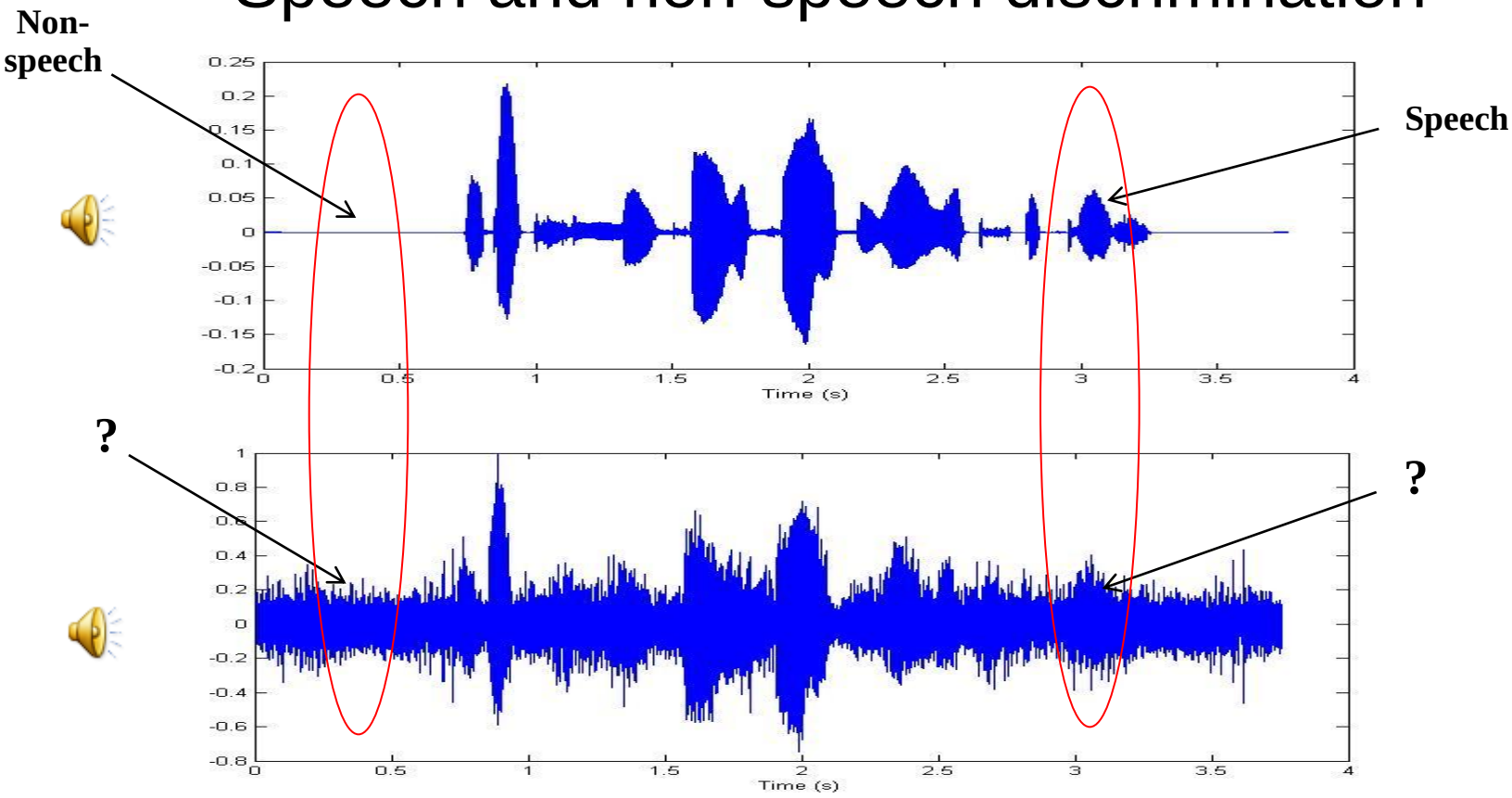
BIEN 2011

Voice Activity Detection

Jonathan Kola
Dr. Tarun Pruthi
Dr. Carol Espy-Wilson

What Is It?

