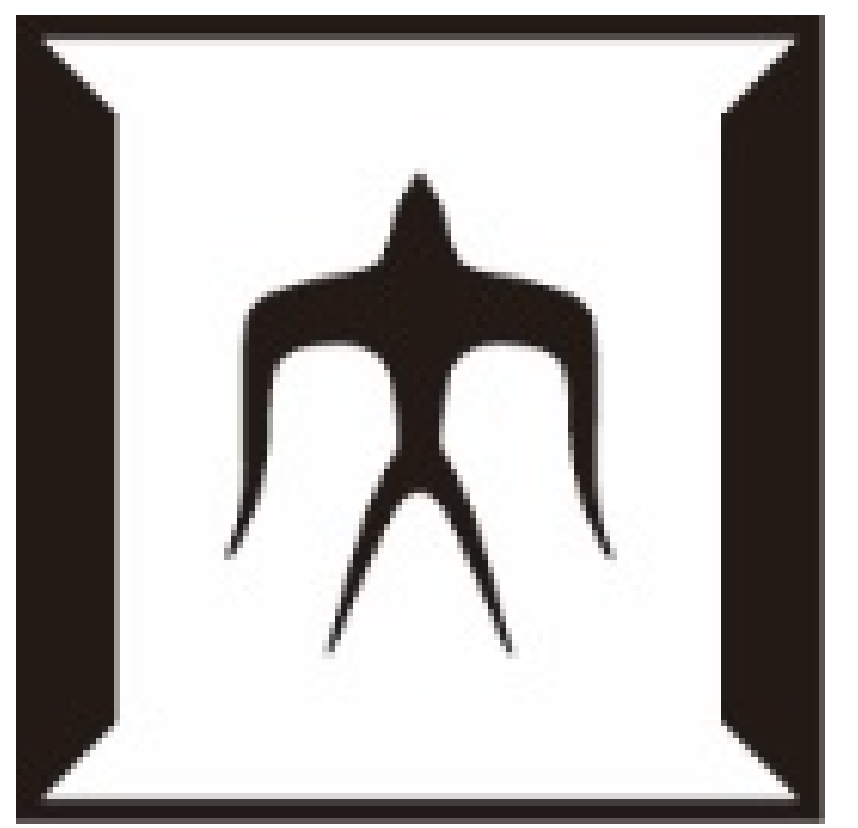


# Multi-Task Autoencoder For Noise-Robust Speech Recognition

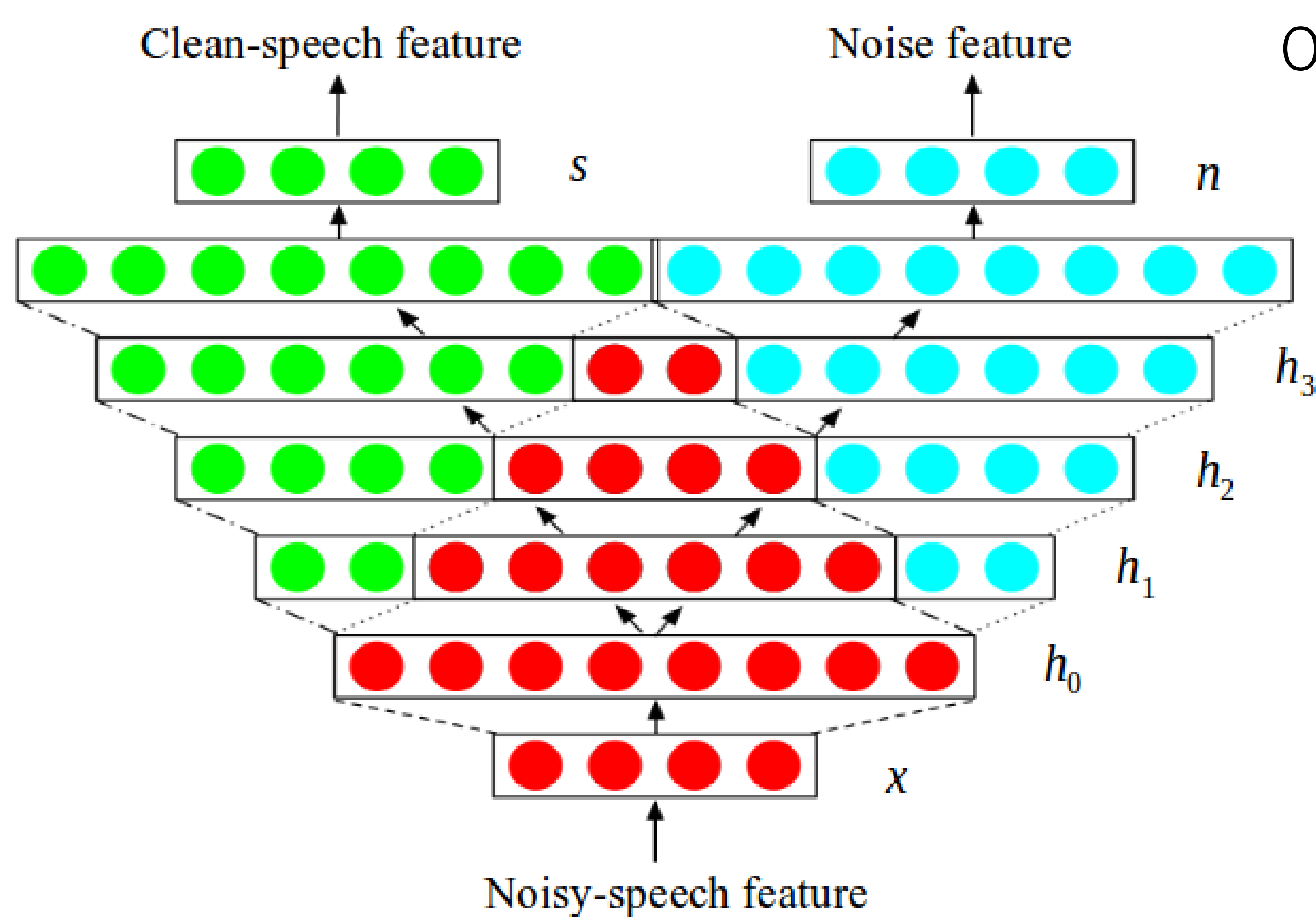


Haoyi Zhang, Liu Conggui, Nakamasa Inoue, Koichi Shinoda  
Tokyo Institute of Technology

## Background

- Deep Denoising Autoencoder (DDAE) is often used as a preprocessor for speech recognition.
- We add another task, “deSpeech” from noisy speech, to form a multitask encoder.
- We expect it will enhance the denoising performance.

## Multi-Task Autoencoder (MTAE)



Objective function:

$$c\|s' - s\| + (1 - c)\|n' - n\|$$

(We use  $c = 0.5$ )

There is no connection between green nodes and blue nodes.

