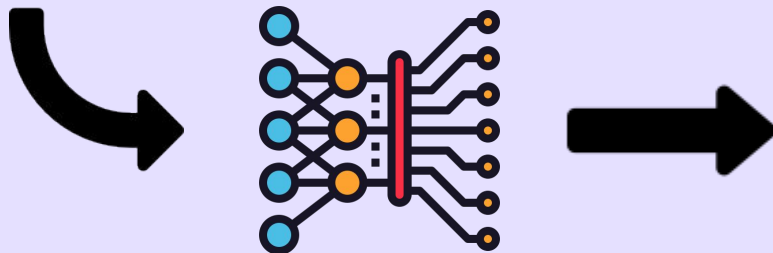


# Automatické rozpoznávání hudebního zápisu pomocí neuronových sítí (Optical music recognition = OMR)

Vedoucí práce: Ing. Michal Hradiš Ph.D.

Vojtěch Vlach (xvlach22)




```
...  
<Note>  
  <durationType>quarter</durationType>  
  <pitch>60</pitch>  
</Note>  
<Note>  
  <durationType>quarter</durationType>  
  <pitch>64</pitch>  
</Note>  
<Note>  
  <durationType>quarter</durationType>  
  <pitch>67</pitch>  
</Note>  
<Rest>  
  <durationType>quarter</durationType>  
</Rest>  
...
```

# Dva základní pojmy

## Monofonní hudba



## Polyfonní hudba

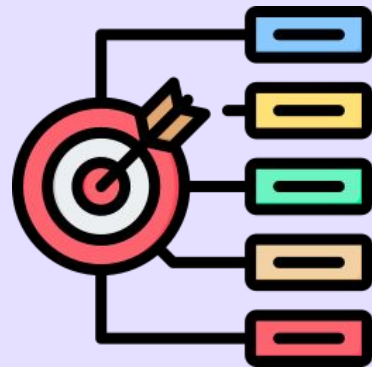
A musical score for a polyphonic piece. It is written on a grand staff (treble and bass clefs) with a key signature of one flat (B-flat). The score shows multiple voices, with some notes beamed together. The piece is marked with a measure number of 64 and a rehearsal mark of 26.

clef-G2 + keySignature-BbM + note-G3\_eighth. note-Bb3\_quarter note-Eb4\_quarter  
note-F4\_quarter + note-F3\_sixteenth + note-G3\_sixteenth note-Eb4\_quarter  
note-F4\_quarter + note-Bb3\_eighth + note-G3\_sixteenth + note-G3\_sixteenth note-Bb3\_eighth  
note-Eb4\_eighth note-F4\_eighth + note-G3\_eighth. + note-F4\_sixteenth + note-G4\_sixteenth +  
note-C4\_eighth note-G4\_sixteenth + note-A4\_eighth. + note-C4\_eighth + barline

[Edirisooriya, H.W. Dong, J. McAuley, and T. Berg-Kirkpatrick. An empirical evaluation of end-to-end polyphonic optical music recognition. In International Society for Music Information Retrieval, 2021](#)

# Cíle práce a prostředky ke splnění

- Nový přístup k OMR pomocí **img2seq transformer** sítě na **tištěnou monofonní** hudbu
- Nová reprezentace **polyfonní** hudby jako sekvence
- Otestovat a vylepšit transformer síť pro **polyfonní** hudbu
- Veřejné datasety
  - PrIMuS<sup>[1]</sup> – monofonní hudba
  - MuseScore<sup>[2]</sup> – polyfonní hudba
- Hodnocení pomocí Symbol error rate<sup>[3]</sup>
  - Levenshteinova vzdálenost (jako Word error rate v NLP)



[1] [J. Calvo-Zaragoza and D. Rizo, "End-to-end neural optical music recognition of monophonic scores," Applied Sciences, vol. 8, no. 4, 2018](#)

[2] <https://github.com/Xmader/musescore-dataset>

[3] [Baró, Arnau & Riba, Pau & Calvo-Zaragoza, Jorge & Fornés, Alicia. \(2019\). From Optical Music Recognition to Handwritten Music Recognition: a Baseline. Pattern Recognition Letters. 123. 10.1016/j.patrec.2019.02.029.](#)

# Vytvořená baseline pro monofonní hudbu

- Framework PERO<sup>[4]</sup>, síť VGG LSTM
- Dataset PrIMuS<sup>[1]</sup>
  - 80 000 obrázků a k nim přepisy s dvěma typy kódování
  - **Semantic**: méně symbolů, vystihují smysl symbolů (např. nota C4 v houslovém klíči)
  - **Agnostic**: více symbolů, popis grafických primitiv (např. nota na 1. pomocné lince)



```
clef-G2      timeSignature-3/4      note-C4_quarter  note-C4_half  
clef.G-L2    digit.3-L4    digit.4-L2    note.quarter-L0  note.half-L0
```

- Minimální dosažená chybovost baseline:
  - **Semantic**: 7 % (SER)
  - **Agnostic**: 9.71 % (SER)

[4] <https://pero.fit.vutbr.cz/about>

# Plán další práce

- Vlastní architektura **img2seq transformer** sítě
- Vylepšení výsledků oproti baseline
- Pomocí datasetu MuseScore vytvořit nový polyfonní zápis
- Dataset MuseScore<sup>[2]</sup> obsahuje:
  - víc než 20 000 stránek notových zápisů, ~100 000 řádků

