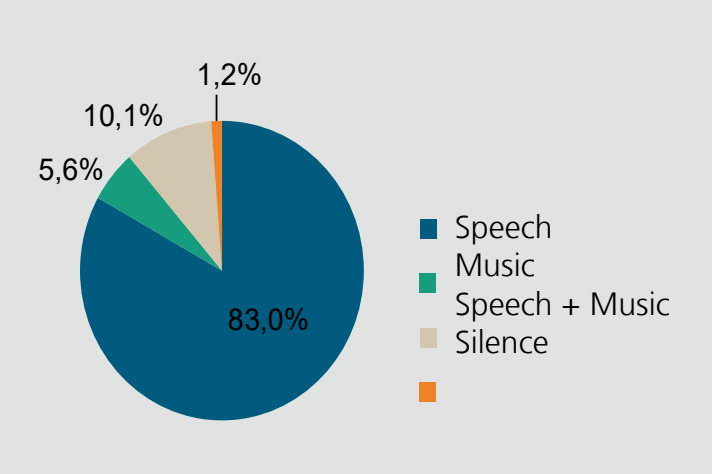
- Determine the exact amount of music and speech in radio and TV programs
- Automatic classification of TV- or Radio content into
  - Silence, Speech, Music, Speech and Music, other

Challenges

- Segment boundaries, mixes and masking, perception boundaries, cleanup for jingles and production music.
- Additional classes depending on use case, e.g., applause detection

Use Cases

- Optimize broadcasting programs
- Provide accurate accounting for copyright agencies



Fraunhofer IDMT

Music and Speech Detection

Current Research

- Compression of deep learning-based methods for large-scale analysis of music archives
- Task-specific treatment of background music
- Improved robustness against
  - Audio quality variations
  - Languages / dialects

Semantic Music Technologies

Research Topics

Music Recommendation

Fraunhofer IDMT

Music Recommendation

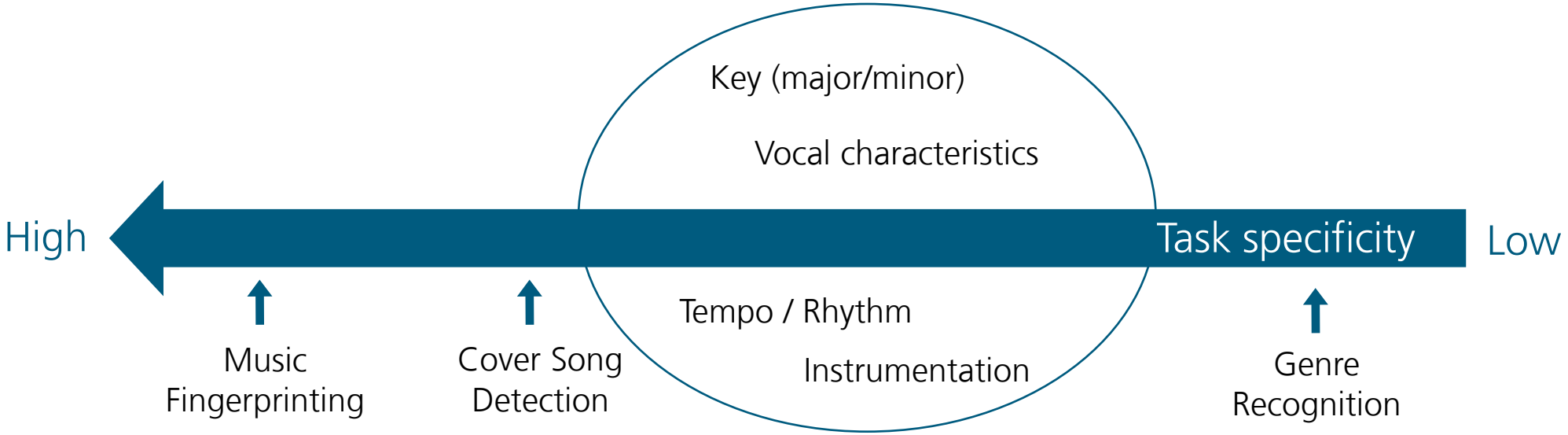
How to search for alternative music?

