

# SARA PAPI

## PhD Candidate in Speech Translation

✉ sara95papi@gmail.com    🐦 @sarapapi    🔗 linkedin.com/in/sara-papi



## EDUCATION

**PhD in Information and Communication Technology** - Simultaneous Speech Translation and Automatic Subtitling

Fondazione Bruno Kessler & University of Trento

📅 November 2020 – now

**Master's Degree in Computer and Automation Engineering** - Robotics and Automation

University of Siena

📅 December 2017 – February 2020

**Bachelor's Degree in Information Engineering - Systems and Automation**

University of Siena

📅 October 2014 – December 2017

## EXPERIENCE

**Internship in Streaming Speech Recognition and Translation**

Microsoft

📅 May 2023 – Aug 2023    📍 Seattle, USA

**Internship in Data-to-Text Generation**

Fondazione Bruno Kessler

📅 Jul. 2020 – Oct. 2020    📍 Trento, Italy

**Internship in Automatic Speech Scoring**

Fondazione Bruno Kessler & University of Siena

📅 Jul. 2019 – Dec. 2019    📍 Siena, Italy

## LANGUAGES

Italian (Native)

English (C1)



## PROGRAMMING LANGUAGES

Python

C/C++

Matlab

## SELECTED PUBLICATIONS

### 👤 Conference Proceedings

- Papi, Sara, Marco Gaido, Alina Karakanta, Mauro Cettolo, Matteo Negri, and Marco Turchi (Nov. 2023). "Direct Speech Translation for Automatic Subtitling". In: vol. 11, pp. 1355–1376. DOI: 10.1162/tac1\_a\_00607.
- Papi, Sara, Matteo Negri, and Marco Turchi (2023a). "AlignAtt: Using Attention-based Audio-Translation Alignments as a Guide for Simultaneous Speech Translation". In: *Proc. Interspeech 2023*.
- Gaido, Marco, Sara Papi, Matteo Negri, and Marco Turchi (2023). "Joint Speech Translation and Named Entity Recognition". In: *Proc. Interspeech 2023*.
- Papi, Sara, Matteo Negri, and Marco Turchi (2023b). "Attention as a guide for Simultaneous Speech Translation". In: *Proc. ACL 2023 (Best Paper Nomination)*.
- Papi, Sara, Marco Gaido, Matteo Negri, and Marco Turchi (Dec. 2022a). "Does Simultaneous Speech Translation need Simultaneous Models?" In: *Findings of EMNLP 2022*.
- Papi, Sara, Alina Karakanta, Matteo Negri, and Marco Turchi (Nov. 2022). "Dodging the Data Bottleneck: Automatic Subtitling with Automatically Segmented ST Corpora". In: *Proc. AACL and IJCNLP*.
- Papi, Sara, Marco Gaido, Matteo Negri, and Marco Turchi (July 2022b). "Over-Generation Cannot Be Rewarded: Length-Adaptive Average Lagging for Simultaneous Speech Translation". In: *Proc. Automatic Simultaneous Translation Workshop*.
- Papi\*, Sara, Marco Gaido\*, Matteo Negri, and Marco Turchi (Nov. 2021). "Speechformer: Reducing Information Loss in Direct Speech Translation". In: *Proc. EMNLP 2021*.
- Papi, Sara, Edmondo Trentin, Roberto Gretter, Marco Matassoni, and Daniele Falavigna (2020). "Mixtures of Deep Neural Experts for Automated Speech Scoring". In: *Proc. Interspeech 2020*.

### Accepted

- Papi, Sara, Peidong Wang, Junkun Chen, Jian Xue, Jinyu Li, and Yashesh Gaur (2023). "Token-Level Serialized Output Training for Joint Streaming ASR and ST Leveraging Textual Alignments". In: *Accepted at ASRU*.

Full list available at: [semanticscholar.org/author/Sara-Papi/2006601535](https://semanticscholar.org/author/Sara-Papi/2006601535).