

MICROTASK CROWDSOURCING FOR MUSIC SCORE TRANSCRIPTIONS: AN EXPERIMENT WITH ERROR DETECTION



Andrea Mauri Cynthia Liem Christoph Lofi Ioannis Petros Samiotis Sihang Qiu Alessandro Bozzon {i.p.samiotis, s.qiu-1, a.mauri, c.c.s.liem, c.lofi, a.bozzon}@tudelft.nl

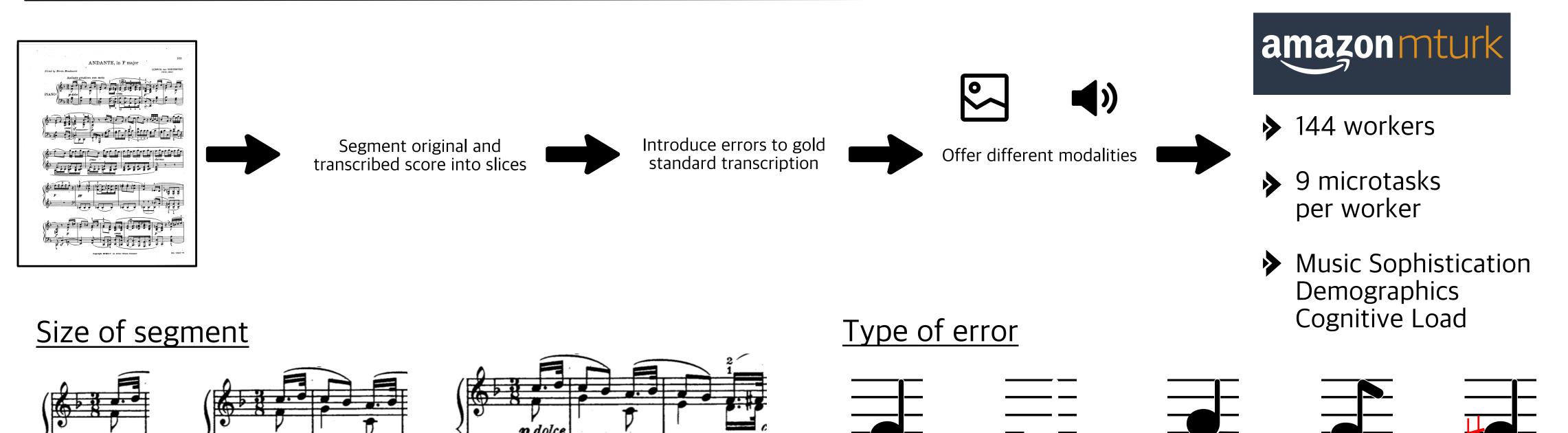
{?} Problem Description

- The majority of transcriptions for professional use involve experts using specialised interfaces.
- State-of-the-art Optical Music Recognition methods show acceptable performance in the case of clean music scores but they still require substantial human intervention to provide results with consistent quality.
- Microtask crowdsourcing is a popular approach for scaling up digital content annotation tasks.



To what extent are workers from microtask crowdsourcing platforms able to detect errors in transcribed music scores?

ᇑ Experimental Design



Modality of segment

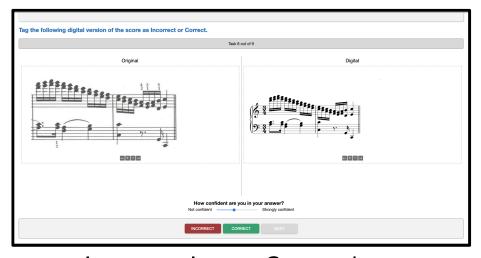
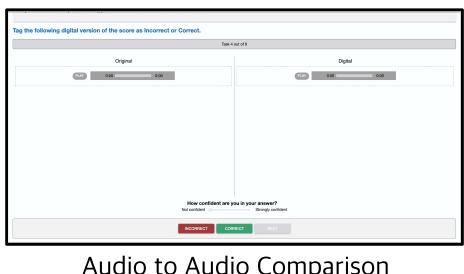
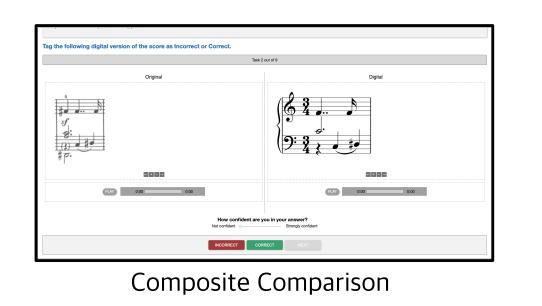