You can search

- by title, artist, album: Access only possible with additional knowledge
- via metadata/descriptions: genre, mood, instrumentation, context, ...
- Search by example: Query by example

Music Recommendation

What does music similarity mean?



Fraunhofer IDMT

SoundsLike

Music Annotation and Music Similarity

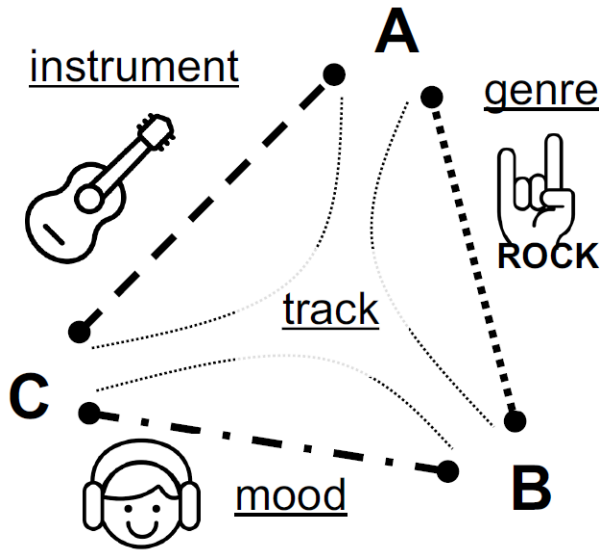
AI-based music similarity search and automatic annotation of music properties

- Music Annotation via automatic metadata extraction (selection)
  - Genre (Classical, Country, Electronica, Jazz, Latin, Pop, Rap, Rock,…)
  - Emotion (Anxious, Depressed, Exuberant, Content)
  - Instrumental Density (Full, Sparse), Dynamic (Continuous, Changing)
- Music Similarity
  - Multidimensional across multiple similarity profiles
  - Timbre, harmony, rhythm
  - Adaptable through AI-based learned embeddings depending on the definition of music similarity

Music Recommendation

Current Research

- Disentanglement Learning for music similarity
- Time-dependent harmonic similarity
- Instrument recognition in loops
- Rhythm pattern similarity



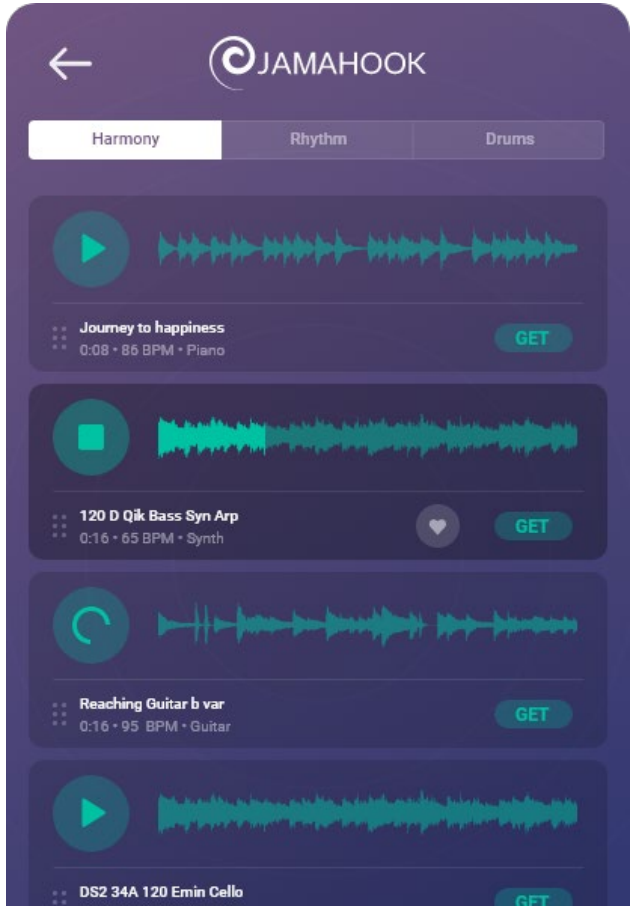
Source: Lee, J. et al (2020). Disentangled Multidimensional Metric Learning for Music Similarity. In Prof. of *ICASSP 2020*.

