

# **Mirco Ravanelli**

Assistant Professor at Concordia University

Adjunct Professor at Université de Montréal

Mila Associate Member

[mirco.ravanelli@gmail.com](mailto:mirco.ravanelli@gmail.com)

## RESEARCH OVERVIEW

- My research is centered on **deep learning** for **Conversational AI**, with a particular focus on **speech processing**.
- I am currently an **assistant professor** at Concordia University (*Department of Computer Science and Software Engineering*), **adjunct professor** at the Université de Montréal (*Département d'informatique et de recherche opérationnelle*), and an **associate member of the Mila - Quebec AI Institute**.
- I did my postdoc at **Mila (Université de Montréal)** supervised by Prof. **Yoshua Bengio**.
- I'm the creator of the **SpeechBrain toolkit**, which is one of the most used open-source toolkits for speech processing.
- During the last few years, I published more than **100 papers on deep learning and speech processing** topics.
- My research has recently been awarded with the 2022 **Amazon Research Award (ARA)**. Moreover, my publications recently got some **international awards** given by **IEEE** and **IBM**.
- My PhD has been awarded with the "**Best PhD** at the University of Trento" and the "**Best PhD Student Award at Fondazione Bruno Kessler (FBK)**".
- My international network of collaborators currently involves researchers from many institutions around the world including scientists from **North American** (McGill, University of Toronto, Ohio State University), **European** (University of Cambridge, Avignon, Trento, Aalto, Ghent), **Asian** (IIT-Madras, Taiwan National University) **institutions**.
- According to Google Scholar, **my h-index is 31** (with more than **1650 citations in 2024**).

## ASSISTANT PROFESSOR

Concordia University  
From January 2022

I'm a tenure-track Assistant Professor at Concordia University (*Department of Computer Science and Software Engineering*). I'm continuing my research on deep learning for Conversational AI. I'm a member of the Applied AI Institute of Concordia.

**Teaching:** Machine Learning (COMP 6321 and COMP 432), Conversational AI (COMP 499/691), and Applied Deep Learning (COMP

## INTERESTS

Deep Learning

Representation Learning

Self-Supervised Learning

Sequence Modeling

Multimodal LLMs

Interpretable Learning

EEG signal processing

Continual Learning

Conversational AI

Discrete Representations

Open-Source Software



6841). My courses have consistently received outstanding evaluations from students in teaching evaluations.

### ADJUNCT PROFESSOR

Université de Montréal / DIRO  
From January 2022

I'm an Adjunct Professor at the Université de Montréal at Département d'informatique et de recherche opérationnelle (DIRO). I'm currently supervising two PhD students at DIRO.



### MILA ASSOCIATE MEMBER

Université de Montréal / DIRO  
From January 2022

I have been nominated to be a Mila Associate Member. This allows me to keep a strong connection with Mila and continue my fruitful collaboration with this institution.



### POSTDOCTORAL RESEARCHER

Mila - Université de Montréal  
Jan 2018 - Dec 2021  
Supervisor: Prof. Yoshua Bengio

I did post-doc research at the Mila (Université de Montréal). Mila is the world's largest academic research center in deep learning.

### MAIN RESEARCH ACHIEVEMENTS

- I proposed a novel convolutional neural network called **SincNet**, that efficiently processes raw audio samples using insights gained from digital signal processing. SincNet is now a **widely adopted** neural network for processing audio. It has also been adopted to process EEG brain signals.
- I was among the first to study the potential of **self-supervised learning** in audio and speech. In particular, I studied **mutual information**-based approaches and I co-developed a Problem-Agnostic Speech Encoder (PASE) that gave extremely promising results in many speech applications.
- I am the founder and leader of the **SpeechBrain project**, which aims to develop a radically new speech processing toolkit. The project was sponsored by Samsung, Dolby, Nvidia, and Nuance

and involved 30 contributors (Mila interns and external collaborators) from different universities in the world.

- I co-developed the **L-MAC** technique for improving interpretability of audio classifiers (in a paper presented at ICML 2024 in an oral session and at NeurIPS 2024).
- I co-developed the **SepFormer** (Separation Transformer), which is a transformer-based neural architecture that currently achieves state-of-the-art **performance** in speech separation.
- I worked in collaboration with the University of Oxford and the Université d' Avignon to study neural architectures based on **quaternion numbers**.
- I studied novel regularization techniques for recurrent neural networks and contributed to deploy novel solutions for end-to-end spoken language understanding and state-of-the-art speech enhancement systems (e.g, MetricGAN+).

## **RESEARCH FUNDING**

- I'm the recipient of 2025 Imminent Research Grants Award from Translated (C\$ 30k)
- I'm the recipient of the 2025 CRBLM Research Incubator Award (C\$10k)
- I'm the recipient of the 2023 Amazon Research Award (C\$ 100k) to work on empathetic conversational AI.
- I'm recipient of the NSERC Discovery (C\$ 150k) for years 2022-2027 to work on cooperative conversational AI.
- In 2020-2022, I raised C\$ 520k for the SpeechBrain project from important industrial sponsors such as Samsung, Dolby, Nvidia, Nuance, HuggingFace, ViaDialogue, Baidu, OVH.
- I am a co-principal investigator of the ResearchMatch project on "*Facilitation of assessment and intervention for speech and language disorders via automatic speech recognition*" funded by the McGill Initiative in Computational Medicine with C\$75k.
- I contributed to the successful proposal of the EU ESPERANTO (Exchanges for SPEech ReseArch aNd TechnOlogies) project and to make Mila one of the non-European nodes of this initiative. The project is framed within Research and Innovation Staff Exchange (RISE) - Marie Skłodowska-Curie Actions. The project officially started in Jan 2021.

