Automatic Volume Leveler for Real Time Speech Applications

Justin Bare, Dr. Carol Espy-Wilson, Dr. Tarun Pruthi

UNIVERSITY OF MARYLAND 1856
MERIT FAIR
BIEN 2011

The Institute for Systems Research

Introduction

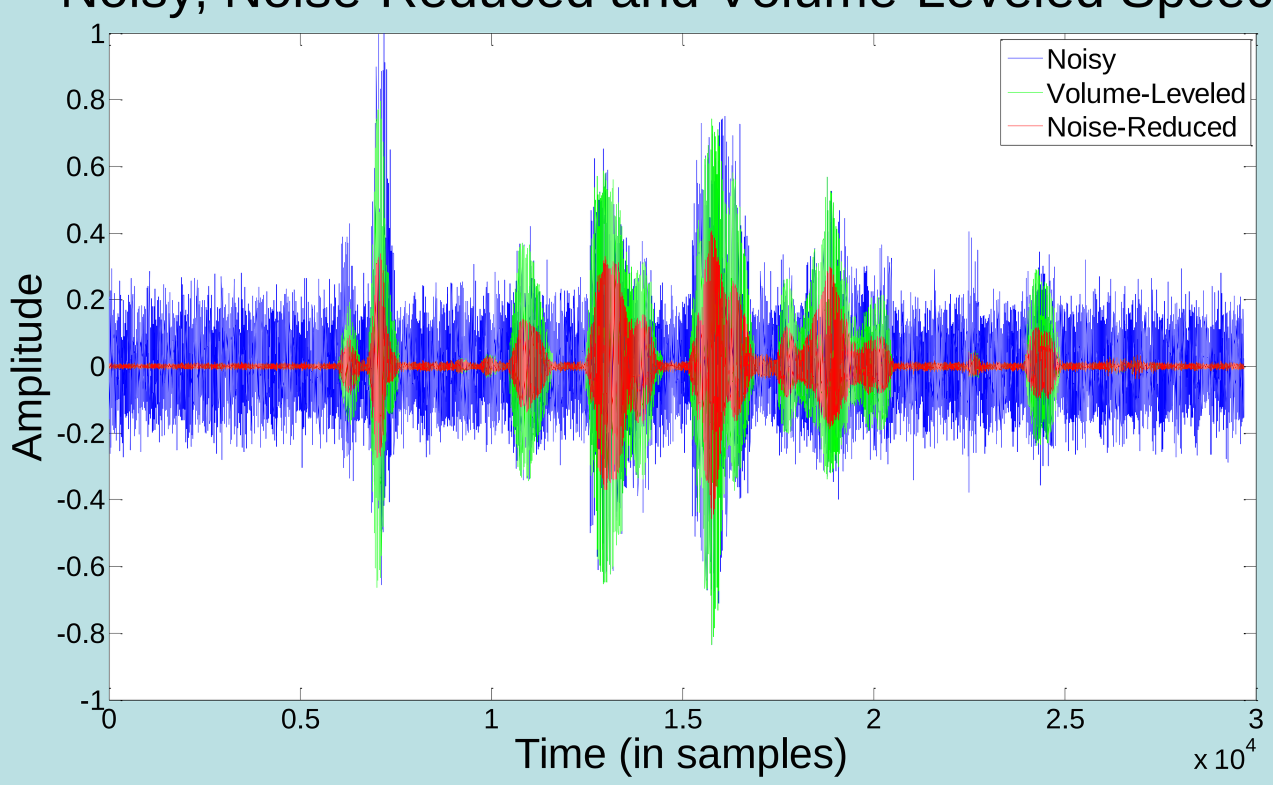
• Undesirable side effects of digital processing of speech signals are:
- •Amplification/attenuation of the target speech
- •Changes in coloring and loudness

•The goal for this project is to develop an automatic volume leveling algorithm which will, under real time constraints,
- •Fix any incorrect attenuation/amplification of the speech signal
- •Ensure there is no clipping in the scaled signal

