Task-Oriented Evaluation for Automated Music Simplification

Universität Bamberg

Andrew McLeod, Hanna Lukashevich, Jakob Abeßer

(Contact: andrew.mcleod@idmt.fraunhofer.de)

Fraunhofer IDMT

Motivation

- No standard evaluation exists for music simplification
- Use cases differ widely
- Evaluation metrics/models suited to one use case often fail on another

Use-Cases

Skill level (US1.1) — Playable simplification preserving "essence" [1]

- Simplify a piece so a lower-skill musician can still play something that captures its essence
- Related to difficulty/playability estimation literature
- Musician learns only the simplified version, never the original

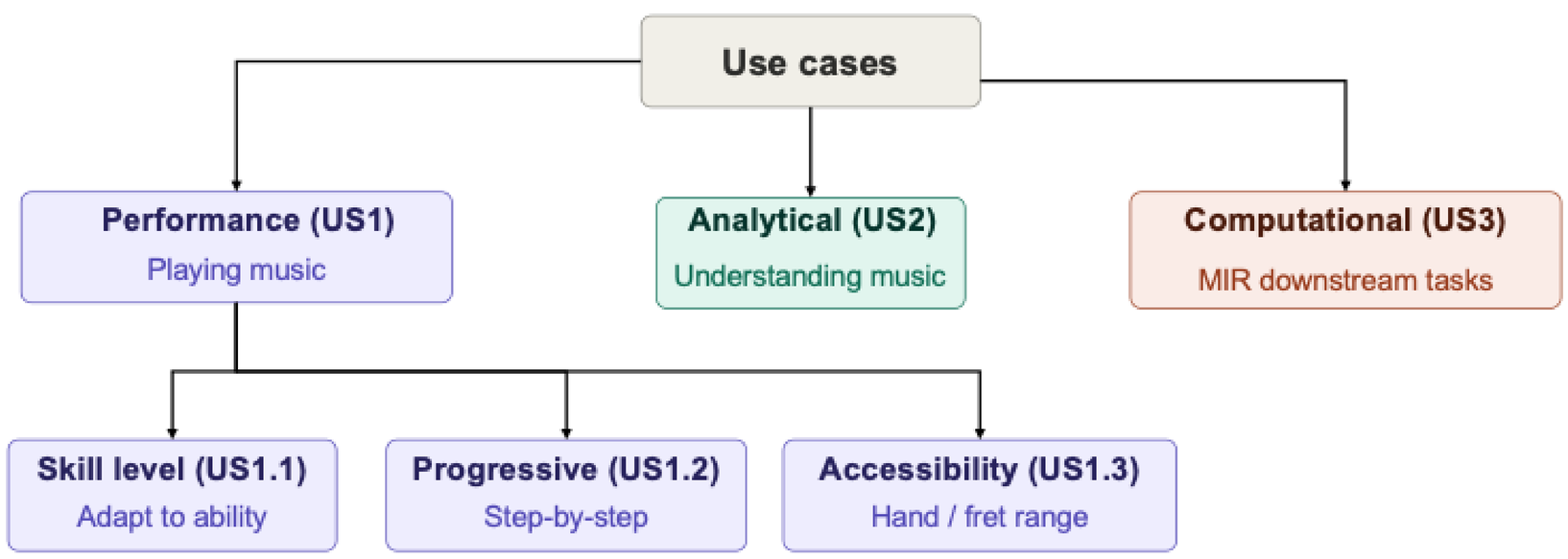
Progressive (US1.2) — Learning scaffold toward the full piece [2]

- Simplified version acts as a stepping stone toward eventually learning the full piece
- Key challenge: simulating/measuring the learning process and its duration
- Musician first learns the simplified version, then the complex original

Accessibility (US1.3) — Motor, cognitive, or sensory constraints [3]

- Reduce polyphony, hand span, tempo, or rhythmic density for users with physical/cognitive limits
- Applies instrument-specific constraints (e.g. one-handed piano, limited fret range on guitar)
- Evaluated via constraint-satisfaction metrics, task success rate, and workload questionnaires