## **TEACHING**

- **2022/2023/2024: Machine Learning**  
(COMP 6321, COMP 432)

I'm teaching machine learning courses for both graduate and undergraduate students. I am also the course coordinator. The course introduces basic concepts such as neural networks, support vector machines, density estimation, gaussian mixture models, unsupervised learning. It provides a strong theoretical and practical view on the files with both frontal lectures and laboratories.



- **2023/2024: Conversational AI**  
(COMP 691, COMP 499)

I'm the teacher and proposer of the Conversational AI course at Concordia. This timely course introduces recent practices used in the field of conversational AI and explains how modern speech assistants and large language models. The students also gain practical experience through lab sessions, tutorials, and projects.

- **2024/2025: Applied Deep Learning**  
(COMP 6841)

I teach the applied deep learning course, emphasizing a hands-on approach. The curriculum includes coding sessions, tutorials, and projects alongside traditional lectures.

- **2021: Advanced Deep Learning Projects**  
(FT6759: Projets avancés en apprentissage automatique)

I taught the “advanced deep learning project” course at the professional master in machine learning organized by Mila and the University of Montreal. This was a mandatory course where students are taught how to work on complex machine learning projects with real data. In 2021, my course had 60 students involved that had to work on 25 real machine learning projects.

<https://admission.umontreal.ca/cours-et-horaires/cours/ift-6759/>  
<https://sites.google.com/view/ift6759/course-description>



## SUPERVISION

- I'm the founder of the Conversational AI Lab (CAIL), which involves students at both Concordia and Université de Montréal. The lab is rapidly growing and I currently involve about 18 students (7 PhD and 4 Master Students, 7 Undergraduate students).
- I **regularly supervise interns** doing research at Mila. In 2020-2022, I supervised about 20 students working in the context of the SpeechBrain project. With many of them, we published papers and we have established a long-term collaboration on research topics on common interest.

- I constantly supervise master students. In 2019-2022, I was the mentor of several students of the professional master organized by Mila. I also followed them during their internship in local AI companies. In particular, I contributed to form new machine learning scientists with both a theoretical and practical background.

## AWARDS

- **Recipient of the 28th Annual Gina Cody School Award for Excellence in Teaching by a Junior Faculty Member**

I received the Excellence in Teaching Award for Junior Faculty for my instruction of the Machine Learning course for undergraduates and the Conversational AI course for both undergraduate and graduate students.

- **2022 Amazon Research Award**

I was one of 79 researchers across 54 universities worldwide who was selected for the prestigious Amazon Research Award (ARS) under the [Amazon Web Services \(AWS\) #AI](#) calls for proposals.

- **2021 IEEE CIS TETCI Outstanding Paper Award**

The IEEE computational intelligence society (CIS) has awarded my 2018 journal paper entitled "*Light Gated Recurrent Units for Speech Recognition*" published on the IEEE Transactions on Emerging Topics in Computational Intelligence with the prestigious 2021 IEEE CIS TETCI Outstanding Paper Award.

- **2024 Italian Ingenuity Award (Special Mention)**

The Italian Ambassador in Canada awarded me a special mention as part of the Italian Ingenuity Award, which aims to recognize Italian excellences conducting research in Canada. The recognition was granted for the SpeechBrain project.

- **Best PhD in ICT at University of Trento**

In May 2018, I received the best doctorate award in Information and Communication Technologies at the University of Trento. It is an important and competitive recognition given every two years by the University of Trento to the best-graduated PhD student.



The Amazon Research Awards logo, featuring the Amazon logo and the text "amazon research awards" in white and orange on a dark blue background.

- **Best PhD Student Award at Fondazione Bruno Kessler (FBK)**

I won the "Best FBK Student Award 2017". It is another very important recognition for the work and the results obtained during my PhD.



- **IBM Best Student Paper Award at IEEE ICASSP 2017**

In the context of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2017), my paper "A network of deep neural networks for distant speech recognition" won the IBM Best Student Paper Award. The work was a joint effort between FBK and Mila. ICASSP is the biggest international conference on signal processing and it is organized by IEEE.

- **Best Paper at AISV 2014**

I won the "Franco Ferrero" award for the best paper at the 2014 AISV (Italian Association of Voice Science) conference with the following paper: "Distant Talking Speech Recognition in Surgery Room: the DOMHOS project".



## EDUCATION

### University of Trento, Italy

PhD in Information and Communication Technologies

Nov 2013 - Dec 2017

- **Mark:** Cum laude distinction
- **Thesis Title:** "Deep Learning for Distant Speech Recognition"
- **Advisor:** M. Omologo



## IBM Research

### Main achievements:

- My PhD work focused on studying **deep learning** techniques for speech recognition, in particular for the challenging field of **distant (far-field) speech recognition**.
- I proposed a novel **cooperative deep learning paradigm** called



“network of deep neural networks”. This work won the “**Best IBM paper award**” at ICASSP 2017.

- I proposed a recurrent neural network for speech processing called **Light GRU**. The related journal paper has been recently awarded by **IEEE Computational Intelligence society** with the IEEE with the “*2021 IEEE TETCI Outstanding Paper award*”.
- I studied **data augmentation** and **contamination** techniques for robust neural speech processing. I contributed to the collection and distribution of the **DIRHA dataset**, a multi-microphone that is now widely used internationally.