## Experiments

Input feature: MFCC 13 dim x 11 frames

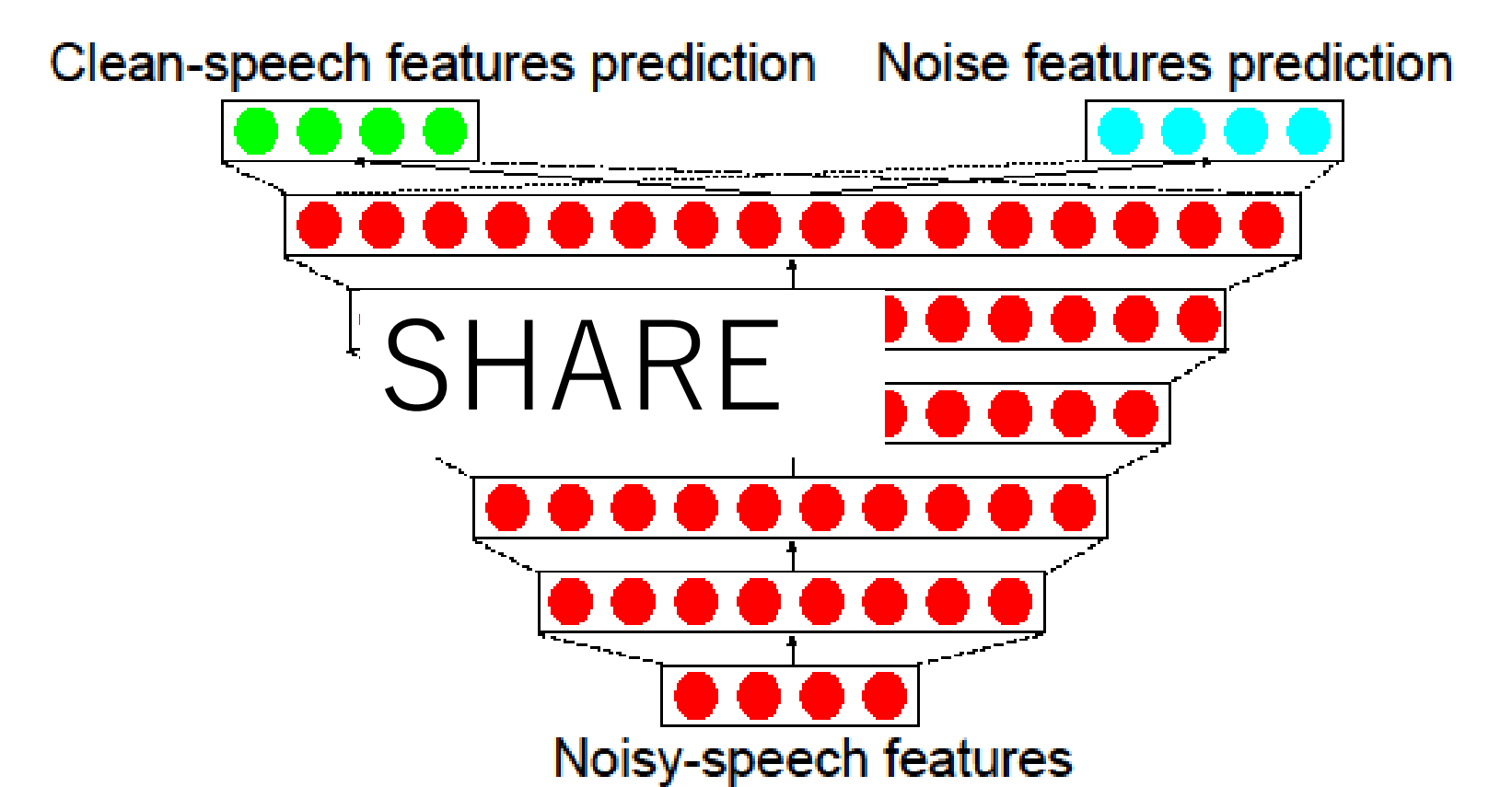
