# ICIEA 2022

**16-19 Dec.**
**Chengdu, China**

# ICIEA22-000346
# Deep Clustering in Complex Domain for Single-Channel Speech Separation

**Runling Liu[1],Yu Tang[1],Hongwei Zhang[2]**

**1. Southwest Jiaotong University**

**2. Harbin Institute of Technology, Shenzhen**

## DEEP CLUSTERING IN COMPLEX DOMAIN

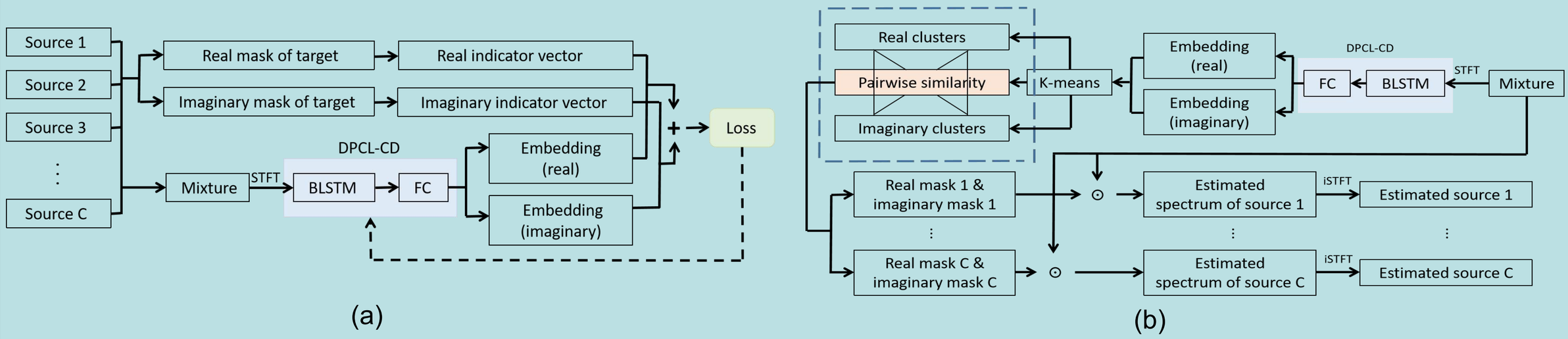


**Fig. 1. Block diagram of speech separation with deep clustering in complex domain: (a) training stage; (b) testing stage.**

Block diagram of the single-channel speech separation using DPCL-CD is shown in Fig. 1. It consists of two stages. In the training stage, the masks of the real and imaginary parts are calculated by using C clean sources, and the indicator vectors are obtained according to the true masks. Neural network is trained to assign two embedding vectors for each time-frequency bin. The network parameters are updated by minimizing the loss of both real and imaginary parts. The new model is thence used in testing stage to generate the embedding vectors. Embedding vectors of the real and imaginary parts are clustered by K-means, and thence the corresponding binary masks are obtained. The similarity scores are calculated to match the binary masks of the real part with the imaginary part. Furthermore, the masks are multiplied with the mixture to obtain the frequency domain estimates for each source. Finally, the iSTFT is used to obtain an estimate of each source in the time domain.

## EXPERIMENTAL RESULTS

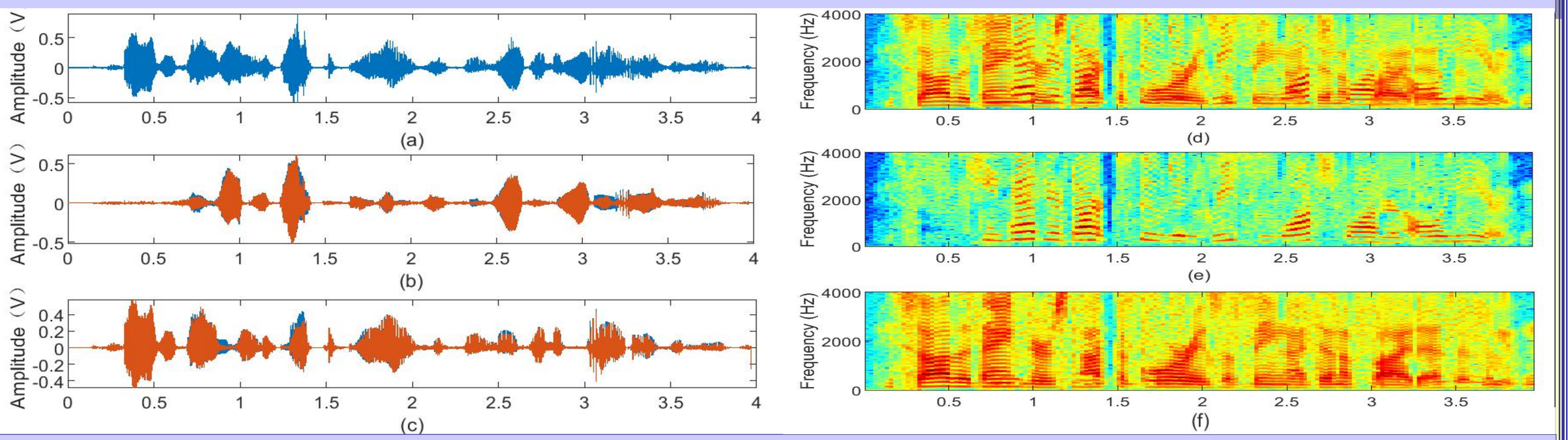


**Fig. 2. An example of two-speaker separation.**

## Summary

- We proposed a novel model to extend deep clustering to complex domain for single-channel speech separation in the speaker-independent case. The network generates two embedding vectors for each T-F bin to estimate the binary mask of the real part and the imaginary part respectively. The permutation problem of the binary masks is handled by calculating the similarity score. Reconstructed signals can be obtained directly by using the complex masks.
- Experimental results show that the proposed method performs well in both unseen speakers and languages, which is superior to the model of DPCL.
- The network structure can be used in all models based on deep clustering, since it is easy in implementation.