## University of Trento, Italy

### Master Degree in Telecommunications

Ott 2009 - Feb 2011

- **Mark:** Cum laude distinction (GPA 30/30)
- **Thesis Title:** “A Multi-Microphone Front-End for Speech Transcription in Court Environment”
- **Advisor:** M. Omologo

## University of Trento, Italy

### Bachelor Degree in Telecommunications

Sett 2006 - Ott 2009

- **Mark:** Full Mark (110/110)
- **Thesis Title:** “Characterization of Avalanche Photomultipliers”
- **Advisor:** Prof. Gian Franco Dalla Betta.

## OTHER PROFESSIONAL EXPERIENCES

### Mila - Visiting Researcher

April 2016 - Ott 2016

**Supervisor:** Prof. Yoshua Bengio

I spent part of my PhD in Montreal, working at the Mila (University of Montreal). During these six months, I contributed to developing **joint training algorithms** for robust distant speech recognition and I started to explore a new paradigm (called **network of deep neural networks**), based on the progressive cooperation and communication across multiple deep learning systems.



## **I.C.S.I (University of California, Berkeley) - Visiting Researcher**

January 2013 - April 2013

**Supervisor:** Prof. Nelson Morgan

During my experience in ICSI, I had the opportunity to directly work with prof. Nelson Morgan, a pioneer of neural network approaches for speech recognition. The topic of my work was the exploration of Pre-Trained Deep Neural Networks (DNNs) for Large Vocabulary Speech Recognition in the context of the Babel project. In particular, I developed a multi-stream **Hierarchical Deep Neural Network** which was able to significantly outperform traditional speech recognition baselines.



## **Fondazione Bruno Kessler (FBK) - Researcher**

November 2011 - November 2013

**Supervisor:** Maurizio Omologo

I was involved in the European Project DIRHA (<http://dirha.fbk.eu/>), which aimed to study and develop a **multi-microphone distant-talking speech recognition** system in a domestic environment. In particular, my research was focused on the development of robust acoustic models in challenging environments characterized by the presence of both non-stationary noises and reverberation. I was also involved in the DomHos project, where I developed a distant-talking ASR system in an operating theater.

## ONGOING PROJECTS

- **SpeechBrain**

<https://speechbrain.github.io/>

I'm the founder and leader of the SpeechBrain project. SpeechBrain is an **open-source** and all-in-one **speech toolkit** relying on PyTorch. The goal is to create a single, flexible, and user-friendly toolkit that can be used to easily develop state-of-the-art speech technologies. SpeechBrain is under development and it is sponsored by important companies such as Samsung, Nvidia, Dolby, Nuance, HuggingFace, Baidu, OVH. It currently involves 30 researchers and developers from different universities in the world.



## PAST PROJECTS

- **ResearchMatch**

I'm currently co-leading a **cross-disciplinary** project funded by the McGill Initiative in Computational Medicine. The project is on the assessment of speech and language disorders via automatic speech recognition. The team is composed of both speech recognition experts and researchers in speech disorder assessment.

- **ESPERANTO (RISE 2020 EU Project)**

I'm contributing to the recently-funded Exchanges for SPEech ReseArch aNd TechnOlogies (ESPERANTO). The project is part of the 2020 RISE Call-Marie Skłodowska-Curie Actions Research and Innovation Staff Exchange. The goal is to promote and develop open-source speech technologies through research exchanges across several academic and non-academic institutions (from both EU and non-EU countries)

- **EU DIRHA (2012 - 2015)**

I contributed to the EU DIRHA project. DIRHA (Distant-Speech Interaction for Robust Home Applications) addressed the



development of **voice-enabled automated home environments** based on distant-speech interaction in different languages. During my PhD I contributed with significant scientific papers that led to international publications. Moreover, I had a primary role in the development of the DIRHA-English corpus, today well-known at international level and distributed by LDC Consortium.



- **DomHos (2012 - 2014)**

I contributed to the DOMHOS. DOMHOS was funded by the Province of Trento through the FESR program (Fondo Europeo di Sviluppo Regionale). The purpose of the project was to lead distant-talking **speech recognition technologies inside the surgery room**, a challenging but fascinating scenario. The activity was conducted with a multidisciplinary team composed of speech recognition experts and surgeons.

- **SWORDFISH (2013)**

The SWORDFISH (Spoken WORDsearch with Rapid Development and Frugal Invariant Subword Hierarchies) aimed at developing speech **recognition technologies for low-resource languages**. The project was part of the BABEL program funded by the US government through IARPA. I contributed to this project during my visit at the International Computer Science Institute (ICSI) of Berkeley.



- **AURORA (2013-2014)**

AURORA was funded by IARPA's ALADDIN program (Automated Low-level Analysis and Description of Diverse INtelligence video), which aims to combine expertise in video extraction, audio extraction, knowledge representation, and search technologies in a revolutionary way, to create fast, accurate, robust, and extensible technology that supports the multimedia analytic needs of the future.

My contribution started during my research visit at the International Computer Science Institute (ICSI), which focused on **acoustic concept detection with deep learning techniques**. The results of my research were published at several international conferences.



UNIONE EUROPEA  
Fondo europeo di sviluppo regionale

## EVENT ORGANIZATION

- **2025 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)**

<https://2025.ieeemlsp.org/en/>

I was the co-general chair of MLSP, which took place in Istanbul in September 2025. MLSP 2024 is a popular annual workshop devoted to the most recent and exciting advances in machine learning for signal processing.