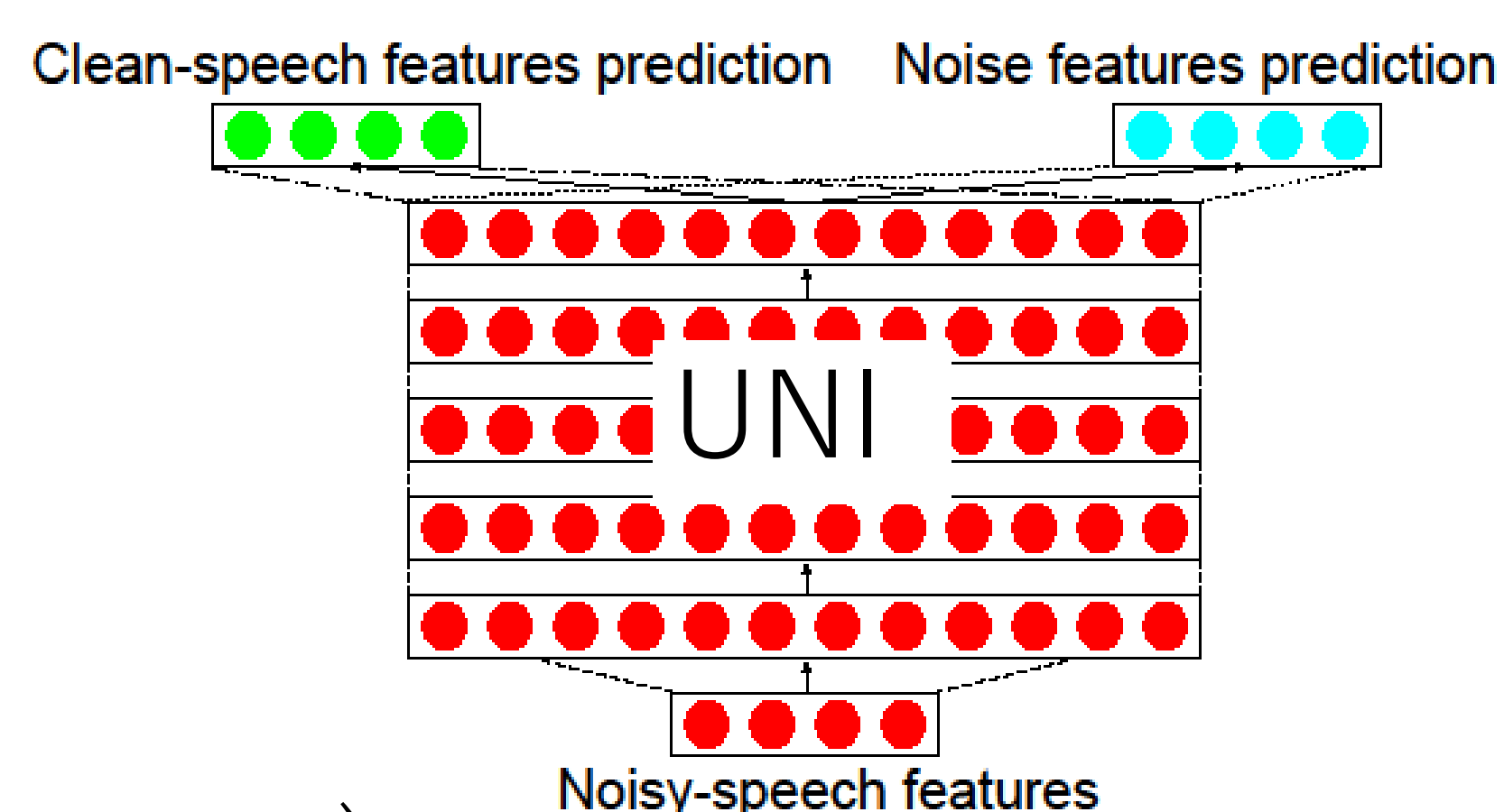
Aurora 2 result (WER %)