Motivation

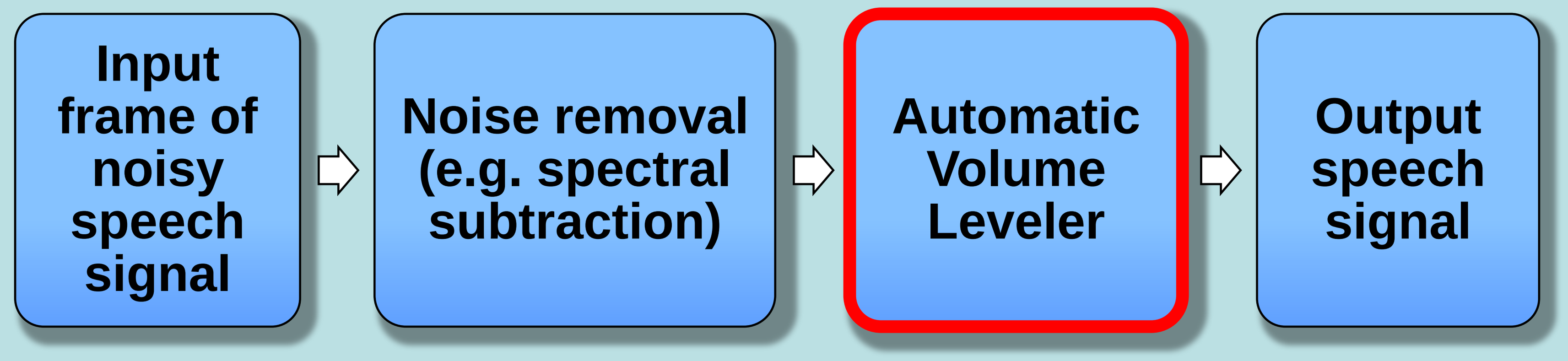
•The type of processing algorithms focused on for this project were noise reduction algorithms
- •While attenuating the noise, these algorithms often attenuate speech as well

Noisy, Noise-Reduced and Volume-Leveled Speech



•Challenges of real time implementations
- •Algorithms must be fast
- •Processing is performed on small time periods (frames) of digital speech data
- •Each frame is processed individually as it is obtained, so only current and past frames are available to the algorithm