- **2025 Montreal Open Tools Symposium (MOTS)**

[https://www.mcgill.ca/channels/section/crlmb/channel\\_event](https://www.mcgill.ca/channels/section/crlmb/channel_event)

I am the creator and co-organizer (with Prof. Caroline Palmer) of the **Montreal Open Source Toolkits**, which will take place on September 25–26 in Montreal at McGill University. The workshop will feature talks by prominent contributors to popular open-source projects.

- **ANNPR 2024 - The 11th IAPR TC3 Workshop on Artificial Neural Networks in Pattern Recognition,**

<https://users.encs.concordia.ca/~annpr24/>

I co-organized the ANNPR 2024 workshop. It is a major forum for international researchers and practitioners working in all areas of neural network- and machine learning-based pattern recognition to present and discuss the latest research, results, and ideas in these areas.



- **ICASSP 2024 Workshop on Explainable AI for Speech and Audio**

<https://xai-sa-workshop.github.io/web/Home.html>

I co-organized the first workshop on Explainable Machine Learning for Speech and Audio at ICASSP 2024. The event aimed at fostering research in the field of interpretability for audio and speech processing with neural networks.

- **ANNPR 2022 - The 10th IAPR TC3 Workshop on Artificial Neural Networks in Pattern Recognition,**

<https://annpr2022.com/>

I co-organized the ANNPR 2022 workshop. It is a major forum for international researchers and practitioners working in all areas of neural network- and machine learning-based pattern recognition to present and discuss the latest research, results, and ideas in

these areas.

- **ICML 2020 Workshop on Self-supervision in Audio and Speech**

<https://icml-sas.gitlab.io/>

In the context of ICML 2020, I co-organized the SAS workshop to promote the discussion on **self-supervised learning** for audio and speech.

- **2019 Sixth Frederick Jelinek Memorial Summer Workshop (JSALT)**

<https://www.clsp.jhu.edu>

In the context of JSALT 2019, I co-organized a team working on "*Cooperative Ad-hoc Microphone Arrays for ASR*". The team was composed of 18 speech experts (professors, researchers, post-docs, PhD students as well as graduate and undergraduate students) and worked for six weeks on distant **speech recognition in challenging acoustic environments**. A noise-robust version of a self-supervised **multi-task speech encoder** has been developed during this workshop.

- **2018 NIPS Workshop on Interpretability and Robustness in Audio, Speech, and Language**

<https://irasl.gitlab.io/>

In the context of NIPS 2018, I co-organized the IRASL workshop to promote the discussion on crucial topics concerning interpretability and robustness of deep learning in Audio, Speech, and Language. The workshop has been attended by more than 300 participants and the debate was enlivened by several invited talks, contributed talks, posters, and panels. 70 papers have been submitted to the workshop 30 of which have been accepted and presented at IRASL.

- Beyond organizing international workshops, I co-organized some special sessions/challenges at some popular speech conferences:
  - SLT 2022 Hackathon on the "Languages of the World"
  - ICASSP 2020 Special Session on "End-to-End Approaches for Spoken Language".
  - EUSIPCO 2014 Special Session on "Acoustic Scene Analysis in Domestic Environments".
  - EVALITA 2014: Challenge on "Speech Activity Detection and Speaker Localization in Domestic Environments"



(SASLODOM) “

## KEYNOTES AND INVITED TALKS

- **2023 World AI Summit**, keynote on “SpeechBrain: An Open-Source toolkit for Conversational AI”.
- **ICML 2022**, invited talk on “Conversational AI” at the “Machine Learning for Audio Synthesis” workshop
- **SPEAKER ODYSSEY 2020**, keynote on “Towards self-supervised learning in Audio and Speech”.  
*Odyssey 2020: The Speaker and Language Recognition Workshop* was hosted by NEC Corporation and Tokyo Institute of Technology in Tokyo, Japan, on May 17-21, 2020. This is a research workshop held in cooperation with the ISCA Speaker and Language Characterization special interest group. Odyssey is an important event in the speaker recognition community that is held every two years.
- **NeurIPS 2020 SaS workshop** on "Self-Supervised Learning for Speech and Audio Processing", invited talk on "*Multi-task self-supervised learning of speech representations*". The event was organized in the context of NeurIPS 2020, which is the largest international conference on machine learning and deep learning.
- **GTC 2020**, invited to talk on “*The SpeechBrain project*”. The GPU Technology Conference (GTC) is a global conference series providing training, insights, and direct access to experts on the topics in computing. This conference is organized by NVIDIA and covers topics like AI and Deep Learning, Accelerated data science, Graphics, simulation, IOT, and much more.
- **DOLBY-SOHO 2019**, invited talk on “*Processing Speech with SincNet*”. DOLBY-SOHO 2019 was an event organized to foster progress in artificial intelligence for audio processing.

## TUTORIALS

- Interspeech 2021: Tutorial on “*SpeechBrain: Unifying Speech Technologies and Deep Learning With an Open Source Toolkit*”.
- ASRU 2021: Tutorial on “*SpeechBrain: Unifying Speech Technologies and Deep Learning With an Open Source Toolkit*”.
- Interspeech 2022: “*A SpeechBrain for Everything: State of the PyTorch Ecosystem for Speech Technologies*”.

