An Alternative Time-Frequency Representation for Speech Signals

Amplitude Modulation Spectrograms

Bryan Labelle and John M. Allen
Univ. of IL, Beckman Inst., Urbana IL

Relevance of AM to Speech Perception

- Amplitude Modulation (AM) is directly related to speech intelligibility (French and Steinberg, Plomp, Houtgast, Fletcher, Mathes and Miller).
- AM in Auditory Nerve responses is important for speech perception (Carlyon/Shamma, Dau).
- Modulation frequencies can be directly related to speech intelligibility (Hrulman).
- A small number of amplitude modulated noise carriers is sufficient for modest intelligibility in quiet conditions (Shannon).
- Frequency modulation (FM) provides significant improvement in intelligibility (Zeng).

The Auditory Filterbank

- Pure Tones
- Noise
- Speech

Auditory Filterbank Based Spectrogram

Stage (ii) Output at Threshold

Stage (vii) at Threshold

Stage (vii) Output the 5 Hz

Stage (vii) Output for 7 Hz

Stage (vii) Output for 10 Hz

Stage (vi) Output

Conclusions

- Robust features which support observed human confusion data have not been found yet.
- These Refinedness need to be made
  - Tune thresholds to match human performance for a wider variety of stimuli.
  - Obtain knowledge of the threshold for each sound, rather than an average over all sounds.
  - Align and scale the speech sounds for comparison.
  - Label the consonant, vowel, and transition regions.
- After the auditory filterbank to have the same response as human neural tuning.
- Perform a statistical analysis of the modulations.

Modulation Filterbank

Goal: Match human modulation detection thresholds.
- Tones
- Noise
- Speech

Frequency Modulation Detection

FM to AM conversion at (vii) for a 7 Hz FM modulated tone.
- A properly chosen FM and AM tone have a very similar percept.
- Demo.

FM to AM conversion for a 7 Hz FM modulated tone.

An Alternative Time-Frequency Tones

Auditory Filterbank Based Spectrogram

Stage (vii) Output the 5 Hz

Stage (vii) Output for 7 Hz

Stage (vii) Output for 10 Hz

Stage (vi) Output

Conclusions

- Robust features which support observed human confusion data have not been found yet.
- These Refinedness need to be made
  - Tune thresholds to match human performance for a wider variety of stimuli.
  - Obtain knowledge of the threshold for each sound, rather than an average over all sounds.
  - Align and scale the speech sounds for comparison.
  - Label the consonant, vowel, and transition regions.
- After the auditory filterbank to have the same response as human neural tuning.
- Perform a statistical analysis of the modulations.

Modulation Detection

stage (i) input components, a tone modulated at 7 Hz in noise.
- stage (vi) EM algorithm output, labeling signal and noise.
- stage (viii) Modulation filterbank (MFB) output.

/modulation Discrimination

Hypothesis: “events” used by HSR for discrimination of speech sounds will be visible in the modulation spectrum.
- How do we separate /m/ and /n/ from the other consonants even at -20 dB SNR?
- What time/frequency structure supports the separation of /m/ and /n/ between -20 dB and -16 dB SNR?

Critical Ratio Demo

The FM in the input signal is converted to AM by the auditory filter.
- The filter skirt slope determines the modulation sensitivity.

Speech Sound Modulations

Noise is speech weighted.
- Sounds are shown at -16 dB and -10 dB SNR because human performance in distinguishing the sounds is very poor at -10 dB but very good at -16 dB.
- Two “typical” examples of /m/ and /n/ are shown.
- Are there any features which are clearly present at -10 dB, but mostly obscured at -16 dB?
- /m/ and /n/ perception demo.

Critical Ratio Demo

The FM in the input signal is converted to AM by the auditory filter.
- The filter skirt slope determines the modulation sensitivity.