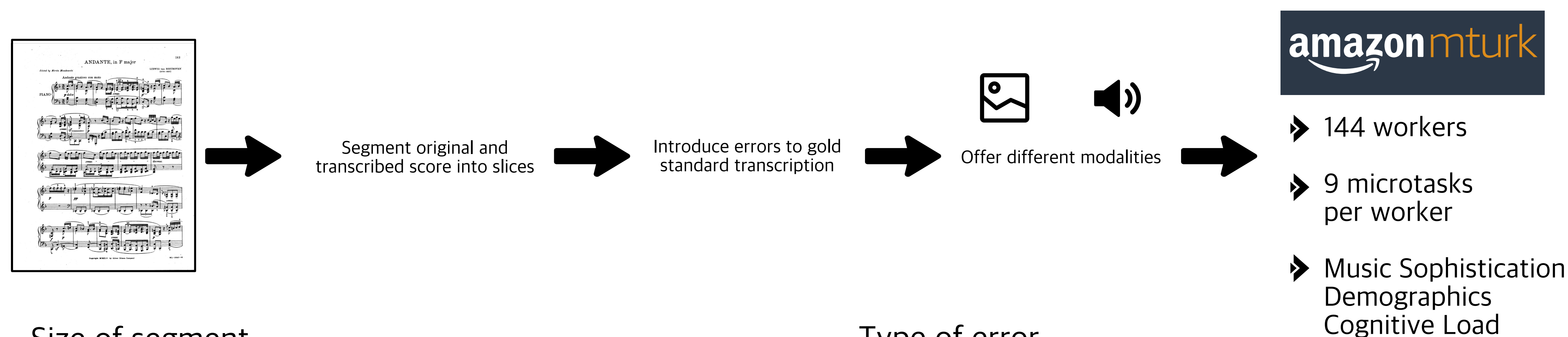


Problem Description

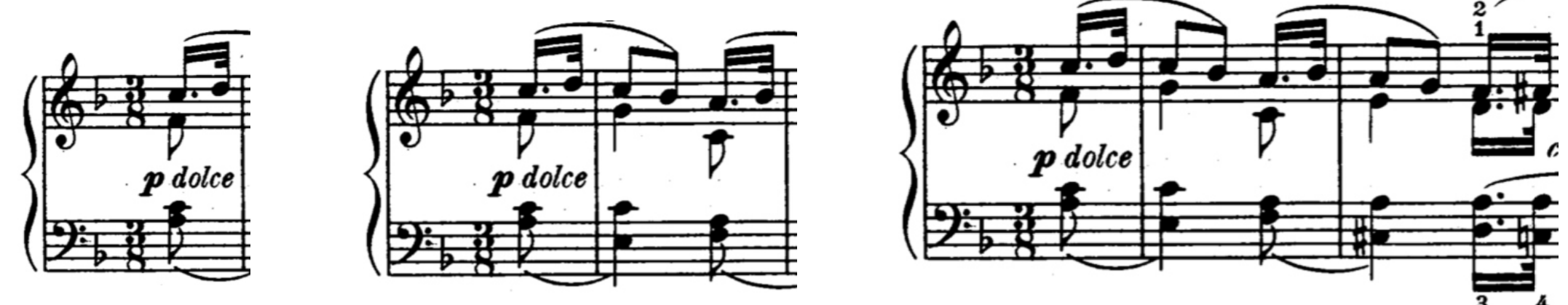
- The majority of transcriptions for professional use involve experts using specialised interfaces.
 - State-of-the-art Optical Music Recognition methods show acceptable performance in the case of clean music scores but they still require substantial human intervention to provide results with consistent quality.
 - Microtask crowdsourcing is a popular approach for scaling up digital content annotation tasks.
- ➡ Music scores are complex artefacts that need specific domain knowledge to read and understand, making the task of transcribing a score complex and cognitively demanding

