

# Umberto Cappellazzo

Website:  
[umbertocappellazzo.github.io/](https://umbertocappellazzo.github.io/)  
Email:  
[umbertocappellazzo@gmail.com](mailto:umbertocappellazzo@gmail.com)  
LinkedIn:  
[umberto-cappellazzo-116093150](https://www.linkedin.com/in/umberto-cappellazzo-116093150)  
GitHub:  
[github.com/umbertocappellazzo](https://github.com/umbertocappellazzo)  
X (formerly Twitter):  
[twitter.com/Umberto\\_Senpai](https://twitter.com/Umberto_Senpai)

## EDUCATION

---

### University of Trento

Ph.D. in Information Engineering and Computer Science [with honours]

Trento, Italy

Nov. 2021–15/01/2025

- Title: *Efficient Knowledge Transfer and Adaptation for Speech and Beyond*
- Advisors: Daniele Falavigna, Alessio Brutti
- Research interests: continual learning for audio and speech processing; multi-modal (i.e., audio-language) continual learning; parameter-efficient transfer learning of audio/speech (e.g., Adapters, Mixture of Adapters, LoRA); multi-modal LLMs for audio-visual speech recognition. This led to 9 publications in top-notch conferences.

### University of Padua

Padua, Italy

MSc in Telecommunication Engineering [110/110 with honours]

2016–2019

- Advisors: Michele Rossi, Matteo Gadaleta
- Thesis Title: A Deep Learning-Based ECG Delineator: Evaluation and Comparison on Standard Databases

### University of Padua

Padua, Italy

BSc in Information Engineering

2013–2016

- Advisor: Nicola Laurenti
- Thesis Title: Message Authentication over an Ideal or Noisy Channel

## WORK EXPERIENCE

---

### Imperial College London

London, UK

Research Associate in the iBUG group (leader: Maja Pantic)

03/2025-ongoing

- My research focuses on advancing audio-visual speech recognition (AVSR) through Multimodal Large Language Models in a parameter-efficient way. [Llama-AVSR](#), presented at ICASSP 2025, represents the very first attempt to unify AVSR and LLMs.
- Proposed: • [Llama-AVSR](#) [ICASSP 2025] • [Llama-SMoP](#) [Interspeech 2025] • [Llama-Matryoshka](#) [IEEE ASRU 2025] • [MoME: Mixture of Matryoshka Experts](#) [NeurIPS 2025] • [Omni-AVSR](#) [submitted to ICASSP 2026] • [Attention Sinks and Massive Activations in AVSR with LLMs](#) [submitted to ICASSP 2026].
- Supervisor: Stavros Petridis (ICL/NatWest AI Research)

### Imperial College London

London, UK

Research Intern, *Audio-visual speech recognition meets LLMs*

February 2024 –November 2024

Supervisor: Stavros Petridis (ICL/Meta AI)

- I investigated the efficient integration of LLMs for the task of audio-visual speech recognition. This culminated in [Llama-AVSR](#), a multimodal LLM with strong audio-visual speech recognition abilities. This work has been accepted at *ICASSP 2025*. More details [here](#).

- Junior researcher for the “*Finite state methods with modern neural Architectures for speech applications and beyond*” group at **JSALT2023** in Le Mans, France. I worked on the integration of early-exit techniques to make the training and inference of CTC/MMI systems dynamical. More information available. This led to a [publication](#) at ICASSPW 2023.

## SKILLS

- **Programming Languages:** Python (advanced), Java (basic), HTML (basic), Matlab (basic)
- **ML/DL Toolkits/Libraries:** PyTorch, HuggingFace Transformers, Pytorch Lightning, NumPy, Matplotlib, Scikit-Learn
- **Distributed Systems:** Hands-on experience with large scale training of models using distributed systems
- **Tools & Platforms:** Git, Docker

## LANGUAGES

- **Italian:** Mother tongue
- **English:** Professional working proficiency

## TALKS & PRESENTATIONS

- “Parameter-Efficient Fine-tuning for Audio and Speech Processing.” *Invited talk at the CUED Speech Group Seminars at the University of Cambridge* (April 2024).