Fraunhofer IDMT

Reference project

Jamahook: Search engine for loops and beats based on SoundsLike

- Developed by Fraunhofer IDMT and Jamahook
- (VST) plugin for digital audio workstations
- Allows music producers to find the most suitable audio loops for a production based on examples extracted from an audio mix
- Based on Fraunhofer IDMT's SoundsLike AI-based music classification technologies



Fraunhofer IDMT

Semantic Music Technologies

Research Topics

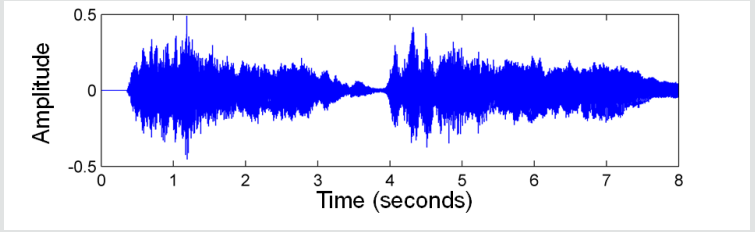
Automatic Music Transcription

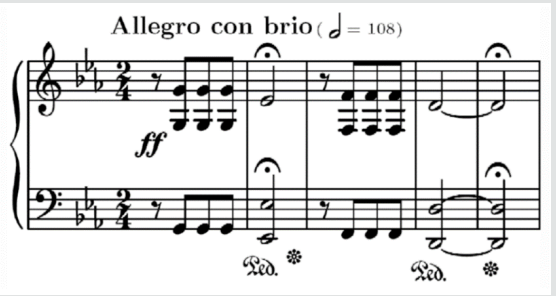
Fraunhofer IDMT

Automatic Music Transcription

Convert audio recordings to musical (score) notations

- Relevant Tasks
  - Predominant Melody Transcription
  - Polyphonic Transcription (Piano, Guitar, etc.)
  - Bass Transcription
  - Drum Transcription
  - Beat Tracking
  - Chord Recognition
  - Music Segmentation (Chorus, Verse, etc.)





Müller, M. (2021). Fundamentals of Music Processing Using Python and Jupyter Notebooks (2nd ed.), p. 2, Fig. 1.1

Fraunhofer IDMT

Automatic Music Transcription

Current Research

- Deep-learning based real-time melody / bass / polyphonic / drum transcription
- Compensating domain shift in real-world music learning application scenarios
- Beat Tracking
- Large-vocabulary chord recognition
- Music instrument recognition & separation

