



PhD in Information Technology and Electrical Engineering
Università degli Studi di Napoli Federico II

PhD Student: Luigi Maria Giordano Orsini

Cycle: XXXIX

Training and Research Activities Report

Academic year: 2024-25 - PhD Year: Second

Luigi Maria Giordano Orsini

Tutor: prof. Francesco Cutugno

Co-Tutor:

Date: October 31, 2025

Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle:

Author:

1. Information:

- **PhD student:** Luigi Maria Giordano Orsini **PhD Cycle:** XXXIX
- **DR number:**
- **Date of birth:** 19/07/1989
- **Master Science degree:** Scienze Statistiche per le decisioni
- **University:** Università degli studi di Napoli Federico II
- **Scholarship type:** PNRR - DM 118/2023 Mis. 4.1: Dottorati Pubblica Amministrazione
- **Tutor:** Francesco Cutugno
- **Co-tutor:**
- **Period abroad:**

2. Study and training activities:

Activity	Type ¹	Hours	Credits	Dates	Organizer	Certificate ²
Participation to the Conference “CLiC-it 2024”	Research	20	3.5	04/12/2023 06/12/2023	CNR Pisa	Y
PYRLATO – A novel methodology to collect online data Lecturer: Giuseppe Magistro	Seminar	2	0.4	04/11/2024	Centro interdipartimentale Urban ECO	Y
Italiano e dialetto a Napoli Lecturer: Pietro Maturi	Seminar	2	0.4	12/11/2024	DSU - Unina	Y
Il fenomeno delle allucinazioni negli LLM Lecturer: Danilo Croce	Seminar	4	0.8	13/11/2024	Centro interdipartimentale Urban ECO	Y
Perché l'Intelligenza Artificiale crede di fare a meno della teoria	Seminar	2	0.4	20/11/2024	Centro interdipartimentale	Y

Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle:

Author:

linguistica, ma in realtà non potrà farlo Lecturer: Alessandro Lenci					Urban ECO	
AI and Enabling Technologies for Social Robots Lecturer: Silvia Rossi	Seminar	2	0.4	03/12/2024	Centro interdipartimentale Urban ECO	Y
Strutture basate su regole e strutture basate su approssimazioni Lecturer: Antonio Lieto	Seminar	2	0.4	10/12/2024	Centro interdipartimentale Urban ECO	Y
Presentazione del Volume “Eidola. Estetiche dei mondi immateriali” Lecturer: Massimo SgROI	Seminar	4	0.8	18/12/2024	Centro interdipartimentale Urban ECO	Y
Conversazioni su Etica e Intelligenza Artificiale	Courses	8	1.6	02/12/2024 12/12/2024	Prof. Lucio Franzese	Y
Evoluzione di musica e linguaggio Lecturer: Andrea Ravignani	Seminar	4	0.8	10/01/2025	Centro interdipartimentale Urban ECO	Y
The Good, the Bad, and the Ugly in Quantum Computing: Computational Power, Intrinsic Noise, and Transient Faults Lecturer: Paolo Rech	Seminar	1	0.2	17/01/2025	DIETI - Unina	Y
Participation to the workshop “Lo stato dell'arte nelle tecnologie per il riconoscimento del parlato. Il quadro italiano della ricerca pubblica e industriale”	Research	8	2	05/02/2025	Università di Urbino	Y

Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle:

Author:

Participation to the Conference “Aisv 2025”	Research	20	3.5	06/02/2025 08/02/2025	Università di Urbino	Y
I segnali discorsivi nell'italiano contemporaneo Lecturers: Violetta Cataldo, Ilaria Fiorentini, Andrea Sansò	Seminar	2	0.4	04/03/2025	DSU – Unina	Y
Matrix Analysis for Signal Processing with MATLAB Example	Courses	14	3	06/05/2025 03/06/2025	Dr. Massimo Rosamilia	Y
SPEECH PROCESSING	Courses	48	6	03/2025 06/2025	Prof. Francesco Cutugno	Y
PhysBERT: Lessons for Building a Language Model for Physics from Scratch Lecturer: Andrea Pollastro	Seminar	2	0.4	27/05/2025	DIETI – Unina	Y
Tutorship for the course of Natural Language Processing at DIETI with professor Francesco Cutugno	Tutorship	8	1.6	03/2025 06/2025	DIETI - Unina	N
La Scienza moderna e il problema della disciplina giuridica dell'IA	Courses	22	6	17/06/2025 22/07/2025	Prof. Lucio Franzese	Y
Big Data Architecture and Analytics	Courses	20	5	07/05/2025 28/05/2025	Prof. Giancarlo Sperli	Y
Text Mining	Courses	48	6	03/2025 06/2025	Prof.ssa Anna Corazza	Y
AI Code Generation: Foundations, Evaluation, and Security	Courses	12	3	07/10/2025 31/10/2025	Dr. Pietro Liguori	Y

- 1) Courses, Seminar, Doctoral School, Research, Tutorship
- 2) Choose: Y or N

Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle:

Author:

2.1. Study and training activities - credits earned

	Courses	Seminars	Research	Tutorship	Total
Bimonth 1	0	3.6	9	0	12.6
Bimonth 2	1.6	1	7	0	9.6
Bimonth 3	0	0.4	5	0	5.4
Bimonth 4	9	0.4	9	1.6	20
Bimonth 5	17	0	5	0	22
Bimonth 6	3	0	3	0	6
Total	30.6	5.4	38	1.6	75.6
Expected	30 - 70	10 - 30	80 - 140	0 - 4.8	

3. Research activity:

The long-term goal of my doctoral research is to explore novel approaches to Automatic Speech Recognition (ASR) based on syllable-like units, motivated by studies suggesting the salience of syllables in human cognitive development [1, 2]. During the first year of my PhD, I began this exploration by implementing several prototype systems, including a modified Wav2Vec2.0 [3] architecture and a Transformer-based model.

Throughout that initial exploratory phase, a recurring and fundamental challenge emerged: the difficulty in effectively representing and comparing variable-length audio segments. My main difficulty was to understand how to handle the variable vector dimensions during modeling without losing significant temporal information. This bottleneck made it clear that before a robust syllable-based ASR system could be built, a more fundamental problem needed to be solved: how do we reliably measure "similarity" between these units?

Therefore, the focus of my second year of research shifted from direct system implementation to a foundational methodological investigation. My goal was to systematically analyze the trade-offs between different distance metrics, using syllabic units as a rigorous test case to understand the relationship between computational cost and representational quality.

To address this challenge, I designed and implemented a comprehensive comparative analysis framework from the ground up. The pipeline was structured to process a large population of several thousand syllabic segments extracted from an annotated speech dataset. For each segment, I computed multiple representations to serve as inputs for the different metrics, such as, the raw waveform, a time series of openSMILE features, and a time series of MFCCs.

With the data prepared, I focused on the core of the investigation: the computation of several different distance matrices on a representative sample of a few hundred segments. The metrics were chosen to cover a wide spectrum of approaches:

- **Time-Aware Metrics:** These included Dynamic Time Warping (DTW) on raw audio and a model-based distance proposed by Piccolo, which is derived from ARIMA time series models.
- **Time-Agnostic, Feature-Based Metrics:** I also included standard Euclidean and Cosine distances on aggregated feature vectors.

Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle:

Author:

- **Proxy Metrics:** Finally, I implemented an Autocorrelation Function (ACF) based distance, designed to capture some temporal structure even from aggregated vectors.

To evaluate the quality of the resulting distance spaces, I developed an evaluation framework based on both quantitative and visual analysis. I calculated key statistics for each matrix, such as its Coefficient of Variation (CV) to measure discriminative power, and Kruskal's Stress-1 to quantify the fidelity of a 2D projection via Multidimensional Scaling (MDS).