System Overview

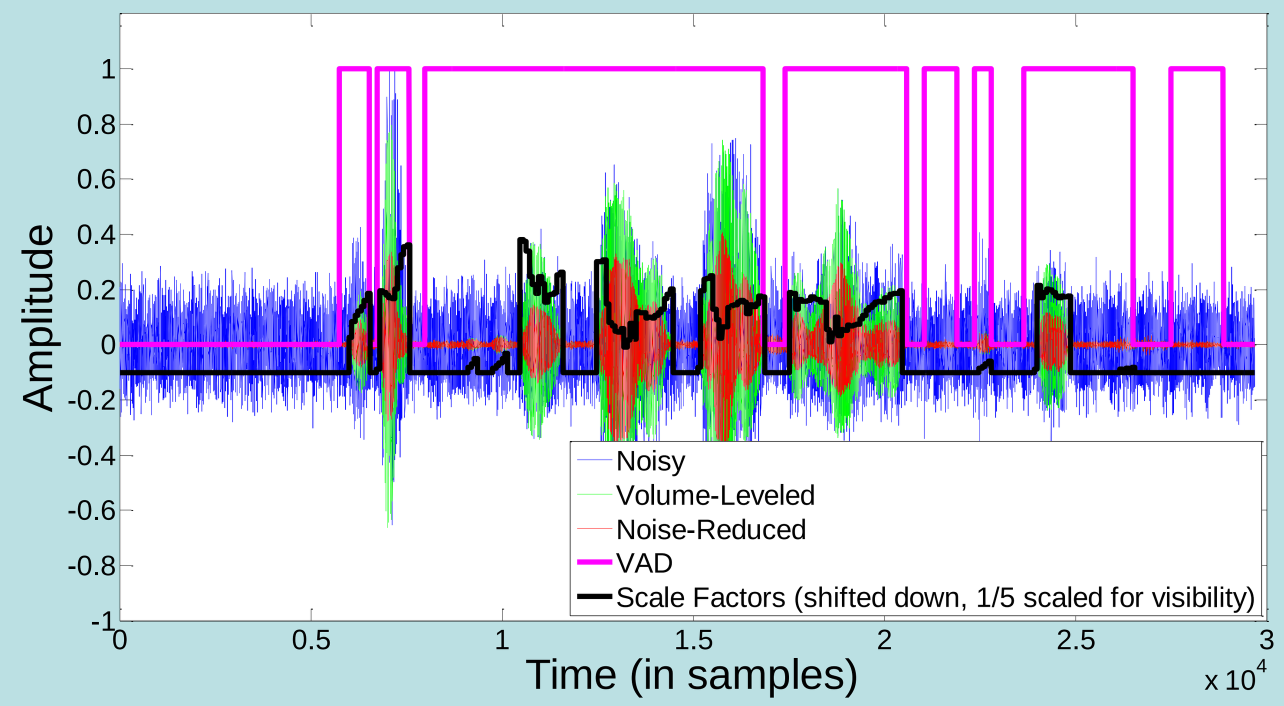


Key Aspects of the Volume Leveler

•Tracks the noise level when the Voice Activity Detector(VAD) does not detect speech, so pure noise regions are not amplified
•Uses a smoothing constant ($\alpha = 0.9$) to mitigate large differences in the multiplying factor (K(i)) between frames:

$$R(i) = mean(noisy(i)) / mean(proc(i))$$
$$K(i) = \alpha * K(i-1) + (1-\alpha) * R(i)$$

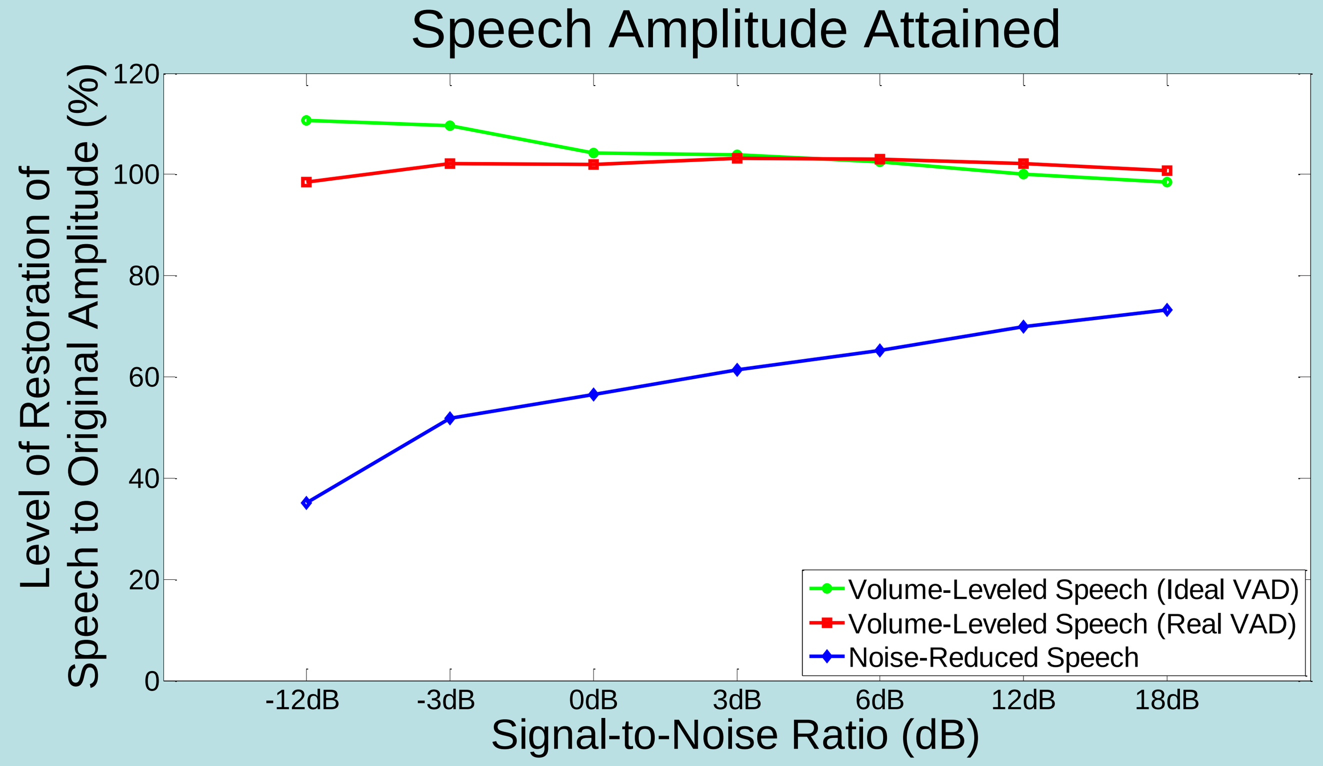
Noisy, Volume-Leveled, and Noise-Reduced Speech with Ideal VAD Decision and Scale Factors