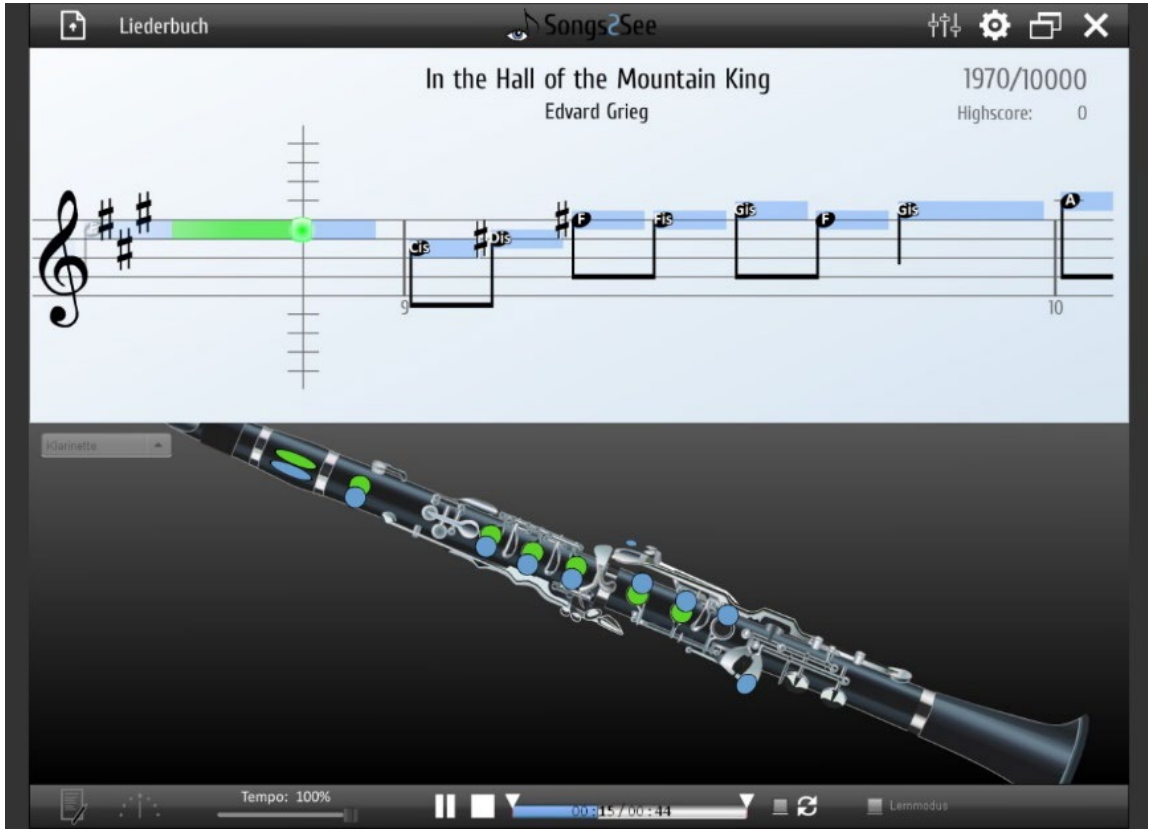
Automatic Music Transcription

Songs2See

Music learning software based on automatic music transcription

Songs2See Game

- Similar to karaoke games
- Listens to the user and gives points
- Can be played with real acoustic instruments
- User trains musical and technical skills
- Real-time polyphonic pitch recognition
- Slowing down, tablature, automatic transposition, fingering



