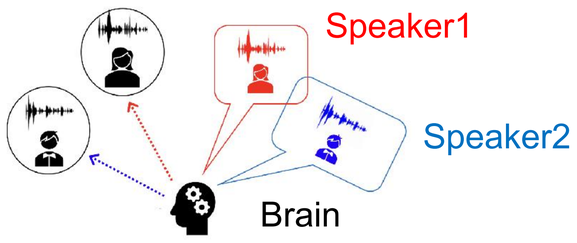
# NOISE-AWARE SPEECH SEPARATION WITH CONTRASTIVE LEARNING

Zizheng Zhang, Chen Chen, Hsin-Hung Chen,
Xiang Liu, Yuchen Hu, Eng Siong Chng

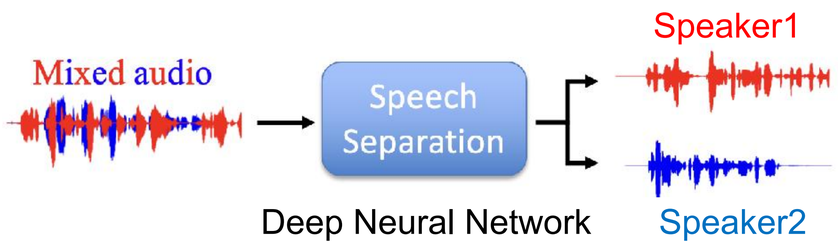
北京大学 PEKING UNIVERSITY
NANYANG TECHNOLOGICAL UNIVERSITY
Georgia Tech