The investigation carried out this year led to a key insight: there appears to be no single "best" metric. Instead, my analysis revealed that different metrics act as **specialized tools for different analytical tasks**.

- I found that the **Piccolo (ARIMA)** distance excels at **structural analysis**. Its ability to produce a high-variance, low-stress distance space suggests it is highly sensitive to the fine-grained generative structure of the time series.
- In contrast, **Dynamic Time Warping (DTW)** functions as a powerful **anomaly detector**. Its extremely high CV indicates a capacity to identify a dense core of "normal" data and cleanly separate structural outliers.
- Interestingly, the simple **Cosine distance on feature embeddings** emerged as a computationally efficient **proxy for DTW's anomaly detection capabilities**, showing a similar statistical profile at a negligible computational cost.
- The metrics based on **MFCCs** consistently produced high-stress projections, which led me to understand that their structure is inherently high-dimensional and not well-suited for low-dimensional visualization without further processing.

These findings have established a solid methodological foundation for the next phase of my research. The immediate goal is to quantitatively validate these observations. To do this, I am applying clustering algorithms to the computed distance matrices and will evaluate the results against multiple ground-truth criteria (phonetic frequency, balanced phonotactic structures, and novel phonetic-articulatory features) using the Adjusted Rand Index (ARI).

Comparing the results across these three experiments will provide a definitive understanding of each metric's strengths and weaknesses. Looking further ahead, this validated framework will be used to benchmark more advanced, learned representations from Autoencoders. The ultimate goal is to leverage these foundational insights to return to the original ASR problem, equipped with a principled and empirically validated methodology for representing and comparing syllabic units, with a potential application case in the analysis of infant vocalizations.

[1] Josiane Bertoncini, Jacques Mehler, Syllables as units in infant speech perception, *Infant Behavior and Development*, Volume 4, 1981, Pages 247-260, ISSN 0163-6383. Url: [https://doi.org/10.1016/S0163-6383\(81\)80027-6](https://doi.org/10.1016/S0163-6383(81)80027-6).

[2] Santolin, C., et al. (2024). Abstract processing of syllabic structures in early infancy. *Cognition*, 244, 105663. Url: <https://doi.org/10.1016/j.cognition.2023.105663>

[3] Alexei Baevski, Henry Zhou, Abdelrahman Mohamed e Michael Auli. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. 22 Oct. 2020. doi: 10.48550/arXiv.2006.11477. arXiv: 2006.11477[cs,ee]. Url: <http://arxiv.org/abs/2006.11477>

[4] Eyben, F., Wöllmer, M., & Schuller, B. (2010). Opensmile: The munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*. Url: <https://doi.org/10.1145/1873951.1874246>

Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle:

Author:

- [5] Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), 43-49. Url: <https://doi.org/10.1109/TASSP.1978.1163055>
- [6] Piccolo, D. (1990). A distance measure for classifying ARIMA models. *Journal of Time Series Analysis*, 11(2), 153-164. Url: <https://doi.org/10.1111/j.1467-9892.1990.tb00048.x>
- [7] Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons. Url: <https://doi.org/10.1111/jtsa.12194>
- [8] Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1), 1-27. Url: <https://doi.org/10.1007/BF02289565>

4. Research products:

-

5. Conferences and seminars attended

*Participation to the Conference “CLiC-it 2024”
Tenth Italian Conference on Computational Linguistics*

Participation to the workshop “Lo stato dell'arte nelle tecnologie per il riconoscimento del parlato. Il quadro italiano della ricerca pubblica e industriale”

*Participation to the Conference “Aisy 2025”
XXI Convegno Nazionale dell'Associazione Italiana Studio della Voce*

Participation to Seminars:

*PYRLATO – A novel methodology to collect online data
Lecturer: Giuseppe Magistro*

