A LOCALLY LINEAR EMBBEDING BASED POSTFILTERING APPROACH FOR SPEECH ENHANCEMENT

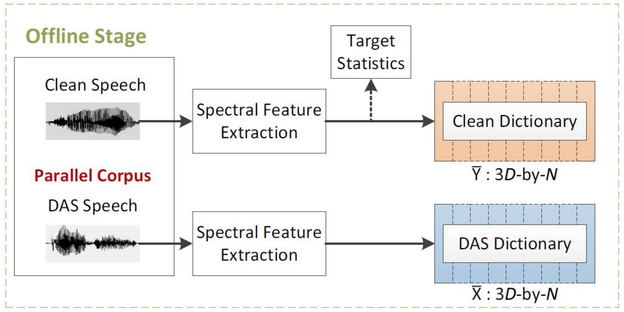
Yi-Chiao Wu, Hsin-Te Hwang, Syu-Siang Wang, Chin-Cheng Hsu,
Ying-Hui Lai, Yu Tsao, and Hsin-Min Wang

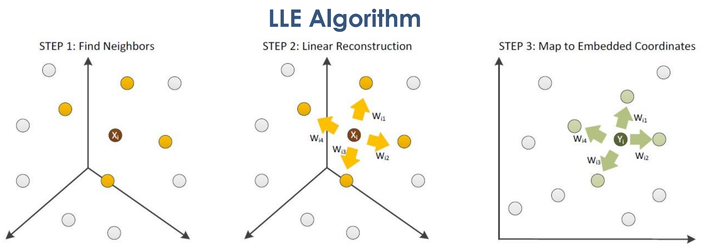
Introduction

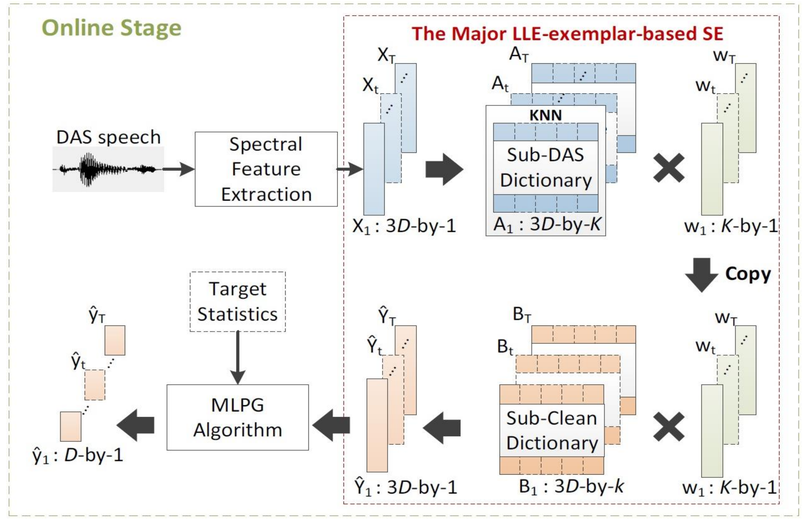
- When a sufficient training data is available, DNN-based approaches get outstanding performance.
- LLE is a manifold learning method for super resolution, which attempts to discover the intrinsic geometry of the high-dimensional data and then embed them onto low-dimensional embedding space.
- We have successfully applied LLE to speaker voice conversion, but employed LLE to directly convert noisy speech to clean speech could not get satisfactory performance
- We noted that LLE-based postfiltering could be combined with DAS to further remove the residual noise components, and thus improve the SNR and speech quality.

System Framework

- The main concept of the proposed system is to perform voice conversion (from DAS-processed speech to clean speech) based on the LLE algorithm. The proposed system includes **offline** and **online** stages.







Experimental Results

- **Corpus:** Mandarin hearing in noise test (MHINT)(300 utterances of single speaker, 250 for training and 50 for testing)
- **DAS system:**
  - Structure: 7 hidden layers with 1200, 300, 300, 514, 300, 300, 1200 hidden nodes
  - Training data: 250 utterances are mixed with -10~20 dB (5dB interval) car/two-talker noise
- **Proposed system:**
  - Five-fold cross validation with 50 utterances from the testing set of MHINT
  - Dictionary: 40 utterances with SNR -10, 0, 10 dB car/two-talker noise
  - Testing data: 10 utterances with SNR -10, -6, -2, 0, 2, 6, 10 dB car/ two talk noise

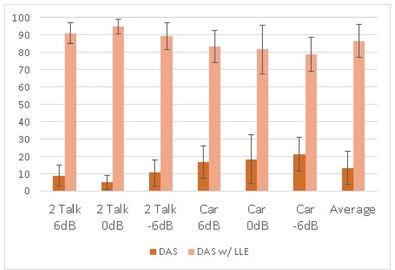
Objective evaluation of two-talker noise

	DAS w/ LLE			DAS		
	PESQ	STOI	SSNR	PESQ	STOI	SSNR
SNR10	**2.22**	0.83	**12.73**	2.21	**0.88**	12.48
SNR6	**2.11**	0.82	**12.08**	2.05	**0.86**	11.76
SNR2	**1.97**	0.80	**10.88**	1.93	**0.84**	10.47
SNR0	**1.86**	0.79	**10.12**	1.83	**0.83**	9.66
SNR-2	**1.78**	0.78	**9.03**	1.75	**0.81**	8.46
SNR-6	1.59	0.75	**6.13**	**1.61**	**0.78**	5.38
SNR-10	1.42	0.69	**2.53**	**1.47**	**0.72**	1.51
Ave	**1.85**	0.78	**9.07**	1.83	**0.82**	8.53

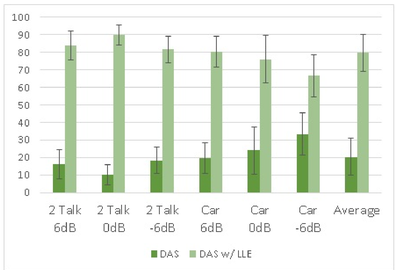
Objective evaluation of car noise

	DAS w/ LLE			DAS		
	PESQ	STOI	PESQ	STOI	PESQ	STOI
SNR10	**2.03**	0.80	**15.73**	1.96	**0.85**	15.04
SNR6	**1.99**	0.79	**14.91**	1.93	**0.84**	14.17
SNR2	**1.92**	0.78	**13.37**	1.89	**0.83**	12.40
SNR0	**1.86**	0.78	**12.34**	1.85	**0.82**	11.40
SNR-2	**1.82**	0.77	**11.05**	1.81	**0.81**	10.00
SNR-6	1.71	0.75	**7.74**	**1.75**	**0.79**	6.34
SNR-10	1.60	0.72	**3.90**	**1.67**	**0.76**	2.22
Ave	**1.85**	0.77	**11.29**	1.84	**0.81**	10.23

Noise reduction capability subjective test



Preference subjective test



Conclusion

- We investigate the use of the LLE algorithm with the paired SE-processed and clean dictionaries for postfiltering SE task.
- Experiment results shows that DAS w/ LLE system gets better subjective preference, and objective PESQ and SSNR than baseline DAS system.