## SEMINARS

I'm often invited by universities and companies to give seminars on my research. Recently, I gave the following seminars:

- University of Trento, May 2023, “*Self-Supervised Learning:*



*Building Intelligent Systems Without Human Labels*".

- University of Sheffield, August 2021, "*SpeechBrain: A General-Purpose Speech Toolkit*".
- University of Siena, July 2020, "*Novel deep learning modalities in speech processing*".
- Université d' Avignon, May 2019, "*Robust deep learning in audio and speech*".
- University of Trento, May 2019, "*Toward self-supervised learning in audio and speech*".
- IBM, April 2019, "*Toward self-supervised learning in audio and speech*".

## SERVICE

I have been nominated chair of the Technical Committee of *Neural Networks and Computational Intelligence* of The *International Association for Pattern Recognition* (IAPR) for the period 2023-2025. I served as a vice-chair for the period 2021-2023.

I lead the Language and Music Models axis in the Centre for Research on Brain, Language and Music (CRBLM). CRBLM is a strategic research group with a unique interdisciplinary focus on language, music, and their intersection. The center gathers members from 4 Canadian Universities: McGill, Concordia, Université de Montréal, Université du Québec à Montréal (UQAM).

## REVIEWER ACTIVITY

I regularly serve as a reviewer for top conferences and journals dealing with deep learning (Nature, NeurIPS, ICML, ICLR, AAAI, NEPL), and speech/language processing (IEEE transactions on Audio, Speech, and Language, Compute Speech and Language, Speech Communication, ICASSP, Interspeech, ASRU, SLT).

## SOCIAL MEDIA

Communicating research results is of crucial importance to gain visibility. I recently started to be active on social media with the goal of **better communicating the most important research results** obtained by me and by my collaborators. I also created a YouTube channel where I upload short presentations of my papers.

## CODING

I strongly believe that **we really understand something when we implement it**. For this reason, I not only dedicate a lot of time trying to

read research articles and supervise students, but I often try to replicate and code many research ideas. I also encourage my students and collaborators to share this vision.

## OPEN, TRANSPARENT, and REPRODUCIBLE RESEARCH

I strongly promote **transparent**, **open**, and **reproducible research**. For this reason, I took the leadership of some popular speech-related open source projects such as SpeechBrain. During my career, I often released the code of my research results to foster replicability of experiments. This is a fundamental step towards a better science.

## PUBLICATION LIST

### 2025

D. Borra, E. Magosso, M. Ravanelli, "A protocol for trustworthy EEG decoding with neural networks", *Neural Networks*, Volume 182, 2025.

G. A. D'Inverno, S. Brugiapaglia, M. Ravanelli, "Generalization Limits of Graph Neural Networks in Identity Affects Learning", *Neural Networks*, Volume 181, 2025.

S. Zaiem, Y. Kemiche, T. Parcollet, S. Essid, M. Ravanelli, "Speech self-supervised representations benchmarking: A case for larger probing heads", *Computer Speech & Language*, Volume 89, January 2025.

P. Mousavi, G. Maimon, A. Moumen, D. Petermann, J. Shi, H. Wu, H. Yang, A. Kuznetsova, A. Ploujnikov, R. Marxer, B. Ramabhadran, B. Elizalde, L. Lugosch, J. Li, C. Subakan, P. Woodland, M. Kim, H.-y. Lee, S. Watanabe, Y. Adi, M. Ravanelli, "Discrete Audio Tokens: More Than a Survey!", *Transactions on Machine Learning Research (TMLR)*, 2025.

F. Öncel, E. Penalzoza, H. Wu, S. Gupta, M. Ravanelli, L. Charlin, C. Subakan, "Audio Prototypical Network For Controllable Music Recommendation," *Proc. of IEEE-MLSP*, 2025.

L. Della Libera, C. Subakan, M. Ravanelli, "Autoregressive Speech Enhancement via Acoustic Tokens," *Proc. of ICASSP*, 2025.

P. Plantinga, J.K. Chen, R. Sattari, M. Ravanelli, D. Klein, "From Black Box to Biomarker: Sparse Autoencoders for Interpreting Speech Models of Parkinson's Disease," *arXiv preprint arXiv:2507.16836*, 2025.

P. Plantinga, B. Cordelle, D. Louër, M. Ravanelli, D. Klein, "Does Language Matter for Early Detection of Parkinson's Disease from Speech?," *Proc. of IEEE-MLSP*, 2025.

P. Mousavi, Y. Wang, M. Ravanelli, C. Subakan, "ALAS: Measuring Latent Speech-Text Alignment For Spoken Language Understanding In Multimodal LLMs," *Proc. of the ICML 2025 Workshop on Machine Learning for Audio*.

P. Mousavi, S. Gupta, C. Subakan, M. Ravanelli, "Listen: Learning Soft Token Embeddings for Neural Audio LLMs," *Proc. of Interspeech*, 2025.

Y. Wang, A. Alhmoud, S. Alsahly, M. Alqurishi, M. Ravanelli, "Calm-Whisper: Reduce Whisper Hallucination On Non-Speech By Calming Crazy Heads Down," *Proc. of Interspeech*, 2025.

E. Mancini, F. Paissan, P. Torroni, M. Ravanelli, C. Subakan, "Investigating the Effectiveness of Explainability Methods in Parkinson's Detection from Speech," *2025 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2025.

Y. Wang, P. Mousavi, A. Ploujnikov, M. Ravanelli, "What Are They Doing? Joint Audio-Speech Co-Reasoning," *2025 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2025.