*Italiano e dialetto a Napoli
Lecturer: Pietro Maturi*

*Il fenomeno delle allucinazioni negli LLM
Lecturer: Danilo Croce*

*Perché l'Intelligenza Artificiale crede di fare a meno della teoria linguistica, ma in realtà non potrà farlo
Lecturer: Alessandro Lenci*

*AI and Enabling Technologies for Social Robots
Lecturer: Silvia Rossi*

Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle:

Author:

Strutture basate su regole e strutture basate su approssimazioni

Lecturer: Antonio Lieto

Presentazione del Volume “Eidola. Estetiche dei mondi immateriali”

Lecturer: Massimo Sgroi

Evoluzione di musica e linguaggio

Lecturer: Andrea Ravignani

The Good, the Bad, and the Ugly in Quantum Computing: Computational Power, Intrinsic Noise, and Transient Faults

Lecturer: Paolo Rech

I segnali discorsivi nell'italiano contemporaneo

Lecturers: Violetta Cataldo, Ilaria Fiorentini, Andrea Sansò

PhysBERT: Lessons for Building a Language Model for Physics from Scratch

Lecturer: Andrea Pollastro

6. Periods abroad and/or in international research institutions

-

7. Tutorship

Tutorship for the course of Natural Language Processing at DIETI with professor Francesco Cutugno

8. Plan for year three

The third year of my doctoral program will be dedicated to the synthesis, validation, and application of the foundational methodological framework developed during the second year. The primary objective is to transition from the current exploratory and comparative phase to a conclusive validation and the development of a practical application, culminating in the drafting of the final thesis.

The first major activity of the third year will be to quantitatively validate the hypotheses of metric specialization that emerged from my second-year analysis. To achieve this, I will move beyond the current exploratory stage by applying clustering algorithms, such as HDBSCAN and K-Medoids, to the full set of distance matrices I have already computed. The quality of the resulting clusters will be objectively measured using metrics like the Adjusted Rand Index (ARI). This validation will be conducted across three distinct experimental conditions, each corresponding to a different sampling strategy: one reflecting the natural frequency of syllables, one artificially balanced by phonotactic structure, and a third balanced by phonetic-articulatory features. By comparing the outcomes of these three experiments, I will be able to draw conclusions about which metrics are best suited for tasks that are primarily structural, content-based, or reflective of real-world distributions.

Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle:

Author:

With the validation framework firmly established, the research will then progress to a more advanced topic: the comparison between engineered features and learned representations. My goal here is to investigate whether a deep learning model can generate more effective embeddings than traditional acoustic features. To do this, I will implement and train a convolutional autoencoder on the entire population of syllabic data. The latent representations produced by the encoder will then be evaluated using the same robust pipeline, allowing for a direct, quantitative comparison. This will answer the critical question of whether learned representations can offer a significant advantage for the tasks of structural and content-based clustering.

The insights gained from both the metric comparison and the representation learning will culminate in the development of a practical application. I plan to build a prototype for a syllable-based similarity search engine. This tool will be designed to leverage my findings by selecting the most appropriate metric for a given user query, whether the goal is to find structurally similar syllables or perceptually similar ones. A planned research period abroad will provide the ideal setting to test this prototype on a new and challenging case study: the syllabic analysis of infant vocalizations and caregiver speech. This collaboration will not only serve to test the generalizability of my framework but also demonstrate its practical utility in a novel and impactful research context.

Draft Topic and Title of the Thesis

The thesis will narrate the research journey, from the initial challenges in ASR that motivated a methodological pivot, to the in-depth comparative analysis of metrics, the validation of learned representations, and the final application.

Draft Titles:

- *A Methodological Framework for the Representation, Comparison, and Clustering of Syllabic Audio Units.*
- *From Metrics to Meaning: Benchmarking Distance Measures for the Analysis of Variable-Length Audio Segments.*