•Finds regions of quiet speech using the noise estimate and scales these regions using $\alpha = 0.99$ to avoid noise amplification for low SNR's
•Identifies if the high amplitudes of the frame will be clipped after scaling and reduces the scale factors accordingly

Results

Percent of Original Speech Amplitude Attained

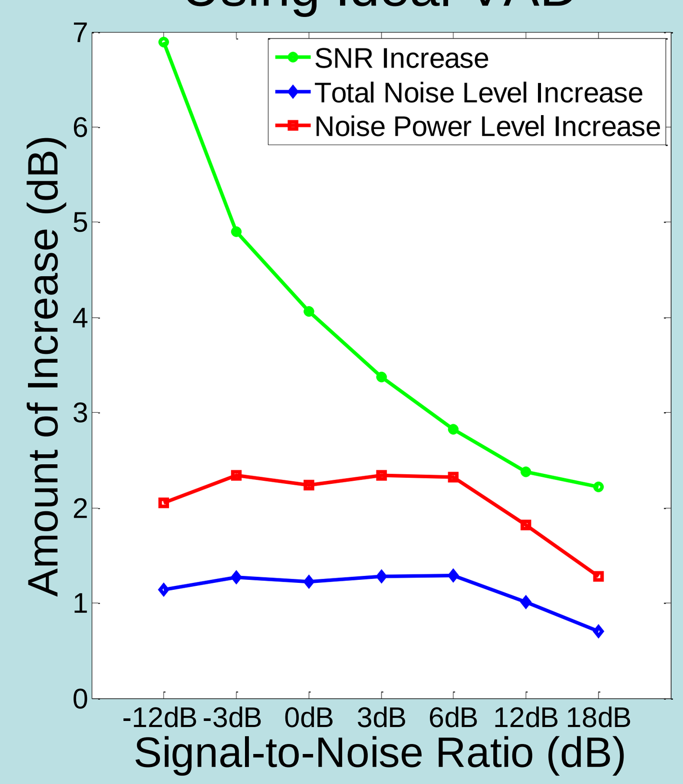


•For all SNR's, the volume in speech regions was restored, on average, to about 100% of the volume in the original signal

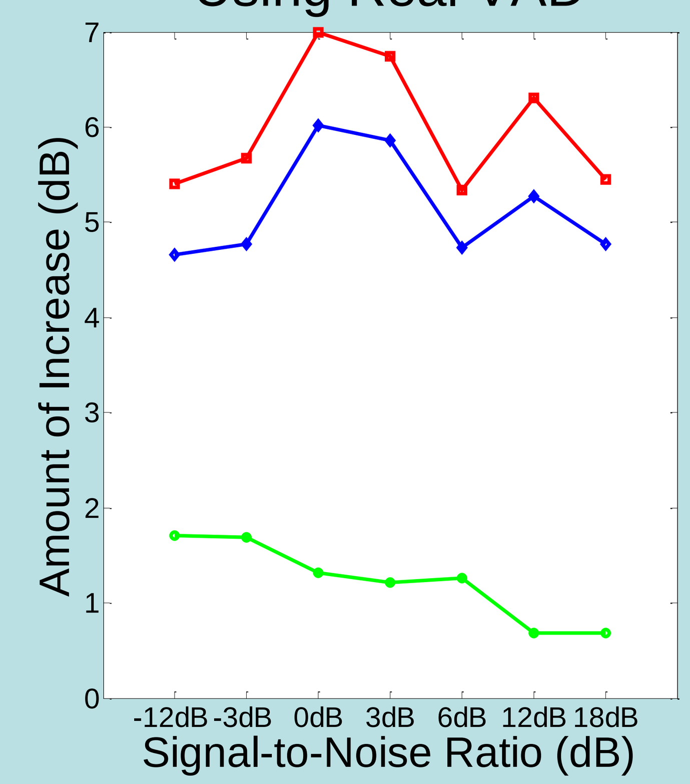
Results (continued)

