QCRI's Live Speech Translation System

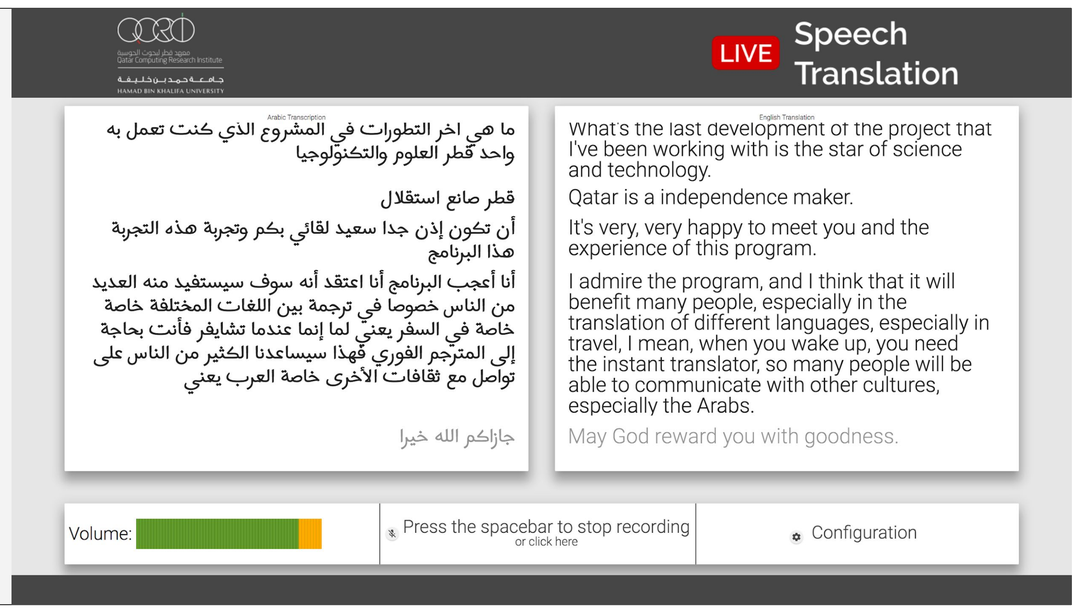
Fahim Dalvi, Yifan Zhang, Sameer Khurana, Nadir Durrani, Hassan Sajjad
Ahmed Abdelali, Hamdy Mubarak, Ahmed Ali, Stephan Vogel

Motivation

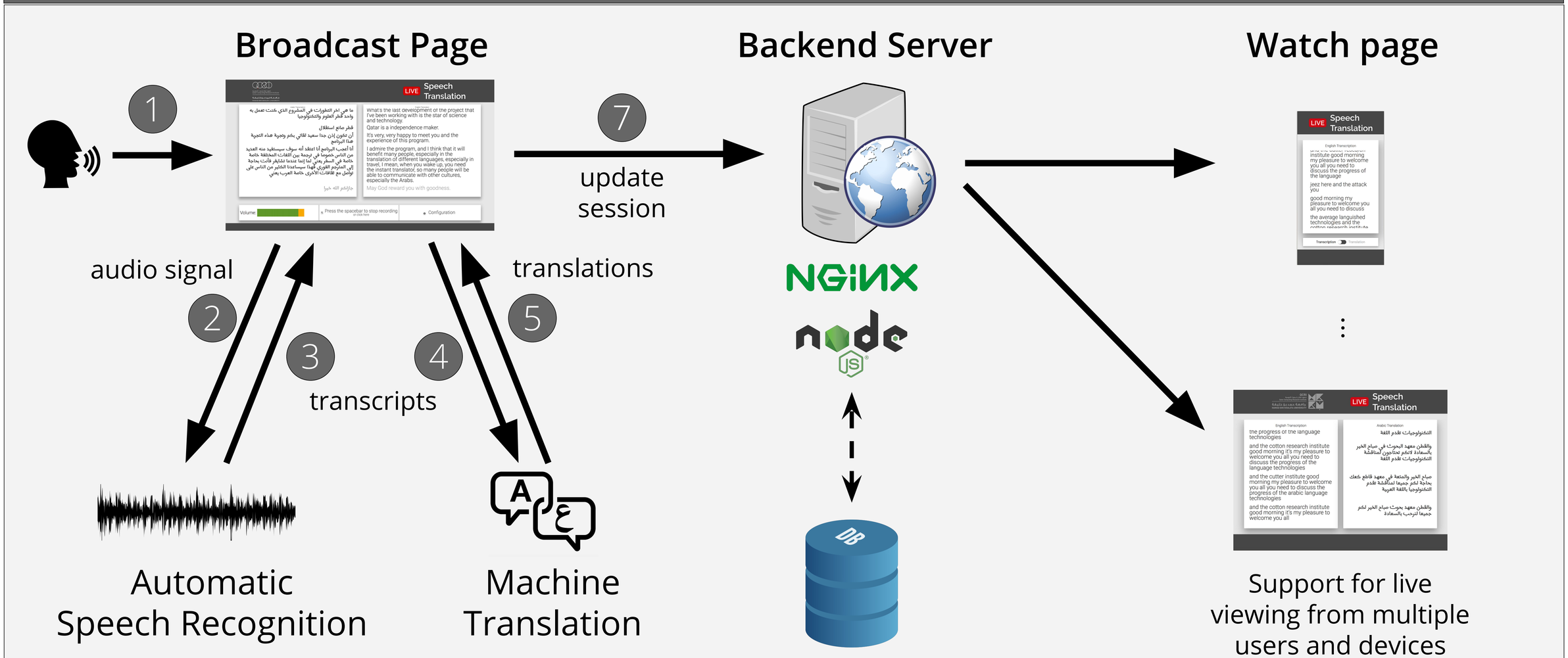
Enable **real-time, low-latency** transcription/translation for Arabic & English

Core System

- Web-based application that streams live audio from a user's browser
- Transcribed and translated results are streamed back with a overall real-time factor of 1.10



Architecture



Speech Recognition

- Grapheme-based pronunciation dictionary of one million words
- Time delayed neural networks (TDNN)
- 1000-dimensional i-Vectors latent variable for speaker adaptation
- Kneser-Ney smoothed tri-gram language model for decoding
- Code-Switching support for frequent English words

Machine Translation

- Phrase-based (PB) (Moses) and Neural Machine Translation (Nematus) systems trained on freely available Arabic and English data
- PB system used phrase-tables, lexicalized reordering, 5-gram language model, all pruned to fit in memory
- Neural MT systems used bilingual LSTM with attention model (50K BPE'd vocab)

This system is available online at https://st.qcri.org/demos/livetranslation

Dalvi, Fahim, Yifan Zhang, Sameer Khurana, Nadir Durrani, Hassan Sajjad, Ahmed Abdelali, Hamdy Mubarak, Ahmed Ali, and Stephan Vogel. "QCRI's live speech translation system." In *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 61-64. 2017.

{faimaduddin,yzhang}@qf.org.qa