NAACL-HLT 2021

# Automatic Simultaneous Translation
# Challenges, Recent Advances, and Future Directions

## The Proceedings of the Second Workshop

June, 10, 2021

©2021 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-954085-29-9

# Introduction

Welcome to the Second Workshop on Automatic Simultaneous Translation (AutoSimTrans)! Simultaneous translation, which performs translation concurrently with the source speech, is widely useful in many scenarios such as international conferences, negotiations, press releases, legal proceedings, and medicine. It combines the AI technologies of machine translation (MT), automatic speech recognition (ASR), and text-to-speech synthesis (TTS), which is becoming a cutting-edge research field.

As an emerging and interdisciplinary field, simultaneous translation faces many great challenges and is considered one of the holy grails of AI. This workshop will bring together researchers and practitioners in machine translation, speech processing, and human interpretation, to discuss recent advances and open challenges of simultaneous translation.

We organized a simultaneous translation shared task on Chinese-English. We released a dataset for open research, which covers speeches in a wide range of domains, such as IT, economy, culture, biology, arts, etc.

Following the tradition of our first workshop, we will have two sets of keynote speakers: Will Lewis, Lucia Specia, and Liang Huang from simultaneous translation, and Hong Jiang from human interpretation research. We hope this workshop will greatly increase the communication and cross-fertilization between the two fields. We have accepted 6 papers that will be presented in the second session.

We look forward to an exciting workshop!

Hua Wu, Colin Cherry, Liang Huang, Zhongjun He, Qun Liu, Maha Elbayad, Mark Liberman, and Haifeng Wang

**Organizers:**

Hua Wu, Baidu Inc.
Colin Cherry, Google
Liang Huang, Oregon State University and Baidu Research
Zhongjun He, Baidu Inc.
Qun Liu, Huawei Noah's Ark Lab
Maha Elbayad, Université Grenoble Alpes

**Steering Committee:**

Mark Liberman, University of Pennsylvania
Haifeng Wang, Baidu Inc.

**Program Chair:**

Mingbo Ma, Baidu Research, USA

**Shared Task Chair:**

Ruiqing Zhang, Baidu Inc.

**Program Committee:**

Naveen Arivazhagan, Google, USA
Chung-Cheng Chiu, Google, USA
Kenneth Church, Baidu Research, USA
Yang Feng, CAS/ICT, China
George Foster, Google, Canada
Alvin Grissom II, Ursinus College, USA
He He, NYU, USA
Alina Karakanta, FBK-Trento, Italy
Wei Li, Google, USA
Hairong Liu, Baidu Research, USA
Kaibo Liu, Baidu Research, USA
Wolfgang Macherey, Google, USA
Jan Niehues, Maastricht U., Netherlands
Yusuke Oda, Google, Japan
Colin Raffel, Google, USA
Elizabeth Salesky, CMU, USA
Jiajun Zhang, CAS/IA, China
Renjie Zheng, Oregon State Univ., USA

**Invited Speakers:**

Will Lewis, Affiliate Faculty, University of Washington
Lucia Specia, Professor, Imperial College London, UK
Liang Huang, Assoc. Professor, Oregon State and Distinguished Scientist, Baidu Research, USA
Hong Jiang, Sr. Lecturer, Dept. of Translation, Chinese University of Hong Kong, China

# Table of Contents

*ICT's System for AutoSimTrans 2021: Robust Char-Level Simultaneous Translation*
Shaolei Zhang and Yang Feng . . . . . 1

*BIT's system for AutoSimulTrans2021*
Mengge Liu, Shuoying Chen, Minqin Li, Zhipeng Wang and Yuhang Guo . . . . . 12

*XMU's Simultaneous Translation System at NAACL 2021*
Shuangtao Li, Jinming Hu, Boli Wang, Xiaodong Shi and Yidong Chen . . . . . 19

*System Description on Automatic Simultaneous Translation Workshop*
Linjie Chen, Jianzong Wang, Zhangcheng Huang, Xiongbin Ding and Jing Xiao . . . . . 24

*BSTC: A Large-Scale Chinese-English Speech Translation Dataset*
Ruiqing Zhang, Xiyang Wang, Chuanqiang Zhang, Zhongjun He, Hua Wu, Zhi Li, Haifeng Wang, Ying Chen and Qinfei Li . . . . . 28

*Findings of the Second Workshop on Automatic Simultaneous Translation*
Ruiqing Zhang, Chuanqiang Zhang, Zhongjun He, Hua Wu and Haifeng Wang . . . . . 36

# Conference Program

**Thursday, June 10, 2021 (all time in PST)**

**07:15–07:30** ***Opening Remarks***

**07:30–10:10** **Session 1 Invited Talks**

07:30–08:10 *Invited Talk by Hong Jiang*

08:10–08:50 *Invited Talk by Lucia Specia*

08:50–09:30 *Invited Talk by Liang Huang*

09:30–10:10 *Invited Talk by Will Lewis*

**10:10–19:00** **Break**

**19:00–20:20** **Session 2: Research Paper and System Description**

19:00–19:10 *ICT's System for AutoSimTrans 2021: Robust Char-Level Simultaneous Translation*
Shaolei Zhang and Yang Feng

19:10–19:20 *BIT's system for AutoSimulTrans2021*
Mengge Liu, Shuoying Chen, Minqin Li, Zhipeng Wang and Yuhang Guo

19:20–19:30 *XMU's Simultaneous Translation System at NAACL 2021*
Shuangtao Li, Jinming Hu, Boli Wang, Xiaodong Shi and Yidong Chen

19:30–19:40 *System Description on Automatic Simultaneous Translation Workshop*
Linjie Chen, Jianzong Wang, Zhangcheng Huang, Xiongbin Ding and Jing Xiao

19:40–19:50 *BSTC: A Large-Scale Chinese-English Speech Translation Dataset*
Ruiqing Zhang, Xiyang Wang, Chuanqiang Zhang, Zhongjun He, Hua Wu, Zhi Li, Haifeng Wang, Ying Chen and Qinfei Li

**Thursday, June 10, 2021 (all time in PST) (continued)**

19:50–20:00 *Findings of the Second Workshop on Automatic Simultaneous Translation*
Ruiqing Zhang, Chuanqiang Zhang, Zhongjun He, Hua Wu and Haifeng Wang

20:00–20:20 *Q&A*

**20:20–20:30** ***Closing Remarks***