Spectro-temporal Fine Structure is Critical for Robust Neural Encoding of Speech in Noise

Nai Ding
Monita Chatterjee
Jonathan Z Simon

2/19/2013
Speech Recognition is Robust to Noise for Normal Hearing Listeners but not for Cochlear Implant Users.
Noise Susceptible Speech Recognition Can be Simulated By Noise Vocoding, Which Removes the Spectro-Temporal Fine Structure.
How Does the Spectro-temporal Fine Structure Contribute to Speech Recognition in Noise?
Hypothesis I: Purely a Problem in Speech Recognition?

Noise may affect the auditory representations of natural speech and vocoded speech similarly, but speech recognition differently.

The amount of auditory degradation (proportional to the SNR)

3 dB SNR

natural speech in quiet

vocoded speech in quiet

vocoded speech

3 dB SNR
Hypothesis II: A Problem in Speech Noise Segregation?

Speech-Noise Mixture → Auditory Streaming → Representation of Speech → Speech Recognition

- Natural speech
- Vocodered speech

The amount of auditory degradation (NOT proportional to the SNR)
We assess whether auditory streaming plays a role in maintaining the robustness of speech recognition, when the acoustic interference is stationary noise.
A noise vocoder reduces the spectral resolution and preserves the temporal envelope.

Reduced Spectral Resolution

Increased Noise

3 dB SNR

in quiet

natural

8-band

4-band

100%

94%

43%

100%

40%

4%

2 seconds

frequency (kHz)
Neural responses are recorded using MEG, while normal hearing listeners listen to a story, either in quiet or in noise, either noise vocoded or not.
The MEG Responses Follows the Slow Temporal Modulations of Speech

Response Phase-Locking Spectrum

Phase-locked MEG Response is Observed below 10 Hz.

Ding & Simon, J Neurophys 2012
For natural speech, the low-frequency neural response is robust to noise at 3 dB SNR.
Response Phase Locking Spectrum: the Effect of Noise

For 8-channel vocoded speech, the neural response is weakened by noise.
Response Phase Locking Spectrum: the Effect of Noise

For 4-channel vocoded speech, the neural response is severely weakened by noise.
Response Phase Locking Spectrum: the Effect of Vocoding

in quiet

3 dB SNR
- natural
- 8-band
- 4-band

inter-trial correlation

frequency (Hz)
Response Phase Locking Spectrum: the Effect of Vocoding

The response spectrum has a lower cut-off frequency for speech with lower frequency resolution.
Neural Phase-Locking is Correlated with Individual Intelligibility Score

4-channel, in quiet

8-channel, in quiet

8-channel, in noise

$R = 0.66 \pm 0.14$

$R = 0.55 \pm 0.14$

$R = 0.71 \pm 0.11$
Speech intelligibility is the amount of auditory degradation (NOT proportional to the SNR).

Noise severely degrades the auditory representation of vocoded speech but not natural speech.
Summary (II)

- The spectro-temporal fine structure plays a critical role in segregating speech from noise and building a robust representation of speech in auditory cortex.
Thank you!