Selective Transparent Headphone

Zichao Wang, Tianyi Yao, Stephen Xia, Shengjia Zhao

Rice University Department of Electrical and Computer Engineering

RICE ece redefining limits

Objective

Explore ways to build a selective transparent headphone that propagates speech signal and attenuates all other types of signals.

Project Overview

- Since our objective is to blindly separate an input signal into its principal/independent components, we will use independent component analysis to separate the signal into its components.
- Then, we will utilize a neural network to select which output signal, if any, contains human speech, and forward it to the output.

Independent Component Analysis

- Problem formulation:
  - *y = input signals (sound mixtures)*
  - *s = original sources*
- ICA attempts to find the independent sources (individual sounds) that comprise an input signal by solving the following formulation for A^{-1}

$$y = As$$

- *A* is the mixing matrix used to combine the original sources, *s*, to obtain the observed input signals, *y*.

Blind Source Separation (BSS)

- Transform *M* microphone inputs into frequency domain by taking the short-time Fourier transform (STFT).

$$x(\omega,t)=[X_1(\omega,t),\dots,X_M(\omega,t)]^T$$

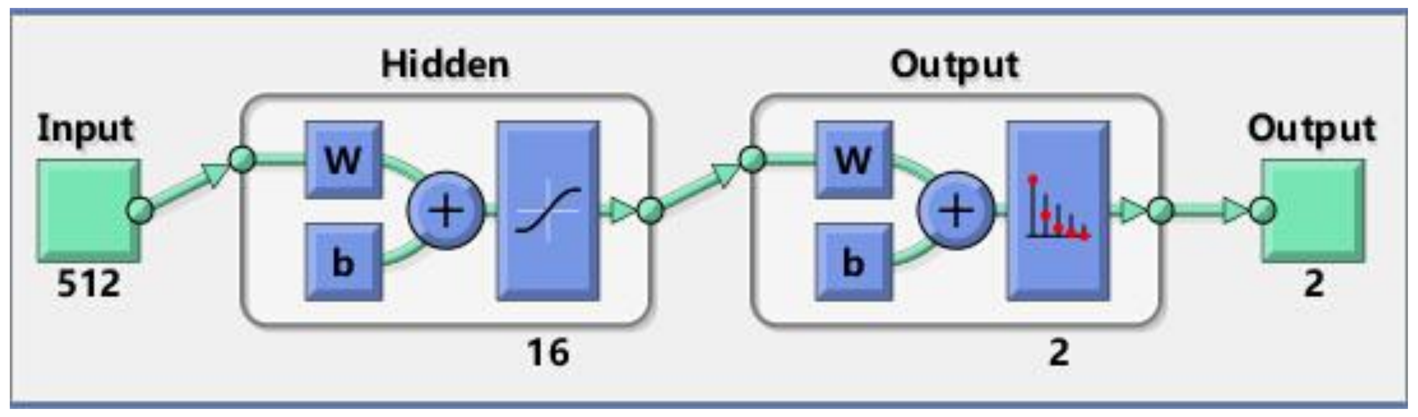
- Find subspace filter to reduce reflections and ambient sound.

$$W(\omega)=\Lambda_s^{-1/2}E_s^H$$

- Find ICA filter matrix, $U(\omega)$, for all sampled frequencies.
- Separation filter: $B(\omega)=U(\omega)W(\omega)$
- Scaling filter: $\tilde{B}_m^{+}(\omega)=diag[B_{m,1}^{+},\dots,B_{m,D}^{+}]$
- Permutation filter: $F(P)=\frac{1}{D}\sum_{n=1}^{D}\cos\theta_n$
- Final filter: $F(\omega)=P(\omega)\tilde{B}_m^{+}(\omega)B(\omega)$