Use-case hierarchy for music simplification



Analytical (US2) — Harmonic analysis & accompaniment flexibility [4]

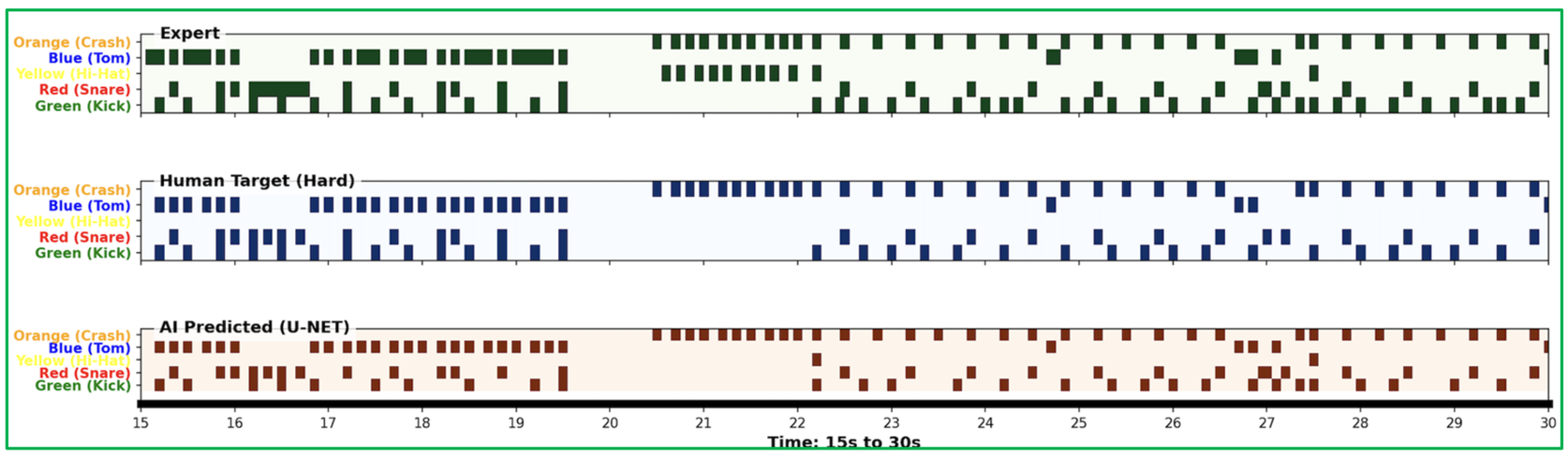
- Simplify to expose harmonic structure, e.g., same progression with varied rhythmic patterns
- Evaluated by running MIR algorithms (chord/scale estimation) with vs. without MS

Computational (US3) — Symbolic music retrieval (query-by-humming) [5]

- Simplify a complex piece before computing a distance measure against a hummed query
- MS = pre-processing step for retrieval
- Evaluated by comparing QbH performance with vs. without simplification

Case Study (US1.1): Instrumental Video Game (Clone Hero)

- **Expert**: Synthesized MIDI drum track of the full-difficulty transcribed drums, given as input to the model. This was created manually, by a user playing the video game
- **Human Target (Hard)**: Slightly easier version of the drum track, created manually by the same user.
- **AI Predicted (U-Net)**: Simplified version output by a U-Net model



[1] Ramoneda et al., "Difficulty-aware score generation for piano sight-reading," Expert Systems with Applications, 2026.
[2] Ramoneda et al., "Difficulty-controlled simplification of piano scores with synthetic data for inclusive music education," arXiv:2511.16228, 2025.
[3] Nakamura & Yoshii, "Statistical piano reduction controlling performance difficulty," APSIPA Transactions on Signal and Information Processing, 7, e13, 2018.
[4] Ericson et al., "MuseReduce: A generic framework for hierarchical music analysis," Music Encoding Conference, 2022.
[5] Kosugi et al., "A practical query-by-humming system for a large music database," ACM Multimedia, 2000.