## SYSTEM FOR AUTOMATIC RECOGNITION **OF CZECH DIALECTS** Author: Bc. Ondřej Odehnal

**ABSTRACT:** Using large self-supervised pre-trained neural models for dialect identification from raw audio with aid of language identification datasets. Part of ongoing NAKI program focused on national and cultural identity heritage.

Supervisor: Ing. Oldrich Plchat, Ph.D.



Langueges dataset – VoxLingua107

- Large dataset for training LID models.
- Short recordings extracted from YouTube.
- 107 world languages.

- Interviewers & respondents conversations.
- Dates back to 1960. Various quality.



## **DATA CLEANSING**

- Overlapping speech detection.
- VAD & speaker diarization.
- Interviewer detection heuristic: "The least speaking person."



O C

Model Training	Loss	Test Accuracy NAKI (%)
20 epochs VoxLingua107 + 20 epochs NAKI	AS1	76.94
	Softmax	77.12
100 epochs NAKI	Softmax	76.40

<sup>1</sup>AS: Angular Softmax Loss

Fiq 5: Keys and values of MHFA frontend per WavLM Layer for AS & Softmax loss.





Fig 6: Accuracy heatmap for each dialect subgroup evaluted on Test part of the NAKI dataset for WavLM trained on 20 epochs of VoxLingua107 & 20 epochs of NAKI with AS loss.

ACKNOWLEDGEMENTS: I would like to thank my supervisor, Ing. Oldřich Plchot, Ph.D., for his invaluable guidance and support throughout this work.