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

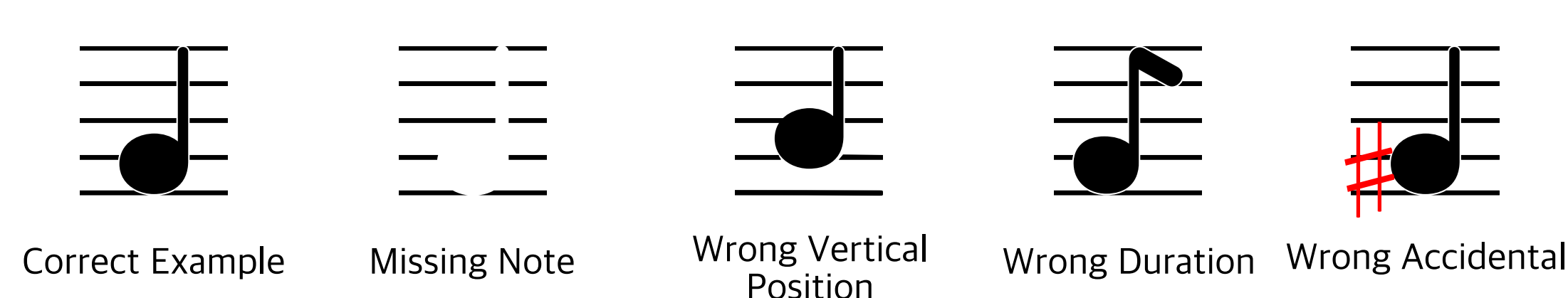
Experimental Design



Size of segment



Type of error



Modality of segment

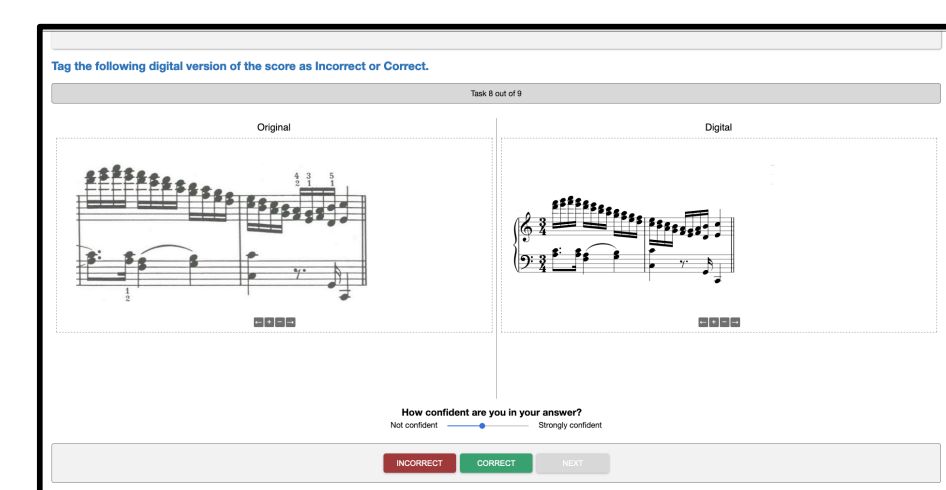
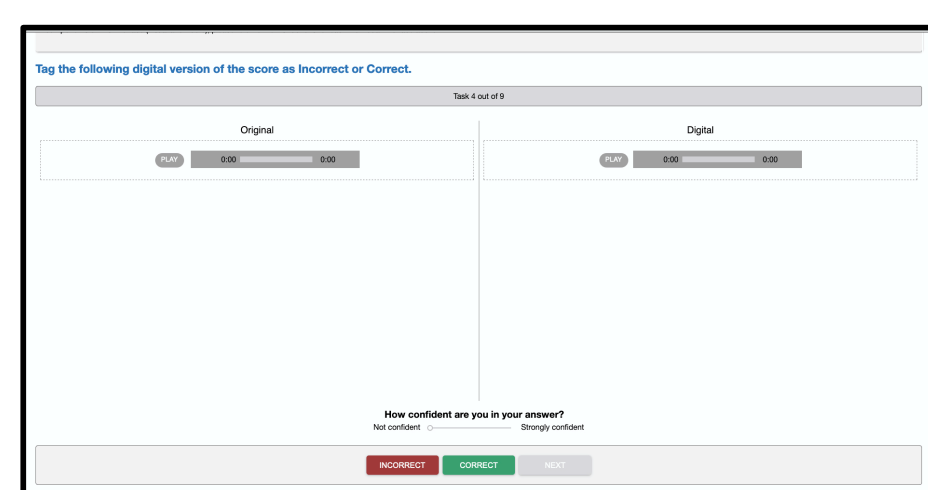
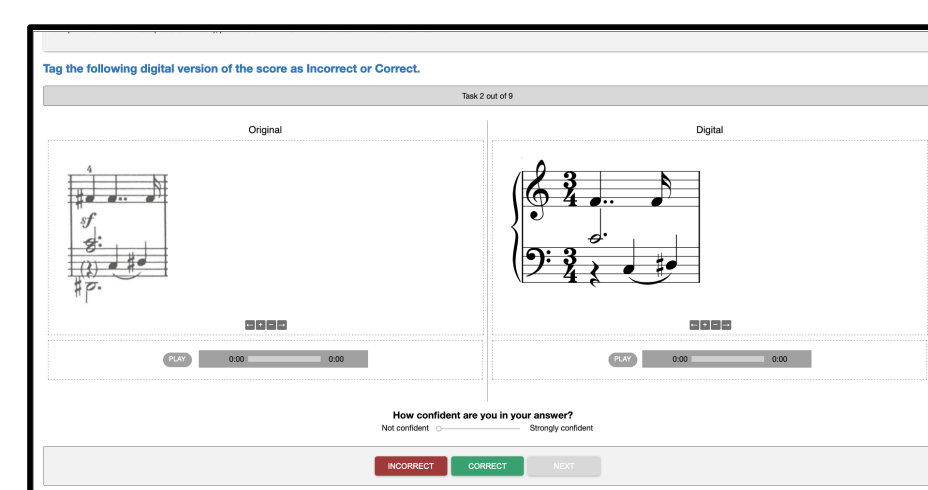


Image to Image Comparison



Audio to Audio Comparison



Composite Comparison

Results

- Best Segment Size
- Best Interface

Modality	Precision	Recall	Accuracy
Individual	68.06%	83.05%	76.92%
Aggregated	94.74%	75.00%	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

