The goal of this report is to document my work on developing and testing a speech recognition system for an airport virtual assistant project. The system is written in Python, and the primary voice recognition capabilities are provided by the Mozilla DeepSpeech software package. The project attempts to handle a variety of issues, such as language selection, dealing with loud surroundings, and assuring robust performance.

Software Used:

- Python: Version 3.8
- Mozilla DeepSpeech: Version 0.9.3
- TensorFlow: Version 2.3.0-6-g23ad988fcd

Language models were configured based on the provided models in the Ex4_models.pdf document. The system supports English, Italian, and Spanish. To mitigate the impact of noisy environments, the system has been configured with audio filters. Gain and amplification adjustments are applied to enhance the signal. Additionally, low-pass filters are employed to improve the error rate in challenging acoustic conditions. In specific, VAD parameters have been fine-tuned to improve the system's ability to handle noisy environments. Adjustments to aggressiveness, frame duration, and padding duration have been made for optimal performance. These configurations collectively enhance the robustness of the speech recognition system in real-world, noisy scenarios. The speech recognition system has been evaluated using a set of audio files provided by the client, along with custom recordings ('your_sentence1.wav' and 'your_sentence2.wav'). The results are presented in the following tables, providing insights into the Word Error Rate (WER), language, inference time, and filename.

ENGLISH EVALUATION

Filename	Language	Inference Time(s)	WER (%)
checkin.wav	e	0.93s	40.00%
parents.wav	n	1.68s	40.00%
passport.wav	g	2.77s	11.11%
suitcase.wav	l	3.71s	33.33%
what_time.wav	i	4.28s	20.00%
where.wav	s	5.17s	16.67%
work.wav	h	5.92s	16.67%

ITALIAN EVALUATION

Filename	Language	Inference Time(s)	WER (%)
checkin_it.wav	i	0.53s	100.00%
parents_it.wav	t	0.97s	60.00%
suitcase_it.wav	a	1.61s	42.86%
what_time_it.wav	l	1.98s	85.71%
where_it.wav	i	2.61s	42.86%

SPANISH EVALUATION

Filename	Language	Inference Time(s)	WER (%)
checkin_es.wav	s	0.73s	100.00%
parents_es.wav	p	1.40s	40.00%
suitcase_es.wav	a	2.20s	50.00%
what_time_es.wav	n	2.78s	100.00%
where_es.wav	i	3.79s	71.43%

The implemented speech recognition system demonstrates effective language support, robustness in noisy environments, and reasonable inference times. The custom configurations, including gain and low-pass filters, significantly contribute to the system's adaptability to challenging acoustic conditions. However, there are instances, particularly in the Spanish language evaluations, where the Word Error Rate is high, indicating areas for potential improvement. Further optimization and model fine-tuning may enhance the overall performance of the system. The inclusion of custom sentences in the evaluation ensures a holistic assessment of the system's practical utility.