Fraunhofer IDMT

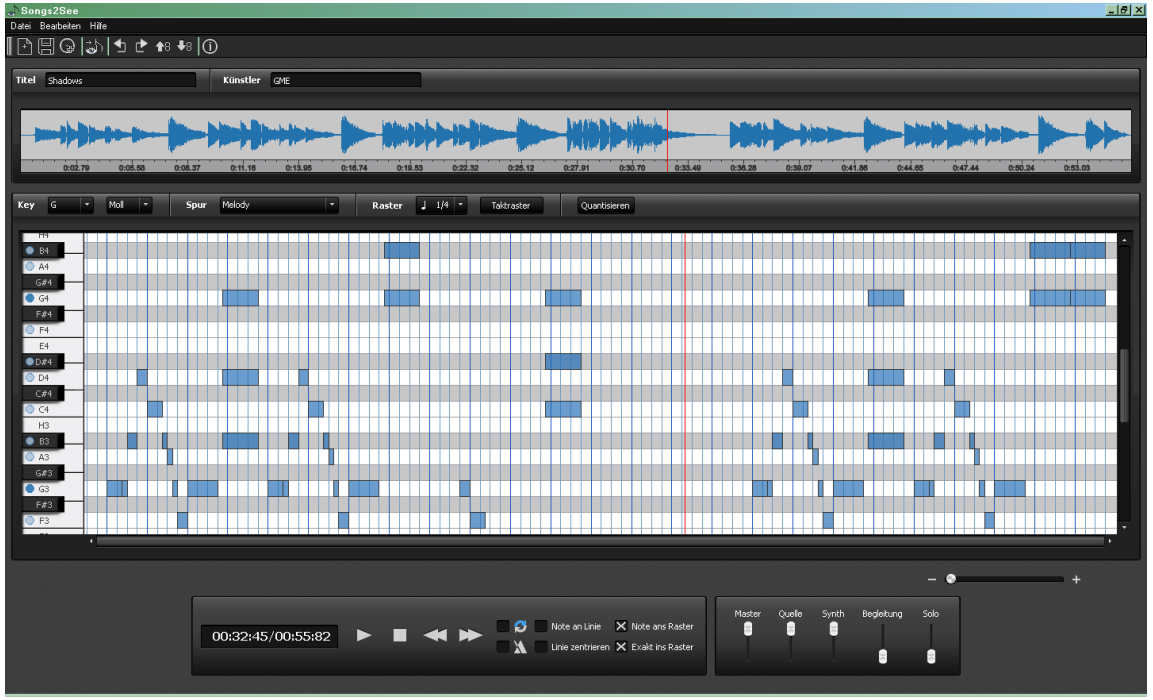
Automatic Music Transcription

Songs2See

Music learning software based on automatic music transcription

Songs2See Editor

- Song2See game content creation
- Determine score from audio input
- Automated transcription of melody and bass, recognition of key, tone gender, meter, chords
- Source separation for solo and backing track
- Also available as library



Fraunhofer IDMT

Environmental Sound Analysis

Environmental Sound Analysis

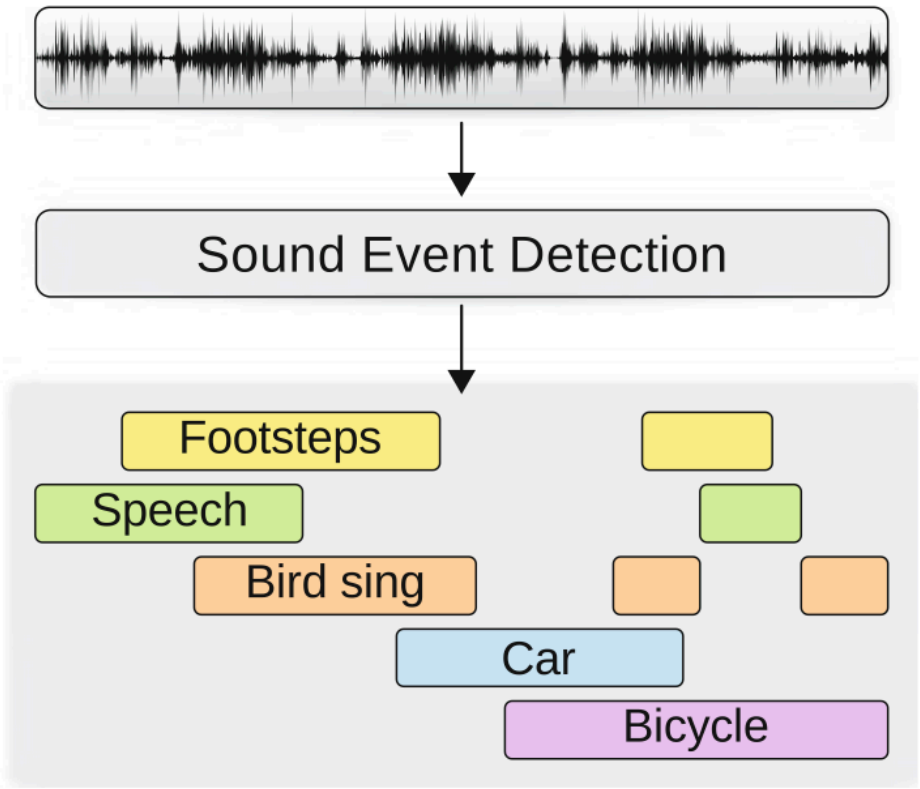
Sound Event Detection and Acoustic Scene Classification

Goals

- Locate and classify individual sounds in complex soundscapes
- Classify environment (park, traffic, etc.)
- Estimate number of simultaneous sounds (polyphony)

Challenges

- High variability and complexity of environmental sounds
- Overlap in time and frequency



Virtanen, T., Plumbley, M. D., & Ellis, D. (Eds.). (2018). *Computational Analysis of Sound Scenes and Events*. Cham, Switzerland: Springer International Publishing, p. 15, Fig. 2.1