E. Mancini, F. Paissan, M. Ravanelli, C. Subakan, "LMAC-TD: Producing Time Domain Explanations for Audio Classifiers," *2025 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2025.

L. Della Libera, F. Paissan, C. Subakan, M. Ravanelli, "FocalCodec: Low-Bitrate Speech Coding via Focal Modulation Networks," *arXiv preprint arXiv:2502.04465*, 2025.

## 2024

F. Paissan, L. Della Libera, M. Ravanelli, C. Subakan, "Listenable Maps for Zero-Shot Audio Classifiers", in Proceedings of NeurIPS 2024.

F. Paissan, M. Ravanelli, C. Subakan, "Listenable Maps for Audio Classifiers", In Proc. of the International Conference on Machine Learning (ICML), oral session (1.5% acceptance).

L. Della Libera, P. Mousavi, S. Zaiem, C. Subakan, M. Ravanelli, "CL-MASR: A Continual Learning Benchmark for Multilingual ASR", accepted in the IEEE Transactions on Audio, Speech and Language Processing, 2024 (to appear)

D. Borra, F. Paissan, M. Ravanelli, "SpeechBrain-MOABB: An open-source Python library for benchmarking deep neural networks applied to EEG signals", Accepted in the Computers in Biology and Medicine Journal (Impact Factor: 7.0).

F. Oncel, C. Subakan, M. Ravanelli, "Adaptation Odyssey in LLMs: Why Does Additional Pretraining Sometimes Fail to Improve?", Accepted at EMNLP 2024.

P. Mousavi, J. Duret, S. Zaiem, L. Della Libera, A. Ploujnikov, Cem Subakan, M. Ravanelli, "How Should We Extract Discrete Audio Tokens from Self-Supervised Models?", Proc. of Interspeech, 2024.

S. Gupta, M. Ravanelli, P. Germain, C. Subakan, "Phoneme Discretized Saliency Maps for Explainable Detection of AI-Generated Voice", Proc. of Interspeech, 2024.

F. Paissan, L. Della Libera, Z. Wang, M. Ravanelli, P. Smaragdis, C. Subakan, Audio Editing with Non-Rigid Text Prompts, Proc. of Interspeech, 2024.

A. Tur, M. Ravanelli, "PROGRES: prompted generative rescoring on asr n-best", accepted at IEEE-SLT 2024

L. Della Libera, C. Subakan, M. Ravanelli, "Focal Modulation Networks for Interpretable Sound Classification", Proc. of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2024.

L. Della Libera, C. Subakan, M. Ravanelli, S. Cornell, F. Lepoutre, F. Grondin, "Resource-Efficient Separation Transformer", Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024

L. Zampierin, G. B. Hacene, B. Nguyen, M. Ravanelli, "SKILL: Similarity-aware Knowledge distillation for Speech Self-Supervised Learning", Proc. of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2024.

S.M. Mousavi, G. Roccabruna, S. Alghisi, M. Rizzoli, M. Ravanelli, G. Riccardi, "Are LLMs Robust for Spoken Dialogues?", In Proc. of the International Workshop on Spoken Dialogue Systems Technology (IWSDS), 2024

S. Mdhaffar, F. Bougares, R. De Mori, S. Zaiem, M. Ravanelli, Y. Estève, "TARIC-SLU: A Tunisian Benchmark Dataset for Spoken Language Understanding", Proc. of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING), 2024.

U. Cappellazzo, D. Falavigna, A. Brutti and M. Ravanelli, "Parameter-Efficient Transfer Learning of Audio Spectrogram Transformers", Proc. of the IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP), 2024.

D. Borra, M. Ravanelli, "Explaining network decision provides insights on the causal interaction between brain regions in a motor imagery task", Proc. of the 11th IAPR TC3 Workshop on Artificial Neural Networks in Pattern Recognition (ANNPR), 2024.

D. Borra, M. Fraternali, M. Ravanelli, E. Magosso, "Multi-modal decoding of reach-to-grasping from EEG and EMG via neural networks", Proc. of the

## 2023

C. Subakan, M Ravanelli, S Cornell, F Grondin, M Bronzi, "Exploring self-attention mechanisms for speech separation", IEEE/ACM Transactions on Audio, Speech, and Language Processing 31, 2169-2180, Vol 12, 2023

Y. Wang, M. Ravanelli, A. Yacoubi, "Speech emotion diarization: Which emotion appears when?", In Proc. of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2023

S. Sagar, M. Ravanelli, B. Kiefer, I. Kruijff-Korbayová, J. van Genabith, "RescueSpeech: A German Corpus for Speech Recognition in Search and Rescue Domain", In Proc. of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2023

J. Hwang, M. Hira, C. Chen, X. Zhang, Z. Ni, G. Sun, P. Ma, R. Huang, V. Pratap, Y. Zhang, A. Kumar, C.-Y. Yu, C. Zhu, C. Liu, J. Kahn, M. Ravanelli, P. Sun, S. Watanabe, Y. Shi, Y. Tao, "TorchAudio 2.1: Advancing speech recognition, self-supervised learning, and audio processing components for PyTorch", In Proc. of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2023

S. Zaiem, Y. Kemiche, T. Parcollet, S. Essid, M. Ravanelli, "Speech Self-Supervised Representation Benchmarking: Are We Doing it Right?", accepted at Interspeech 2023

