

Title

AI Models for Automated Speech Recognition from Phone Calls

Supervision

Hugo Pedro Proença (UBI-DI)

Description

Automated speech recognition (ASR) has progressed rapidly due to self-supervised speech representations, large-scale weak supervision and sequence-to-sequence neural architectures. However, telephone calls remain challenging because of narrow-band audio, background noise, speaker variability, accents, interruptions, overlapping speech and privacy-sensitive content.

The main goal of this dissertation is to study, implement and compare AI-based ASR models for the automatic transcription of phone calls. The work should evaluate existing models and adaptation strategies, with special attention to robustness, domain adaptation, punctuation/restoration, diarisation cues and the reliability of transcripts when the audio quality is poor.



Figure 1: Automated speech recognition from phone calls.

Hence, the main objectives of this Master dissertation are:

- Review state-of-the-art ASR models and datasets relevant to telephone speech.
- Define an experimental protocol with suitable metrics, such as word error rate, character error rate and robustness under degraded audio conditions.
- Implement and evaluate baseline systems based on representative modern ASR architectures.
- Study adaptation strategies for phone-call audio, including fine-tuning, data augmentation and confidence estimation.
- Publish the obtained results in an international journal or conference.

Workplan



Proposal of Masters Dissertation

- 1- Comparative study of the methods previously published in ASR, self-supervised speech learning and telephone-speech transcription.
- 2- Collection/preparation of representative public or institutional audio data, respecting privacy and ethical requirements.
- 3- Development, implementation and test of selected ASR approaches and adaptation strategies.
- 4- Quantitative and qualitative evaluation of transcription accuracy, robustness and practical limitations.
- 5- Preparation of the dissertation and submission of a scientific publication.

Chronogram

Task	M1-M3	M4-M6	M7-M9	M10-M12
Literature review and problem formalisation	X	X		
Dataset preparation, baselines and metrics		X	X	
Model development, evaluation and ablation studies		X	X	X
Dissertation writing and scientific publication			X	X

Prerequisites

- Basic skills in English writing/reading.
- Good programming skills, preferably in Python.
- Previous knowledge about machine learning, artificial intelligence and deep learning.
- Interest in speech processing and responsible handling of sensitive data.

References

- [1] Radford, A. et al., Robust Speech Recognition via Large-Scale Weak Supervision, International Conference on Machine Learning (ICML), 2023.
- [2] Baevski, A., Zhou, Y., Mohamed, A. and Auli, M., wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations, NeurIPS, 2020.
- [3] Gulati, A. et al., Conformer: Convolution-augmented Transformer for Speech Recognition, Interspeech, 2020.