Training: train, bubble, car, exhibition

Test Set A: train, bubble, car, exhibition

Test Set B: restaurant, street, airport, station

Test Set C: train, street with MIRS filter (others G.712)



	Multi-Task AutoEncoder (MTAE)									MFCC	ETSI -AFE	NMCC	DDAE	UNI	SHARE
	Clean	20dB	15dB	10dB	5dB	0dB	-5dB	Ave.							
Set A	0.7	1.0	1.3	2.3	5.0	17.9	53.6	<b>5.5</b>	38.7	12.2	16.8	6.4	7.0	7.3	
Set B	0.7	1.3	2.5	7.3	20.1	48.0	73.1	15.8	44.3	<b>12.9</b>	14.5	20.4	-	-	
Set C	0.7	1.1	2.2	5.4	10.9	36.8	68.3	<b>11.3</b>	33.9	14.2	17.5	17.2	-	-	
Ave.	0.7	1.1	2.0	5.0	12.0	34.2	55.0	<b>10.9</b>	38.9	13.1	16.4	14.7	-	-	

CHiME3 result (WER %)

Training: café, bus, street, pedestrian area

Test Set : real noisy data “et05\_real”

Averaged over 0-20dB

	MFCC	DDAE	MTAE
et05_real	50.8	29.7	26.9

## Summary

- Extracting noise help extracting speech in noisy speech.
- The partly-shared structure is effective in multi-task learning.
- The use of raw or log spectral features may be simpler and more effective.