## PROFESSIONAL SERVICES & MENTORSHIP

### Reviewer

- Conferences: ICLR, ACMMM, ICASSP, Interspeech, IJCNN, IEEE MLSP.
- Journals: IEEE Signal Processing Letters, Neurocomputing, International Journal of Computer Vision, Transactions on Image Processing, Knowledge-based Systems.

### Mentorship

- Anand [University of British Columbia, MSc student][summer 2025]: *attention sinks and massive activation in audio-visual LLMs*.
- Lidia Prokopovych [MIT, B.A. student][summer 2025]: *robustness evaluation of Llama-AVSR under different acoustic and visual noise*.
- Stefano Ciapponi [University of Bologna, MSc student][summer 2023]: *prompting techniques to parameter-efficiently fine-tune neural models for speech processing*.

## PUBLICATIONS

- [1] Anand, U. **Cappellazzo**, S. Petridis, and M. Pantic, “[Mitigating Attention Sinks and Massive Activations in Audio-Visual Speech Recognition with LLMs](#)”, *Under review*, 2025.
- [2] U. **Cappellazzo**, M. Kim, H. Chen, P. Ma, S. Petridis, D. Falavigna, A. Brutti, and M. Pantic, “[Large Language Models are Strong Audio-Visual Speech Recognition Learners](#)”, *ICASSP*, 2025.
- [3] U. **Cappellazzo**, M. Kim, P. Ma, H. Chen, X. Liu, S. Petridis, and M. Pantic, “[MoME: Mixture of Matryoshka Experts for Audio-Visual Speech Recognition](#)”, *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.
- [4] U. **Cappellazzo**, M. Kim, and S. Petridis, “[Adaptive Audio-Visual Speech Recognition via Matryoshka-Based Multimodal LLMs](#)”, *IEEE ASRU*, 2025.

- [5] U. **Cappellazzo**, M. Kim, S. Petridis, D. Falavigna, and A. Brutti, “Scaling and Enhancing LLM-based AVSR: A Sparse Mixture of Projectors Approach”, *Interspeech*, 2025.
- [6] U. **Cappellazzo**, X. Liu, P. Ma, S. Petridis, and M. Pantic, “Omni-AVSR: Towards Unified Multimodal Speech Recognition with Large Language Models”, *Under review*, 2025.
- [7] U. **Cappellazzo**, D. Falavigna, and A. Brutti, “Efficient Fine-tuning of Audio Spectrogram Transformers via Soft Mixture of Adapters”, *Interspeech*, 2024.
- [8] U. **Cappellazzo**, D. Falavigna, A. Brutti, and M. Ravanelli, “Parameter-Efficient Transfer Learning of Audio Spectrogram Transformers”, *IEEE MLSP Workshop*, 2024.
- [9] U. **Cappellazzo**, E. Fini, M. Yang, D. Falavigna, A. Brutti, and B. Raj, “Continual Contrastive Spoken Language Understanding”, *ACL Findings*, 2024.
- [10] G. A. Wright, U. **Cappellazzo**, S. Zaiem, D. Raj, L. Ondel Yang, D. Falavigna, and A. Brutti, “Training dynamic models using early exits for automatic speech recognition on resource-constrained devices”, *Self-supervision in Audio, Speech and Beyond (SASB) Workshop, ICASSP*, 2024.
- [11] M. Yang, U. **Cappellazzo**, X. Li, S. Watanabe, and B. Raj, “Improving continual learning of acoustic scene classification via mutual information optimization”, *ICASSP*, 2024.
- [12] M. Yang, X. Li, U. **Cappellazzo**, S. Watanabe, and B. Raj, “Towards Unified Evaluation of Continual Learning in Spoken Language Understanding”, *Interspeech*, 2024.
- [13] U. **Cappellazzo**, D. Falavigna, and A. Brutti, “An Investigation of the Combination of Rehearsal and Knowledge Distillation in Continual Learning for Spoken Language Understanding”, *Interspeech*, 2023.
- [14] U. **Cappellazzo**, M. Yang, D. Falavigna, and A. Brutti, “Sequence-Level Knowledge Distillation for Class-Incremental End-to-End Spoken Language Understanding”, *Interspeech*, 2023.

See [Google Scholar](#) for my Google Scholar profile.