SINGLE SOURCE AUDIO SEPARATION

Author(s) Name(s)

Author Affiliation(s)

ABSTRACT

In this paper we propose a Deep Attractor based audio separation method from single sources.

***Index Terms*—** Single Source Audio Separation, Deep Attractor, Wasserstein GAN

1. INTRODUCTION

2. RELATED WORK

2.1. LSTM Denoising autoencoders

Previous works have used the Long-Short Term Memory (LSTM) denoising autoencoders, which has shown better performance than feature-mapping networks. A disadvantage of this method is that it requires a fixed number of sources.

2.2. Deep clustering-based separator

DC can solve both permutation and output dimension problem to produce the state of the art separation performance. However, the main drawback of DC is its inefficiency to perform end-to-end mapping, because the objective function is the affinity between the sources in the embedded space and not the separated signals themselves. Minimizing the separation error is done with an unfolding clustering system and a second network, which is trained iteratively and stage by stage to ensure convergence.

3. MODEL DESIGN

4. DATA COLLECTION

It contains a 30 h training set and a 10 h validation set generated by randomly selecting utterances from different speakers in the Wall Street Journal (WSJ0) training set sitrs, and mixing them at various signal-to-noise ratios (SNR) randomly chosen between 0 dB and 10 dB.

Thanks to XYZ agency for funding.

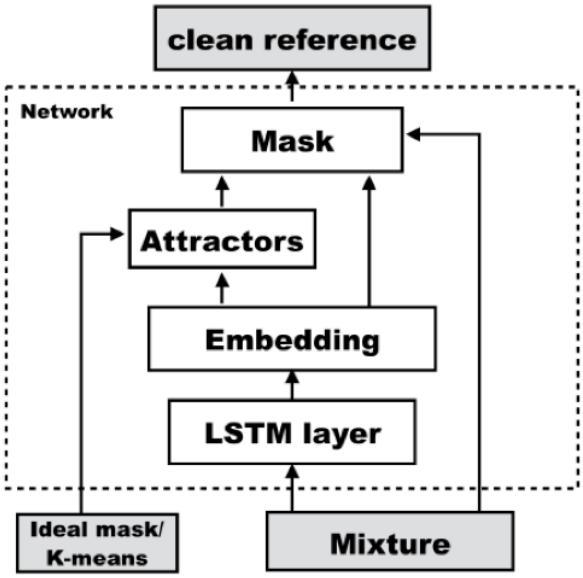

Fig. 1. The system architecture. In the training time, a ideal mask is applied to form the attractor, while during the testing time, Kmeans is used to form the attractor. Alternatives for Kmeans is further discussed in Section 2.3

(a) Result 1

Fig. 1. System architecture.

5. EXPERIMENTATION

6. EVALUATION

7. CONCLUSION

8. REFERENCES