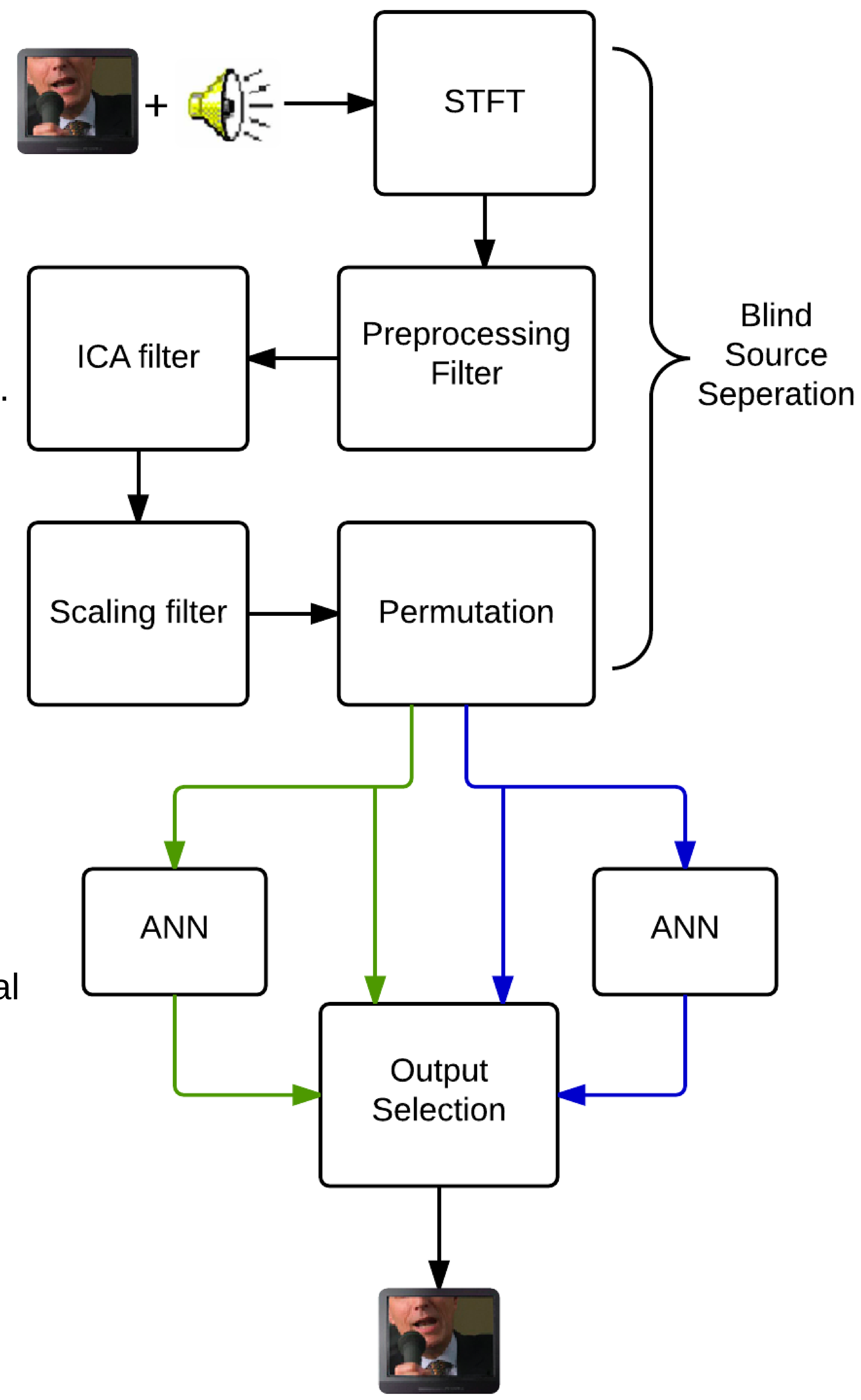
Neural Network

- We used a three-layer neural network to distinguish between human speech and background noise.
- More specifically, our neural network consists of an input layer, a hidden layer, and an output layer and uses the backpropagation algorithm and gradient descent to minimize the cost function.



System Block Diagram

- BSS separates input signal into two components.
- Neural network places a weight on each output based on how much human speech they contain.
- The system outputs the signal with the highest weight above a threshold.