## SPEECH SEPARATION TASK

**For Human:**

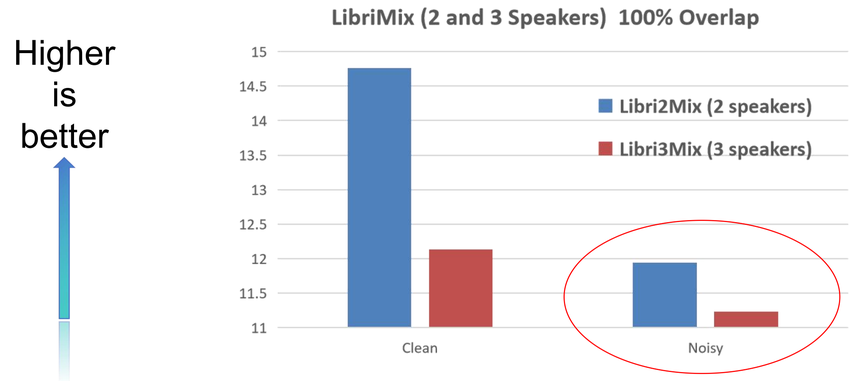



**For Machine:**



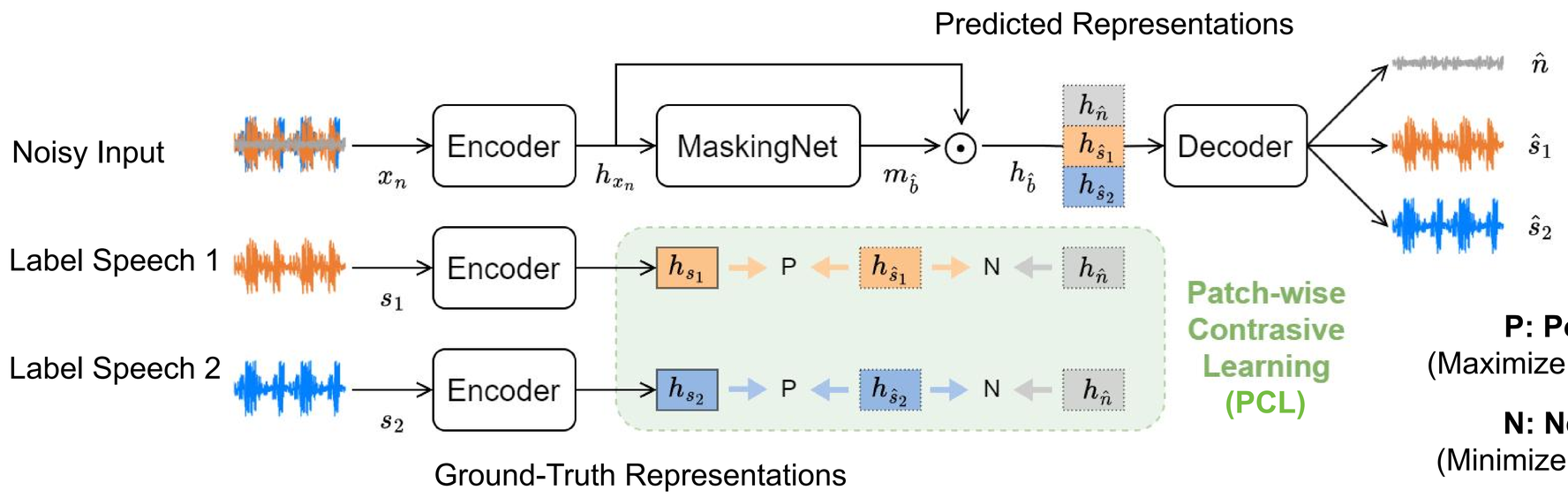

## PRIOR WORK

**Sepformer Model on LibriMix Dataset:**




We are focusing on improving the performance under **noisy condition**

## OVERALL PIPELINE OF PROPOSED METHODS (NASS)




**Additional Noise Output (ANO)**

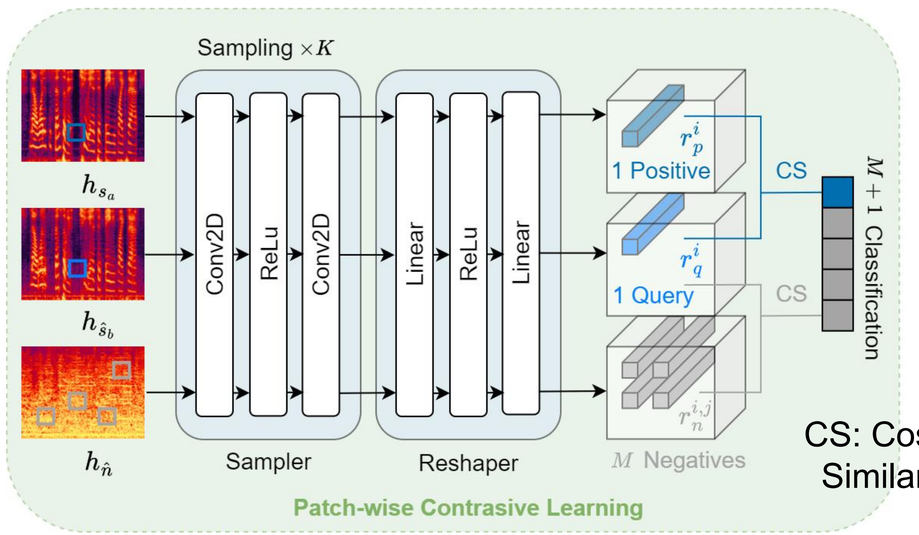
Separated Speech 1

Separated Speech 2

**P: Positive Examples**
(Maximize the Mutual Information)

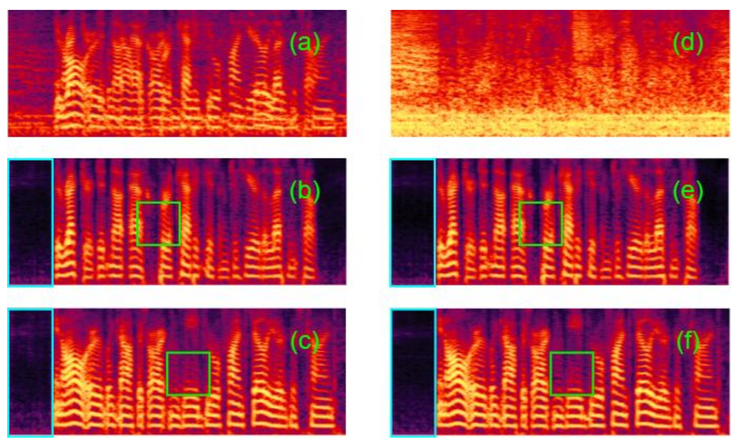
**N: Negative Examples**
(Minimize the Mutual Information)

## DETAILS OF PCL




CS: Cosine Similarity

## SPECTRUM RESULTS




(a): Noisy Mixture
(b): Spk1 (Sepformer)
(c): Spk2 (Sepformer)
(d): Noise (NASS)
(e): Spk1 (NASS)
(f): Spk2 (NASS)

## LIBRIMIX & WHAM! RESULTS

**For LibriMix:**

| Method | Spk | SI-SNRi (dB) | SDRi (dB) | Params (M) |
|---|---|---|---|---|
| DPRNN* | 2 | 13.1 | 13.7 | 14.8 |
| | 3 | 11.3 | 11.9 | 14.9 |
| Sepformer* | 2 | 13.2 | 13.8 | 25.8 |
| | 3 | 10.0 | 10.5 | 25.9 |
| DPRNN (NASS) | 2 | **13.5** | **14.1** | 14.8 |
| | 3 | **12.4** | **12.9** | 14.9 |
| Sepformer (NASS) | 2 | **14.4** | **15.0** | 25.8 |
| | 3 | **12.1** | **12.7** | 25.9 |

**For WHAM!:**

| Method | SI-SNRi (dB) | SDRi (dB) | Params (M) |
|---|---|---|---|
| DPRNN [3] | 13.7 | 14.1 | 2.7 |
| Sepformer [4] | 14.4 | 15.0 | 26.0 |
| DPRNN (NASS) | **15.8** | **16.1** | 14.9 |
| Sepformer (NASS) | **16.5** | **16.8** | 25.8 |

## CONTRIBUTIONS

- We propose a noise-aware speech separation (NASS) method to make use of the noise information and the mutual information.
- NASS consists of an additional noise output (ANO) and patch-wise contrastive learning (PCL) to further separate in detail.
- NASS achieves **1 to 2dB SI-SNRi or SDRi** over DPRNN and Sepformer on WHAM! and LibriMix noisy datasets, with **less than 0.1M parameter increase**.

Repo Link

Paper Link