A. Sarfi, Z. Karimpour, M. Chaudhary, N. Khalid, M. Ravanelli, S. Mudur, E. Belilovsky, "Simulated Annealing in Early Layers Leads to Better Generalization", Proc. of the IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR)

S. Zaiem, R. Algayres, T. Parcollet, S. Essid, M. Ravanelli, "Fine-tuning strategies for faster inference using speech self-supervised models: a comparative study", Proc. of ICASSP 2023

D. Beaini, S. Huang, J. A. Cunha, G. Moisescu-Pareja, O. Dymov, S. Maddrell-Mander, C. McLean, F. Wenkel, L. Müller, J. Hussein Mohamud, A. Parviz, M. Craig, M. Koziarski, J. Lu, Z. Zhu, C. Gabellini, K. Klaser, J. Dean, C. Wognum, M. Sypetkowski, G. Rabusseau, R. Rabbany, J. Tang, C. Morris, I. Koutis, M. Ravanelli, G. Wolf, P. Tossou, H. Mary, T. Bois, A. Fitzgibbon, B. Banaszewski, C. Martin, D. Masters, "Towards foundational models for molecular learning on large-scale multi-task datasets", Proc. of ICLR 2023

## 2022

Z. Wang, C. Subakan, X. Jiang, J. Wu, E. Tzinis, Efthymios, M. Ravanelli, P. Smaragdis, "Learning Representations for New Sound Classes With Continual Self-Supervised Learning," in *IEEE Signal Processing Letters*, vol. 29, pp. 2607-2611, 2022

A. Ploujnikov, M. Ravanelli, "SoundChoice: Grapheme-to-Phoneme Models with Semantic Disambiguation, Proc. of Interspeech 2023

C. Subakan, M. Ravanelli, S. Cornell, F. Grondin, "REAL-M: Towards speech separation on real mixtures", Proc. of ICASSP 2022

S.-W. Fu, C. Yu, K.-H. Hung, M. Ravanelli, Y. Tsao, "Metricgan-u: Unsupervised speech enhancement/dereverberation based only on noisy/reverberated speech", Proc. of ICASSP 2022

## 2021

L. Lugosch, P. Papreja, M. Ravanelli, A. Heba, T. Parcollet, "Timers and Such: A Practical Benchmark for Spoken Language Understanding with Numbers", accepted at NeurIPS 2021 Datasets and Benchmarks Track.

J. M. Mayor-Torres, M. Ravanelli, S. E. Medina-DeVilliers, M. D. Lerner, G. Riccardi, "Interpretable SincNet-based Deep Learning for Emotion Recognition from EEG brain activity", accepted at the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2021

S-W. Fu, C. Yu, T-A. Hsieh, P. Plantinga, M. Ravanelli, X. Lu, Y. Tsao, "MetricGAN+: An Improved Version of MetricGAN for Speech Enhancement", Proc. of Interspeech 2021

N. Dawalatabad, M. Ravanelli, F. Grondin, J. Thienpondt, B. Desplanques, H. Na, "ECAPA-TDNN Embeddings for Speaker Diarization", accepted at Interspeech 2021

T. Parcollet, M. Ravanelli, "The Energy and Carbon Footprint of Training End-to-End Speech Recognizers", Proc. of Interspeech 2021

C. Subakan, M. Ravanelli, S. Cornell, M. Bronzi, J. Zhong, "Attention is All You Need in Speech Separation", Proc. of ICASSP 2021.

## 2020

M. Ravanelli, J. Zhong, S. Pascual, P. Swietojanski, J. Monteiro, J. Trmal, Y. Bengio "Multi-task self-supervised learning for Robust Speech Recognition", Proc. of ICASSP 2020

L. Lugosch, B. Meyer, D. Nowrouzezahrai, M. Ravanelli, "Using Speech Synthesis to Train End-to-End Spoken Language Understanding Models", Proc. of ICASSP 2020

X. Qiu, T. Parcollet, M. Ravanelli, N. Lane, M. Morchid, "Quaternion Neural Networks for Multi-channel Distant Speech Recognition", In Proc. of

Interspeech 2020.

## 2019

S. Pascual, M. Ravanelli, J. Serrà, A. Bonafonte, Y. Bengio " Learning Problem-agnostic Speech Representations from Multiple Self-supervised Tasks", in Proc. of Interspeech 2019.

M. Ravanelli, Y. Bengio, "Learning Speaker Representations with Mutual Information", in Proc. of Interspeech 2019.

L. Lugosch, M. Ravanelli, P. Ignoto, V. S. Tomar, Y. Bengio, "Speech Model Pre-training for End-to-End Spoken Language Understanding", in Proc. of Interspeech 2019.

M. Ravanelli, T. Parcollet, Y. Bengio, "The PyTorch-Kaldi Speech Recognition Toolkit", in Proc. of ICASSP 2019.

T. Parcollet, M. Ravanelli, M. Morchid, G. Linarès, C. Trabelsi, R. De Mori, Y. Bengio, "Quaternion Recurrent Neural Networks", in Proc. of ICLR 2019.

## 2018

M. Ravanelli, Y. Bengio, "Interpretable Convolutional Filters with SincNet", in Proc. of NIPS@IRASL 2018.

T. Parcollet, M. Ravanelli, M. Morchid, G. Linarès, R. De Mori, "Speech Recognition with Quaternion Neural Networks", in Proc. of NIPS@IRASL 2018

M. Ravanelli, Y. Bengio, "Speaker Recognition from raw waveform with SincNet", in Proc. of SLT 2018

M. Ravanelli, D. Serdyuk, Y. Bengio, "Twin Regularization for online speech recognition", In Proc. of Interspeech 2018.

M. Ravanelli, M. Omologo, "Automatic context window composition for distant speech recognition", Speech Communication, 2018.

