Noise-Robustness of Human Speech Recognition

The Role of Cochlear Processing

Marion Régnier\textsuperscript{1}
Sandeep Phatak\textsuperscript{2}
Feipeng Lee
Jont Allen

University of Illinois at Urbana-Champaign

\textsuperscript{1}Currently at BearingPoint, NY
\textsuperscript{2}Currently at the Army Audiology and Speech Center, WRAMC, DC
Motivation

- Recognizing speech in background noise is a big challenge for hearing-impaired (HI) listeners.

Phatak and Allen (2007)
Speech-Weighted Noise
Motivation

- Recognizing speech in background noise is a big challenge for hearing-impaired (HI) listeners.
- Amplification does not solve the noise-robustness problem (SNR-Loss).
Motivation

• Recognizing speech in background noise is a big challenge for hearing-impaired (HI) listeners.

• Amplification does not solve the noise-robustness problem (SNR-Loss).

• Normal hearing (NH) listeners can extract speech from background noise.
Motivation

- Recognizing speech in background noise is a big challenge for hearing-impaired (HI) listeners.
- Amplification does not solve the noise-robustness problem (SNR-Loss).
- Normal hearing (NH) listeners can extract speech from background noise.
- What are noise-robust features of speech?
Finding Noise-Robust Features

PHYSICAL DOMAIN → LISTENER → PERCEPTUAL DOMAIN

Acoustic Speech Signal → Percept of Speech
Spectro-temporal Analysis → Confusion Analysis
Finding Noise-Robust Features

- **Noise Masking**
  Phatak and Allen (2007); Phatak et al. (2008)

- **Time-truncation**
  Régnier and Allen (2008)

- **Filtering**
  Miller and Nicely (1955)
Psychophysics - Time Truncation
(Régnier and Allen, 2008)
• The /t/ feature is
  – short (≈5-10 ms).
  – noise-robust (0 and 12 dB SNR)
  – consistent across multiple utterances.
Psychophysics - Filtering
(Miller and Nicely, 1955)
Psychophysics and Stimuli Correlation

/t/ by talker fl13 @ 0 dB
Psychophysics and Stimuli Correlation

/t\ by talker fl13 @ 0 dB
Psychophysics and Stimuli Correlation

Cooper *et al.* (1952); Stevens (1980)

/t/ by talker fl13 @ 0 dB
Psychophysics and Stimuli Correlation

Cooper et al. (1952); Stevens (1980)

Is it the most noise-robust cue?

/tə/ by talker fl13 @ 0 dB
Psychophysics and Stimuli Correlation

The 4-step Method (Régnier and Allen, 2008)

Step 2: AI-gram of m112 /te/ at 0 dB SNR

Step 4: Event-gram of m112 /te/ at $t = 26.25$ cs

Step 3: Short-time AI above 2 kHz for m112 /te/ at 0 dB SNR

Step 1: Confusion patterns for /te/ by m112

$P(C\text{ heard}|\text{spoken}=/Tay/)$

$\text{Distance along BM [mm]}$
Psychophysics and Stimuli Correlation

The 4-step Method (Régnier and Allen, 2008)

Step 1: Confusion patterns for /te/ by m112

Step 2: AI−gram of m112 /te/ at 0 dB SNR

Step 3: Short−time AI above 2 kHz for m112 /te/ at 0 dB SNR

Step 4: Event−gram of m112 /te/ at \( t^* = 26.25 \) cs
Utterance Variability
(Régnier and Allen, 2008)

Step 4: Event−gram of m112 /te/ at $t^* = 26.25$ cs

Step 1: Confusion patterns for /te/ by m112

Step 4: Event−gram of m117 /te/ at $t^*$

Step 1: Confusion patterns for /te/ by m117
Utterance Variability
(Régnier and Allen, 2008)

Step 4: Event-gram of m112 /te/ at $t$

Step 4: Event-gram of m117 /te/ at $t$

Step 1: Confusion patterns for /te/ by m112

Step 1: Confusion patterns for /te/ by m117
Effect of Noise Spectrum
(Régnier and Allen, 2008)

Event-gram in SWN at $t^* = 15$ cs, BW=570, $T=0.335$

Confusion patterns for f106ta in SWN
Effect of Noise Spectrum
(Régnier and Allen, 2008)

Event-gram in SWN at $t^* = 15$ cs, BW=570, T=0.335

Event-gram in WN at $t^* = 15$ cs, BW=450, T=0.125

Confusion patterns for f106ta in SWN

Confusion patterns for f106ta in WN
Psychophysics and Stimuli Correlation

Correlation between perceptual and physical domains

$SNR_e$ vs $SNR_{90}$

- SWN, BW=570 Hz, T=0.335
- WN, BW=450 Hz, T=0.125

(Régnier and Allen, 2008)
Role of Onsets in Noise-Robust Consonant Identification

DEMO
Onsets and Noise-Robustness

- Across-frequency onsets integration characterizes
  - stop release bursts and frication onsets.
  - formant onsets and formant transitions.
  - voice-onset time.
Onsets and Noise-Robustness

- Across-frequency onset integration characterizes
  - stop release bursts and frication onsets.
  - formant onsets and formant transitions.
  - voice-onset time.

- Onsets are audible at very low SNRs.

- Speech onsets are important not only for speech recognition but also for auditory streaming (Darwin and Ciocca, 1992) and source localization (Litovsky et al., 1999).
Onsets and Noise-Robustness

- Across-frequency onset integration characterizes
  - stop release bursts and frication onsets.
  - formant onsets and formant transitions.
  - voice-onset time.

- Onsets are audible at very low SNRs.

- Speech onsets are important not only for speech recognition but also for auditory streaming (Darwin and Ciocca, 1992) and source localization (Litovsky et al., 1999).

- These onsets are neurally coded with very high accuracy.
  - auditory nerve (Loebach and Wickesberg, 2006).
  - primary auditory cortex (Heil, 2003).
Role of Cochlear processing

Fahey and Allen (1985)
Role of Cochlear processing

- Delay in cochlear filter adaptation.
Role of Cochlear processing

0−15 ms: low thresholds, narrow tuning
75−200 ms: high thresholds, wide tuning

Delgutte (1980); Fahey and Allen (1985)

• Delay in cochlear filter adaptation ⇒ Onset enhancement.
Role of Cochlear processing

Delgutte (1980); Fahey and Allen (1985)

- Delay in cochlear filter adaptation \(\Rightarrow\) Onset enhancement.

- Onset enhancement missing for HI listeners with wider cochlear filters.
Role of Cochlear processing

Delgutte (1980); Fahey and Allen (1985)

- Delay in cochlear filter adaptation $\Rightarrow$ Onset enhancement.

- Onset enhancement missing for HI listeners with wider cochlear filters.

- Enhancing onsets can improve performance in background noise (Hazan and Simpson, 1998).
Acknowledgements

• Yang-soo yoon and David Gooler, Speech and Hearing Science, UIUC.

• Etymotic Research, Starkey Labs, Phonak and Oticon Foundation.
References