Fraunhofer IDMT

Sound Monitoring

Use Cases

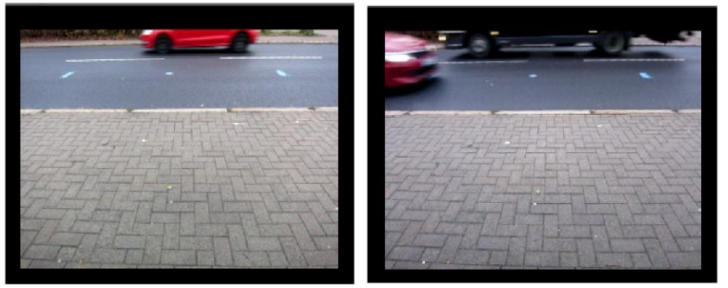
Urban Noise Monitoring

- Distributed noise level measurement using mobile sensors
- Classify prominent noise sources
- Compress deep-learning based sound analysis algorithms

»StadtLärm« sensor units

Traffic Monitoring

- Detect and classify passing vehicles (cars, bus, truck, motorcycles)
- Estimate speed and direction of movement

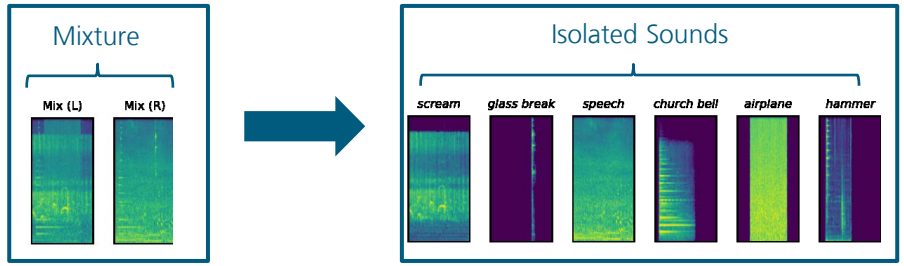


Passing vehicles

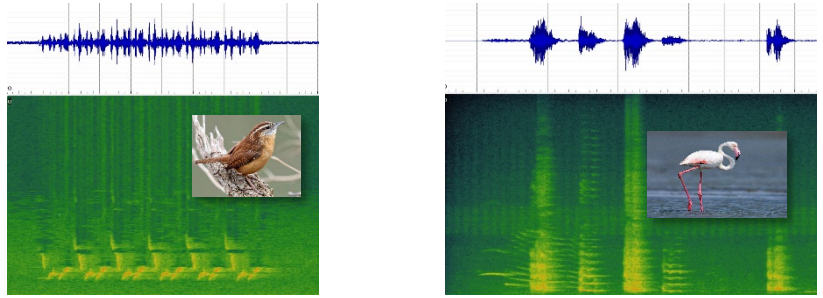
Environmental Sound Analysis

Current Research

- Polyphonic sound separation and detection
- Domain shift / domain adaptation
- Bioacoustics Monitoring
- Model Interpretability



Polyphonic soundscape from »USM-SED« dataset



Vocalizations of different bird species

Fraunhofer IDMT

Semantic Music Technologies

More information

- Audio Matching
  - Query-based Matching
- Music and Speech Detection
- Music Recommendation
  - SoundsLike, Jamahook, AI4media
- Automatic Music Transcription
  - Songs2See, Skoove, Melody Extraction for Music Games
- Environmental Sound Analysis
  - The "StadtLärm" (CityNoise) Project

- Business Unit Media Management and Delivery at Fraunhofer IDMT

Semantic Music Technologies
Publications

Music and Speech Detection

- Grollmisch, S.; Cano, E.; Kehling, C.; Taenzer, M., "Analyzing the Potential of Pre-Trained Embeddings for Audio Classification Tasks," 2020 28th European Signal Processing Conference (EUSIPCO), 2021, pp. 790-794, doi: 10.23919/Eusipco47968.2020.9287743.
- Draghici, Alexandra; Abeßer, Jakob; Lukashevich, Hanna, „A study on spoken language identification using deep neural networks“, In Proceedings of the 15th International Conference on Audio Mostly (AM '20). Association for Computing Machinery, New York, NY, USA, 253–256. DOI:https://doi.org/10.1145/3411109.3411123

