

# Does Speech enhancement of publicly available data help build robust Speech Recognition Systems?

Bhavya Ghai,<sup>1</sup> Buvana Ramanan,<sup>2</sup> Klaus Mueller<sup>1</sup>

<sup>1</sup>Department of Computer Science, Stony Brook University, USA

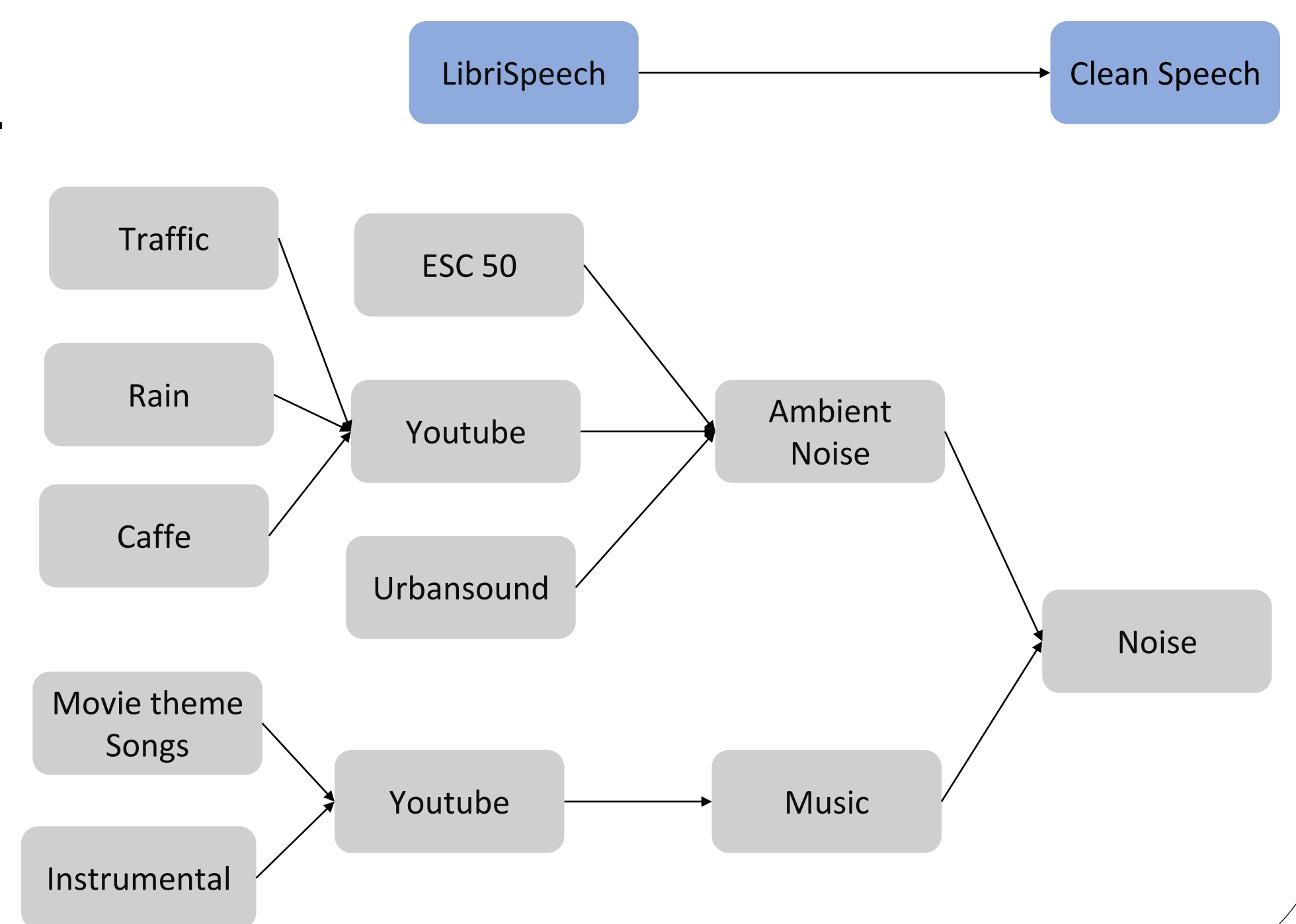
<sup>2</sup>Nokia Bell Labs, Murray Hill, USA

## Objective

- Build robust Automatic Speech Recognition (ASR) systems comparable to the likes of big organizations like Google, Microsoft, etc. without access to huge private repositories of clean labeled dataset
- This will provide a level playing field for startups, academics and other small companies and research organizations to compete with larger organizations.

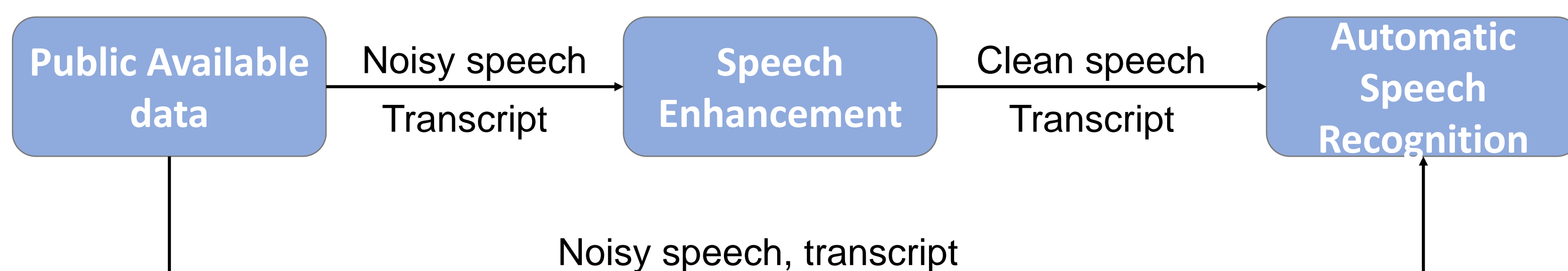
## Dataset

- Existing datasets for speech enhancement are pretty limited in size. So, we decided to curate our own dataset.
- For clean speech, we used LibriSpeech dataset which is derived from public domain audiobooks (~460 hours)
- For ambient noise, we used popular datasets like Urbansound, ESC50 and youtube videos consisting of background noise in rain, traffic, restaurant, etc.
- For background music, we used youtube videos to extract movie theme songs and instrumental music belonging to diverse genres like Latin, African, Heavy metal, Native American, Japanese, Indian, etc.



## Our Approach

- Speech from publicly available sources is first cleaned using SEGAN Algorithm.
- Cleaned speech is contaminated with different kinds of ambient noise and background music to simulate different real-world conditions.
- ASR based on DeepSpeech model is trained using noisy & clean speech together (Multi-condition training).



## Results

- We observed **9.5%** mean reduction in Word error rate (WER) using our approach compared to the baseline i.e. training on noisy data while evaluating for both clean and noisy test datasets.
- Our approach also performs at par with ideal case on Noisy test data. On clean test data, our WER is a bit higher which might be due to artifacts introduced during speech enhancement.