SNR, TNL, and NPL Increase for Volume-Leveled Signal as Compared to Noise-Reduced Signal

Using Ideal VAD

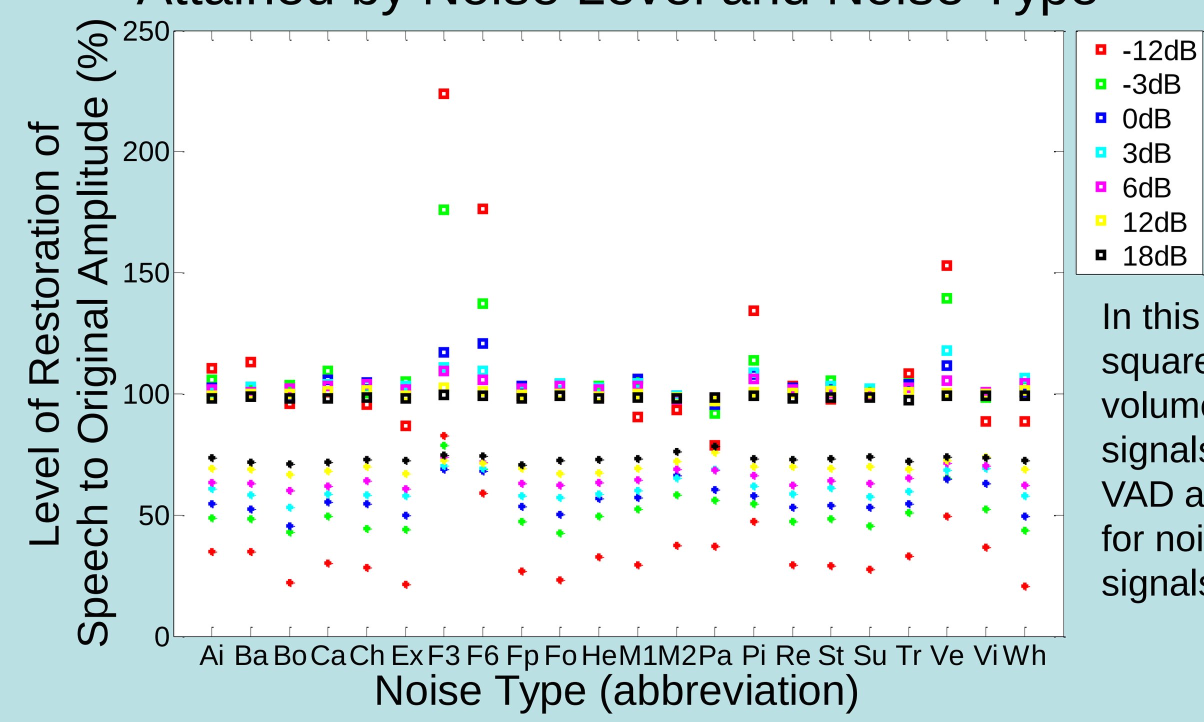


Using Real VAD



•Using an ideal VAD, the volume-leveled signal attains a greater SNR increase than noise level increase
•Unfortunately, using a real VAD, the noise level increased more than the SNR

Percent of Original Speech Amplitude Attained by Noise Level and Noise Type



In this figure, squares are for volume-leveled signals using ideal VAD and dots are for noise-reduced signals

Conclusions and Future Work

•Demonstrated an algorithm to restore speech volume and correct attenuation/amplification problems that have been introduced during processing
•Future work:
- •Find a way to remove VAD from the algorithm
- •Implement the system for use in C programming
- •Restore coloring of speech in addition to volume

Acknowledgments

•Thanks to the National Science Foundation for supporting this work with OCI award #1063035