Results

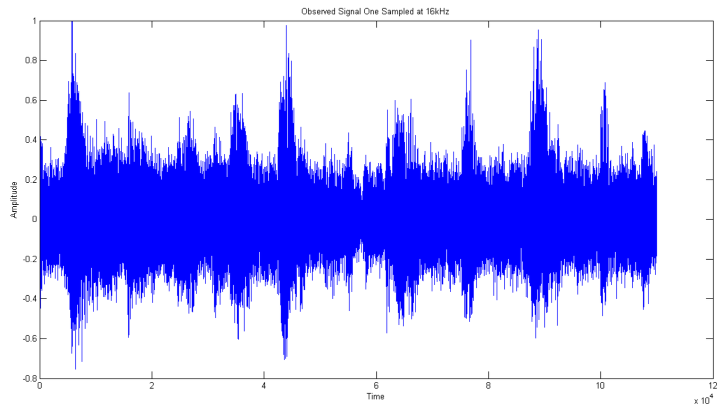


Figure 1: Mixed signal observed at 1st microphone

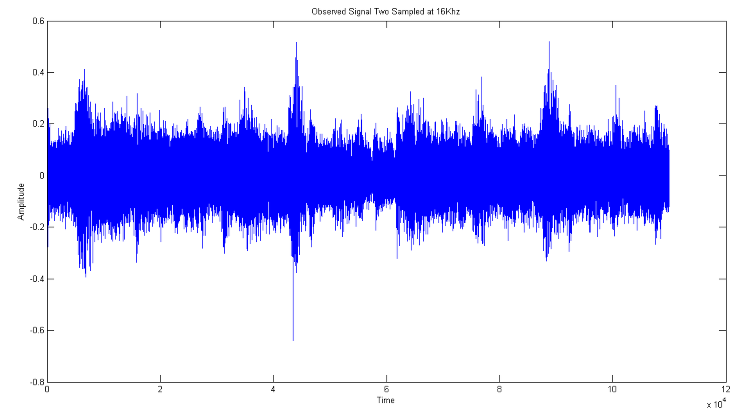


Figure 2: Mixed signal observed at 2nd microphone

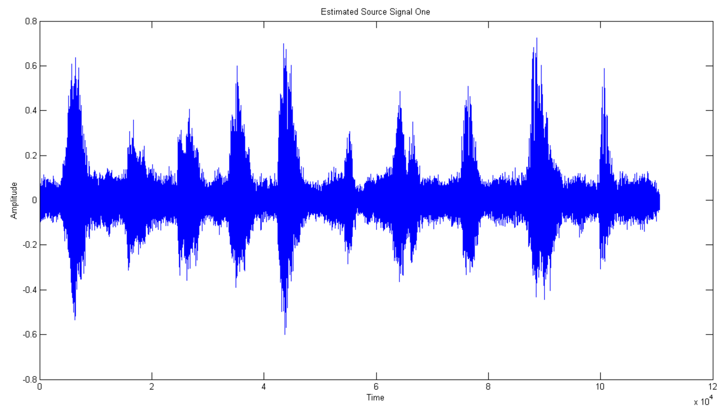


Figure 3: Estimate source signal one – human speech

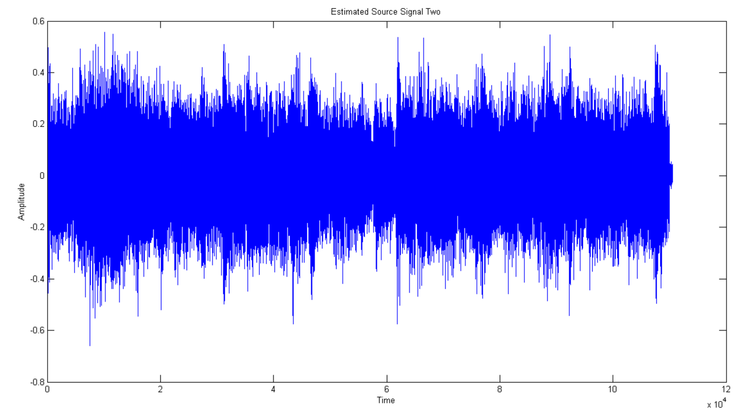


Figure 4: Estimated source signal two – music

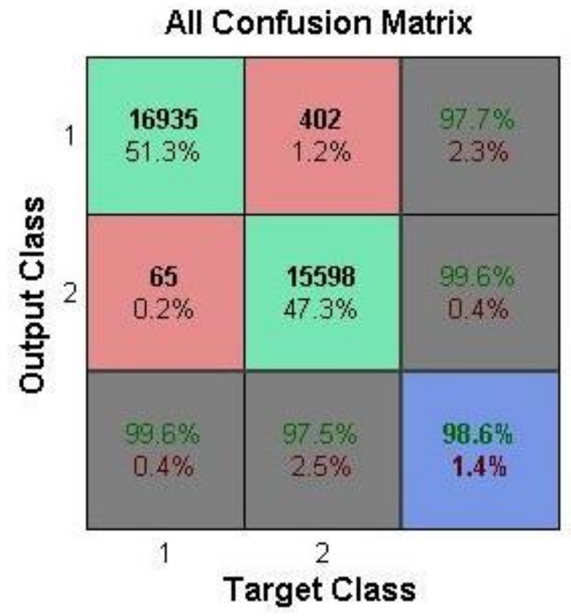


Figure 5: Accuracy of neural network

	Parameters
Sampling Rate	16 kHz
Length of STFT	512
Number of FFT Points	4096
Overlap	492
Number of inputs	2
Permutation Reference Range	5

Figure 6: Input parameters into the system.

Conclusion and Future Steps

- Figures 3 and 4 show that our BSS method is able to separate an input signal into two independent and distinct sources.
- Our neural network can perform binary classification with 98.6% accuracy.
- Implementation in real-time.
- Explore possibility to separate sources from more than two inputs.

References and Acknowledgements

- Hyvärinen, Aapo, and Erkki Oja. "Independent component analysis: algorithms and applications." *Neural networks* 13.4 (2000): 411-430.
- Asano, Futoshi, et al. "Combined approach of array processing and independent component analysis for blind separation of acoustic signals."*Speech and Audio Processing, IEEE Transactions on* 11.3 (2003): 204-215.

- Special thanks to Professor Richard Baraniuk and Dr. Mohammad Golbabaee.

For questions, please contact:
Zichao.Wang@rice.edu; Tianyi.Yao@rice.edu; Stephen.Xia@rice.edu; Shengjia.Zhao@rice.edu