Speech and non-speech discrimination



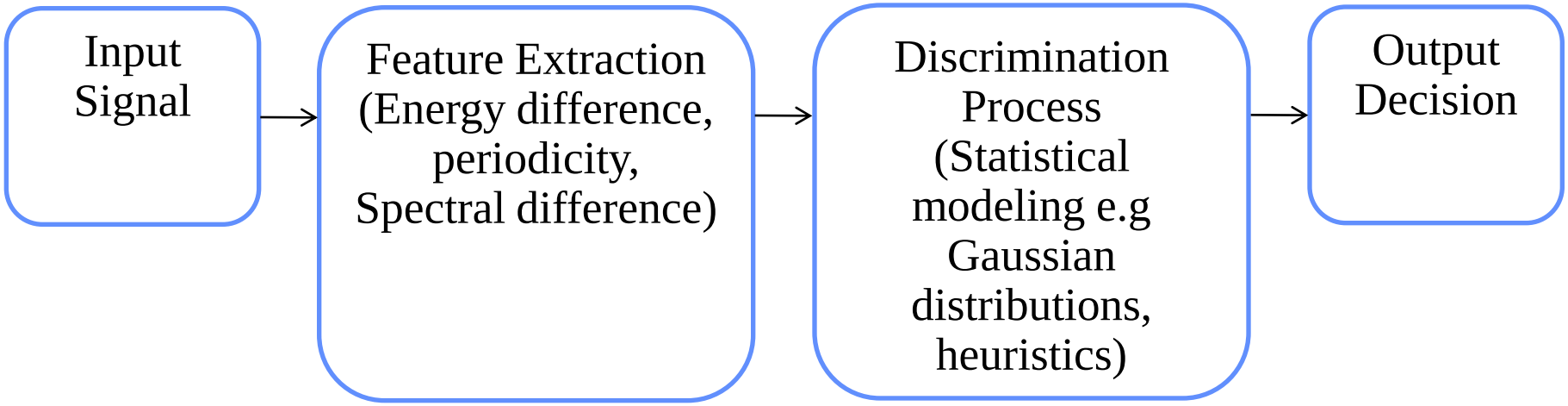
Applications

Speech enhancement

Speech coding

Speaker recognition

Detection Process



My Research

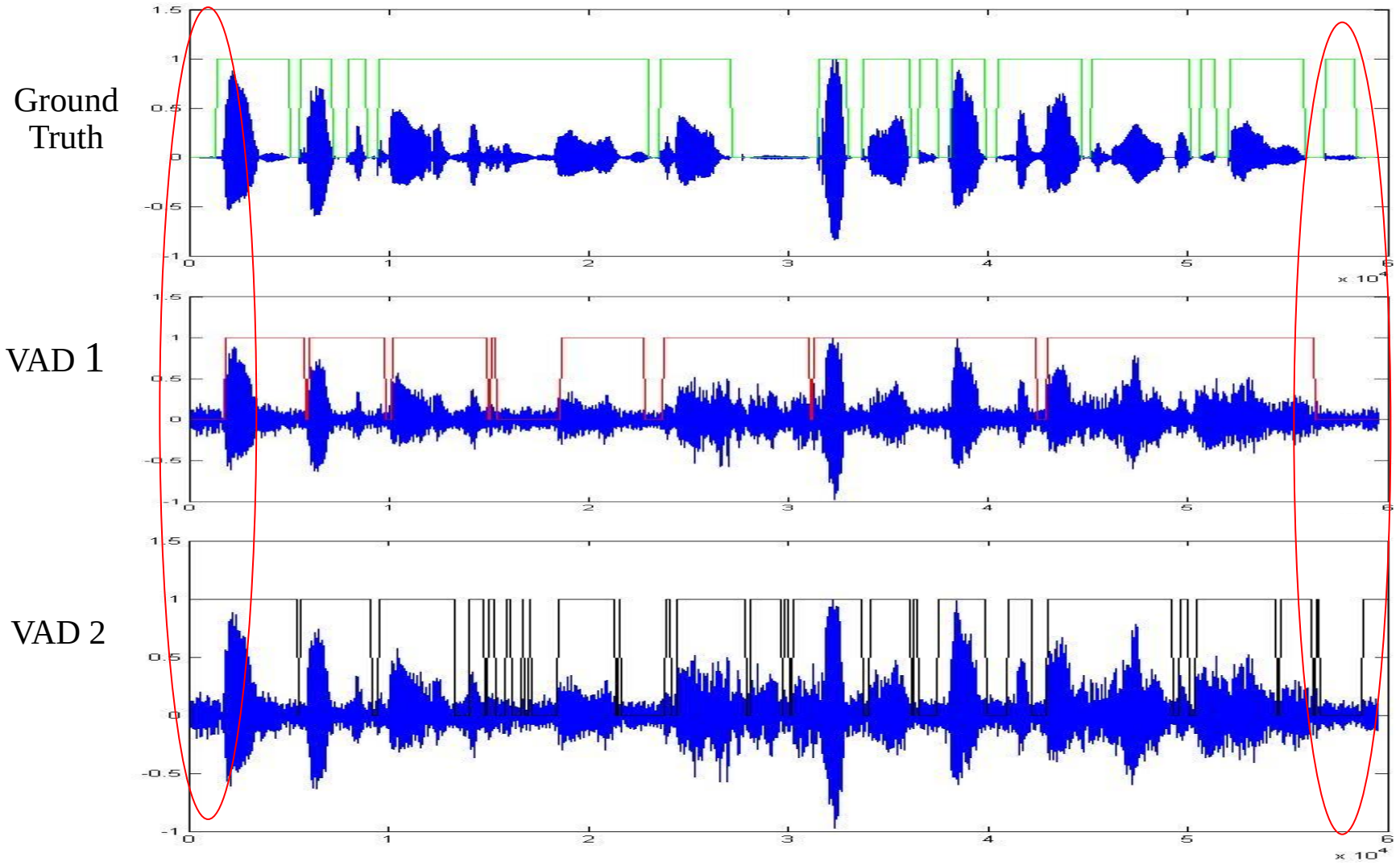
Tested 5 different VAD algorithms

- Used a combination of features including spectral distribution differences, zero-crossing rates and periodicity
- Used either a heuristic model or a statistical model to make a binary decision

Characteristics of a good VAD

- Good decision rule
- Adaptability to background noise
- Low computational complexity

VAD Output

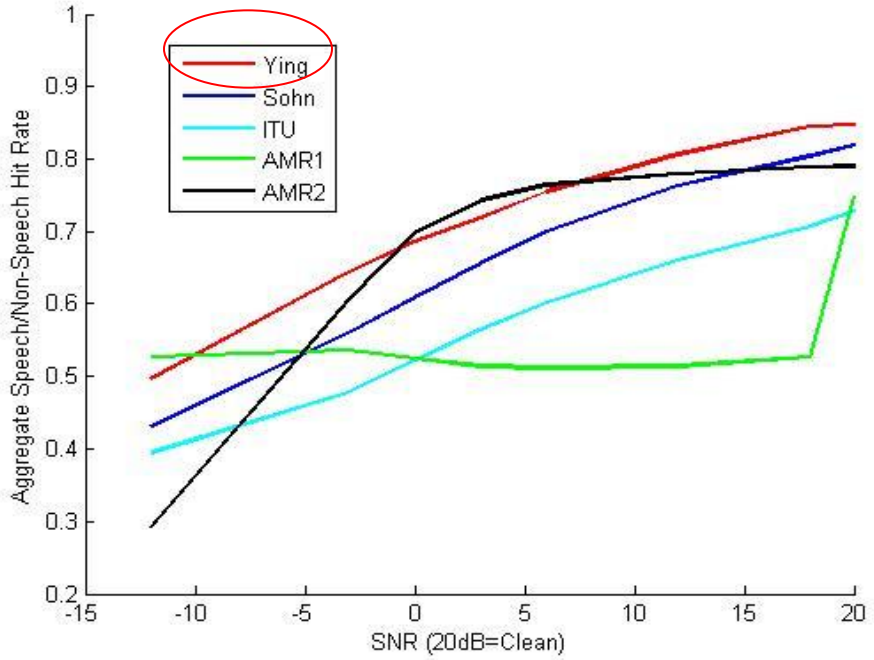


Results

The best VAD (Ying) used:

- Spectral distribution of energy
- Gaussian Mixture Model to describe energy distribution, trained in an unsupervised manner

VAD aggregate hit rates



Speech/non-speech hit rate variance due to different noise types

	Ying	Sohn	AMR 2	ITU	AMR 1
-12 dB	14.1	36.9	51.1	7.8	0.5
-3 dB	5.2	23.0	13.8	3.5	0.9
0 dB	4.0	16.4	5.4	1.7	0.3
3 dB	3.8	9.9	2.4	0.85	0.08
6 dB	2.6	5.9	0.9	0.82	0.0083
12 dB	1.7	3.4	0.1	1.9	0.099
18 dB	0.6	2.5	0.04	3.0	1.6
Average (from 0 – 18dB)	2.54	7.62	1.77	1.65	0.42

Conclusion

Robust Voice Activity Detection can be achieved based on:

- Single parameter (energy based measurements)
- Gaussian modeling of spectral energy distribution
- Unsupervised modeling techniques

Acknowledgements

- Dongwen Ying
- National Science Foundation OCI award #1063035