M. Ravanelli, P. Brakel, M. Omologo, Y. Bengio, "Light Gated Recurrent Units for Speech Recognition", in IEEE Transactions on Emerging Topics in Computational Intelligence, 2018.

## 2017

M. Ravanelli, "Deep Learning for Distant Speech Recognition", PhD Thesis, Unitn 2017

M. Ravanelli, P. Brakel, M. Omologo, Y. Bengio, "Improving speech recognition by revising gated recurrent units", in Proceedings of Interspeech

2017

M. Ravanelli, P. Brakel, M. Omologo, Y. Bengio, "A network of deep neural networks for distant speech recognition", in Proceedings of ICASSP 2017 (best IBM student paper award)

## 2016

M. Ravanelli, P. Brakel, M. Omologo, Y. Bengio, "Batch-normalized joint training for DNN-based distant speech recognition", in Proceedings of STL 2016

M. Ravanelli, P. Svaizer, M. Omologo, "Realistic Multi-Microphone Data Simulation for Distant Speech Recognition", in Proceedings of Interspeech 2016.

M. Matassoni, M. Ravanelli, S. Jalalvand, A. Brutti, "The FBK system for the CHiME-4 challenge" In Proceedings of the CHiME 4 challenge 2016.

## 2015

M. Ravanelli, L. Cristoforetti, R. Gretter, M. Pellin, A. Sosi, M. Omologo, "The DIRHA-English corpus and related tasks for distant-speech recognition in domestic environments", in Proceedings of ASRU 2015.

M. Ravanelli, B. Elizalde, J. Bernd, G. Friedland, "Insights into Audio-Based Multimedia Event Classification with Neural Networks", in Proceedings of ACM@MMCOMMONS.

M. Ravanelli, M. Omologo, "Contaminated speech training methods for robust DNN-HMM distant speech recognition", in Proceedings of INTERSPEECH 2015, Dresden, pp. 756-759.

E. Zwysig, M. Ravanelli, P. Svaizer, M. Omologo, "A multi-channel corpus for distant-speech interaction in presence of known Interferences", in Proceedings of ICASSP 2015, Brisbane, Australia.

## 2014

M. Ravanelli, M. Omologo, "On the selection of the impulse responses for distant-speech recognition based on contaminated speech training", in Proceedings of INTERSPEECH 2014, Singapore, pp. 1028-1032.

M. Matassoni, R. Astudillo, A. Katsamanis, M. Ravanelli, "The DIRHA-GRID corpus: baseline and tools for multi-room distant speech recognition using distributed microphones", in Proceedings of INTERSPEECH 2014, Singapore, pp.1613-1617.

A Brutti, M. Ravanelli, M Omologo, "SASLODOM: Speech Activity detection

and Speaker Localization in DOMestic environments" Proceedings of Evalita 2014, Pisa, Italy.

M. Ravanelli, V.H. Do, A. Janin, "TANDEM-Bottleneck Feature Combination using Hierarchical Deep Neural Networks", in Proceedings of the International Symposium on Chinese Spoken Language Processing, ISCSLP-2014, Singapore.

M. Ravanelli, B. Elizalde, K. Ni, G. Friedland, "Audio Concept Classification with Hierarchical Deep Neural Networks", in Proceeding of the European Signal Processing Conference, EUSIPCO 2014, Lisbon, Portugal.

B. Elizalde, M. Ravanelli, K. Ni, D. Borth, G. Friedland, "Audio-Concept Features and Hidden Markov Models for Multimedia Event Detection", in Proceedings of the Interspeech Workshop on Speech, Language and Audio in Multimedia, SLAM 2014, Penang, Malaysia.

A. Brutti, M. Ravanelli, P. Svaizer, M. Omologo, "A speech event detection and localization task for multiroom environments", in Proceedings of the 4th Joint Workshop on Hands-free Speech Communication and Microphone Array, HSCMA 2014, Nancy, France.

L. Cristoforetti, M. Ravanelli, M. Omologo, A. Sosi, A. Abad, M. Hagmueller, P. Maragos, "The DIRHA simulated corpus", in Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland.

## 2013

A. Sosi, F. Brugnara, L. Cristoforetti, M. Matassoni, M. Ravanelli, M. Omologo, "Embedding speech recognition to control lights", in Proceedings of INTERSPEECH 2013, Lion, France. [pdf] [video]

B. Elizalde, M. Ravanelli, G. Friedland, "Audio Concept Ranking for Video Event Detection on User-Generated Content", in Proceedings of the Interspeech Workshop on Speech, Language and Audio in Multimedia, SLAM 2013, Marseille, France

M. Ravanelli, A. Sosi, M. Matassoni, M. Omologo, M. Benetti, G. Pedrotti "Distant Talking Speech Recognition in Surgery Room: the DOMHOS project", in Proceedings of AISV 2013, Venice, Italy. [pdf] [awarded]

M. Ravanelli, A. Sosi, P. Svaizer, M. Omologo, "Impulse response estimation for robust speech recognition in a reverberant environment", in Proceeding of the European Signal Processing Conference, EUSIPCO 2012, Bucharest, Romania.