Image to Image Comparison



Correct Example

Audio to Audio Comparison



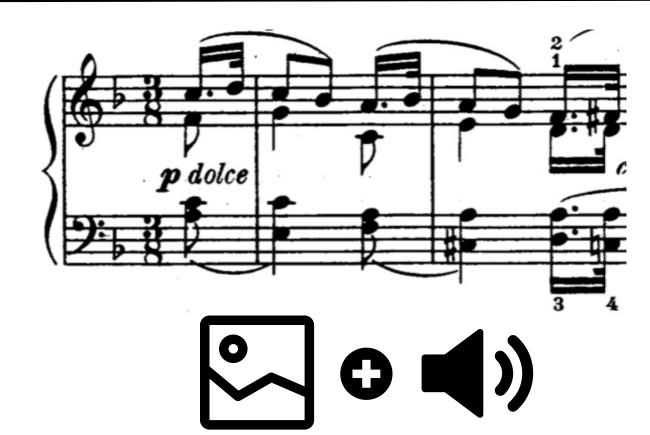
Wrong Vertica

Position

Missing Note

Results

Best Segment Size



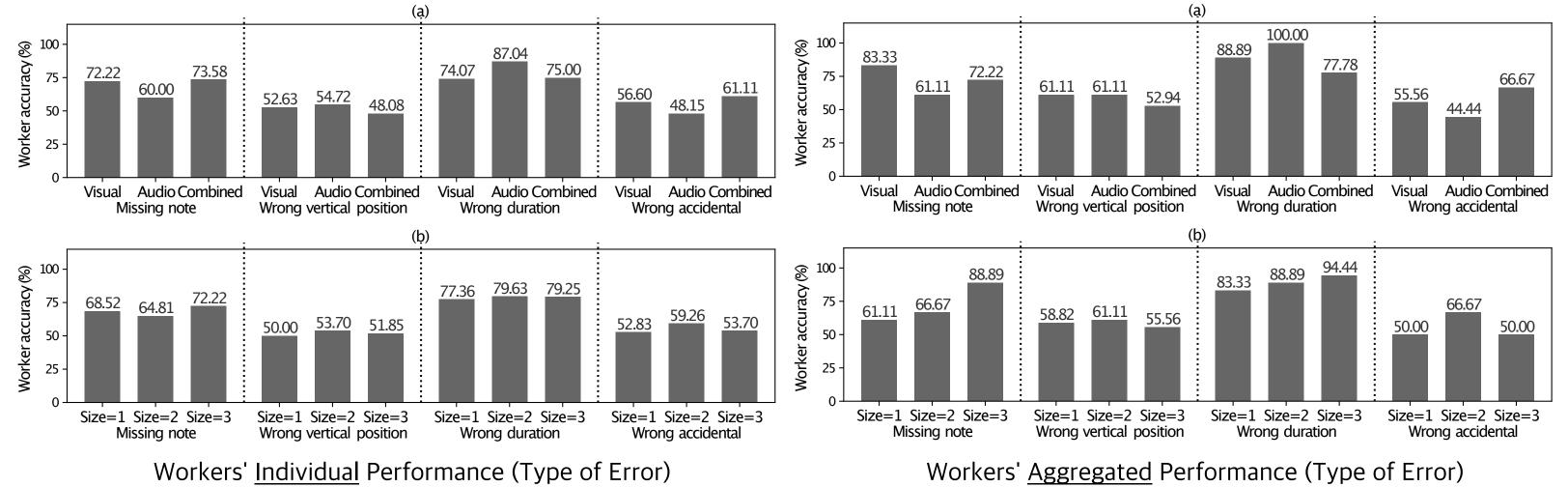
Best Interface

Individual

Precision: 68.06% Accuracy: 76.92%

Aggregated

Precision: 94.74% Recall: 75.00% **Accuracy: 85.42%**



Conclusion

- Microtask crowdsourcing can be used to scale up specific transcription activities
- Worker interfaces that combine visual and audio modalities allow the evaluation of longer score segments
- Focusing on the error detection task, results show that crowd workers can achieve high precision and recall, especially with Missing Note and Wrong Duration errors



Wrong Duration Wrong Accidental

This project has received funding from the European Union's Horizon 2020 research and innovation programme H2020-EU.3.6.3.1. - Study European heritage, memory, identity, integration and cultural interaction and translation, including its representations in cultural and scientific collections, archives and museums, to better inform and understand the present by richer