Automatic Volume Leveler for Real Time Speech Applications

Justin Bare
Dr. Carol Espy-Wilson
Dr. Tarun Pruthi
The Volume Leveler

- Noise Reduction
- Volume Leveling

Noisy → Noise Reduced → Volume Leveled
• Noise reduction is useful
• But it has problems:
  – Attenuation of speech
  – Changes in color or timbre and loudness
Overview

- My goals for the automatic volume leveler
  - Fix speech attenuation
  - Restore loudness
  - Ensure no clipping occurs in output signal
  - Do all of the above under real-time constraints
    - Only present and past frames of signal are available

- For each frame (0.01 seconds) of signal:
Algorithm

- **Input noise-reduced signal frame**
  - **Voice Activity Detector (VAD)**
    - Frame has no speech
      - Update noise tracking
      - Do not amplify
    - Frame has speech
      - Estimate the Signal-to-Noise Ratio (SNR)
        - High SNR: Amplify to original level
        - Middle SNR: Amplify conservatively
        - Low SNR: Do not amplify

Prevent clipping
Ideal vs. Real VAD

- Ideal VAD (not obtained in real time)
  - Clean Signal
  - Phonetic Transcription
    | Time | Sound |
    |------|-------|
    | 1.78 | ‘f’   |
    | 1.9  | ‘aa’  |
    | 2.02 | ‘r’   |
  - VAD Decision

- Real VAD (obtained in real time)
  - Noisy Signal
  - VAD Algorithm
Ideal VAD vs. Real VAD: SNR and Noise Level Increase

Using Ideal VAD

Using Real VAD
Volume Restoration

Percent of Original Speech Amplitude Attained

Level of Restoration of Speech to Original Amplitude (%)

Signal-to-Noise Ratio (dB)

-12dB  -3dB  0dB  3dB  6dB  12dB  18dB

Volume-Leveled Speech (Ideal VAD)
Volume-Leveled Speech (Real VAD)
Noise-Reduced Speech
Future Work

• Less reliance on VAD accuracy
• Incorporate coloring/timbre restoration
• Implement in a fast, low level language such as C

Acknowledgments

• National Science Foundation OCI award #1063035
• Dr. Carol Espy-Wilson
• Dr. Tarun Pruthi
• MERIT BIEN faculty and staff