Music Recommendation

- Ribecky, Sebastian; Abeßer, Jakob; Lukashevich, Hanna, "Multi-input architecture and disentangled representation learning for multi-dimensional modeling of music similarity," in Proceedings of the Audio Engineering Society (AES) Convention 151, 2022
- Grollmisch, S. & Lukashevich, H., (2017). Soundslike. In: Eibl, M. & Gaedke, M. (Hrsg.), INFORMATIK 2017. Gesellschaft für Informatik, Bonn. (S. 139-150). DOI: 10.18420/in2017

Semantic Music Technologies
Publications

Automatic Music Transcription

- Taenzer, Michael; Mimilakis, Stylianos I.; Abeßer, Jakob, "Informing Piano Multi-Pitch Estimation with Inferred Local Polyphony Based on Convolutional Neural Networks" Electronics 10, no. 7: 851, 2021, https://doi.org/10.3390/electronics10070851
- Abeßer, Jakob; Müller, Meinard, "Jazz Bass Transcription Using a U-Net Architecture" Electronics 10, no. 6: 670, 2021, https://doi.org/10.3390/electronics10060670
- Nowakowski, M.; Weiß, C.; Abeßer, J., „Towards Deep Learning Strategies for Transcribing Electroacoustic Music", In: Kronland-Martinet R., Ystad S., Aramaki M. (eds) Perception, Representations, Image, Sound, Music. CMMR 2019. Lecture Notes in Computer Science, vol 12631. Springer, Cham. https://doi.org/10.1007/978-3-030-70210-6_3
- Abeßer, J.; Müller, M., "Fundamental Frequency Contour Classification: A Comparison between Hand-crafted and CNN-based Features," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 486-490, doi: 10.1109/ICASSP.2019.8682252.
- Grollmisch, Sascha; Cano, Estefania, "Automatic Chord Recognition in Music Education Applications", 16th Sound and Music Computing Conference (SMC2019), Málaga, Spain. https://doi.org/10.5281/zenodo.3249362
- Nadar, Christon-Ragavan; Abeßer, Jakob; Grollmisch, Sascha, „Towards CNN-based Acoustic Modeling of Seventh Chords for Automatic Chord Recognition," 16th Sound and Music Computing Conference (SMC2019), Málaga, Spain. https://doi.org/10.5281/zenodo.3249472

Semantic Music Technologies

Publications

Environmental Sound Analysis

- Abeßer, J., "Classifying Sounds in Polyphonic Urban Sound Scenes". In Proceedings of the 152nd Audio Engineering Society (AES) Convention, Online, 2022
- Abeßer, J.; Gourishetti, S.; Kátai, A.; Clauß, T.; Sharma, P.; Liebetrau, J., „IDMT-Traffic: An Open Benchmark Dataset for Acoustic Traffic Monitoring Research“, in Proceedings of the European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 2021.
- Johnson, David S.; Lorenz, Wolfgang; Taenzer, Michael; Mimilakis, Stylianos; Grollmisch, Sascha; Abeßer, Jakob; Lukashevich, Hanna, „DESED-FL and URBAN-FL: Federated Learning Datasets for Sound Event Detection“, 2021, arXiv:2102.08833
- Abeßer, Jakob,"A Review of Deep Learning Based Methods for Acoustic Scene Classification" Applied Sciences 10, no. 6: 2020. https://doi.org/10.3390/app10062020
- Grollmisch, S.; Abeßer, J.; Liebetrau, J.; Lukashevich, H. "Sounding Industry: Challenges and Datasets for Industrial Sound Analysis," 2019 27th European Signal Processing Conference (EUSIPCO), 2019, pp. 1-5, doi: 10.23919/EUSIPCO.2019.8902941.
- Abeßer, J.; Götze, M.; Kühnlenz, S.; Gräfe, R.; Kühn, C.; Clauß, T.; Lukashevich, H., „A Distributed Sensor Network for Monitoring Noise Level and Noise Sources in Urban Environments,“ in Proceedings of the IEEE 6th International Conference on Future Internet of Things and Cloud, Barcelona, Spain, 2018.

Contact

Hanna Lukashevich
Head of Semantic Music Technologies

Phone +49 3677 467-224
hanna.lukashevich@idmt.fraunhofer.de

Fraunhofer Institute for Digital Media Technology
www.idmt.fraunhofer.de/m2d