# Temporal Processing of Vocoded and Degraded Speech in Human Auditory Cortex

Nai Ding[1], Monita Chatterjee[2], Jonathan Z. Simon[1,3]

[1]Department of Electrical & Computer Engineering, [2]Department of Speech & Hearing, [2]Department of Biology
University of Maryland College Park

Computational Sensorimotor Systems Lab

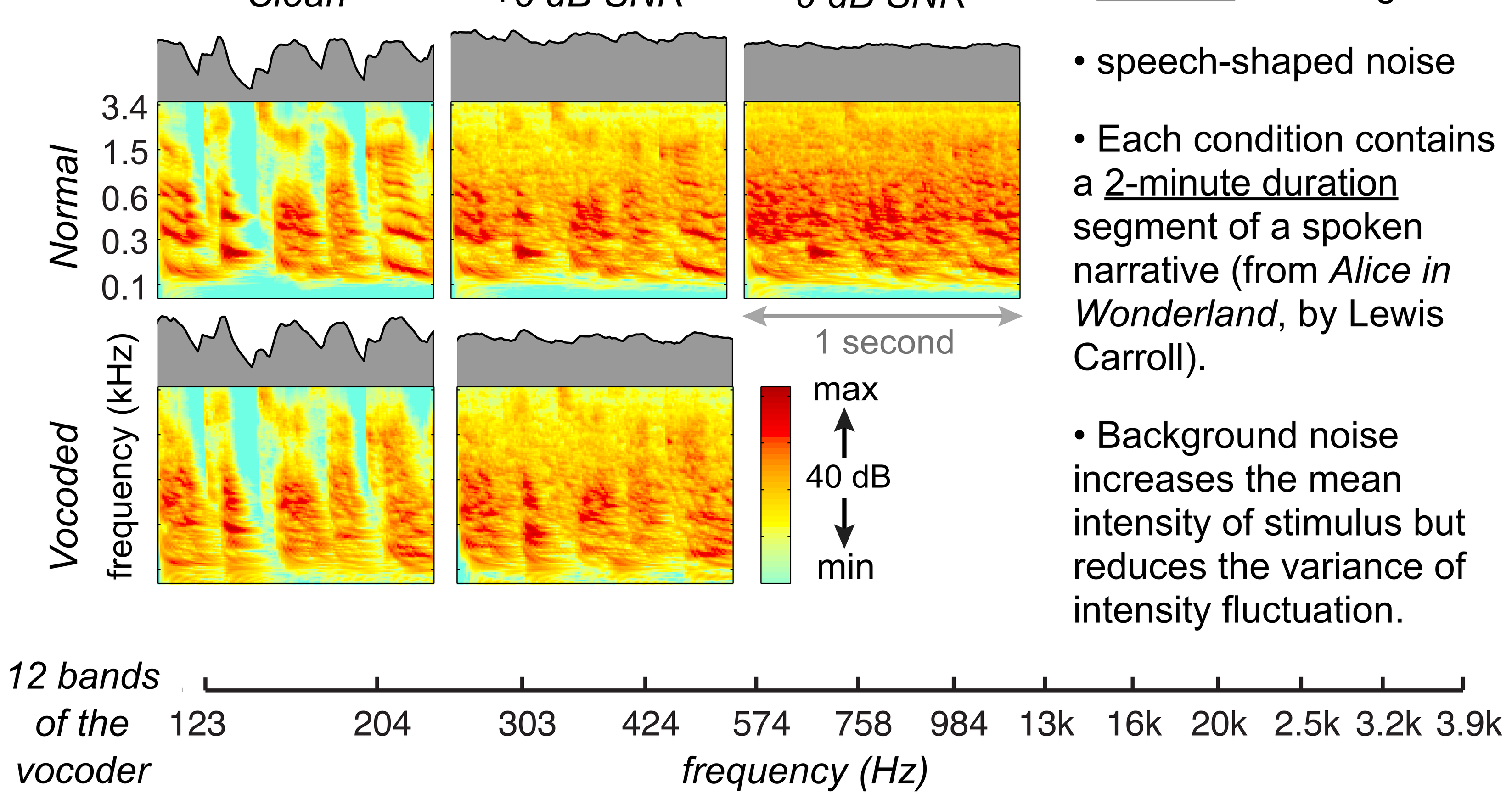
## Introduction

- **The perception of speech is robust to acoustic degradations such as noise vocoding and addition of background noise.**
- **This study addresses how the temporal neural processing of speech is influenced by acoustic degradations that significantly change the acoustics of speech but not the perception of speech.**

We recorded the cortical response using *magnetoencephalography* (**MEG**) from human subjects listening to spoken narratives. MEG is an non-invasive neural recording tool, with millisecond level time resolution. Recently, it has been demonstrated that MEG can reliably measure the cortical activity tracking the temporal modulations of speech.

## Stimuli & Data Analysis

### Stimuli




- 12-band vocoding
- speech-shaped noise
- Each condition contains a 2-minute duration segment of a spoken narrative (from *Alice in Wonderland*, by Lewis Carroll).
- Background noise increases the mean intensity of stimulus but reduces the variance of intensity fluctuation.

### Procedures & Behavioral Results

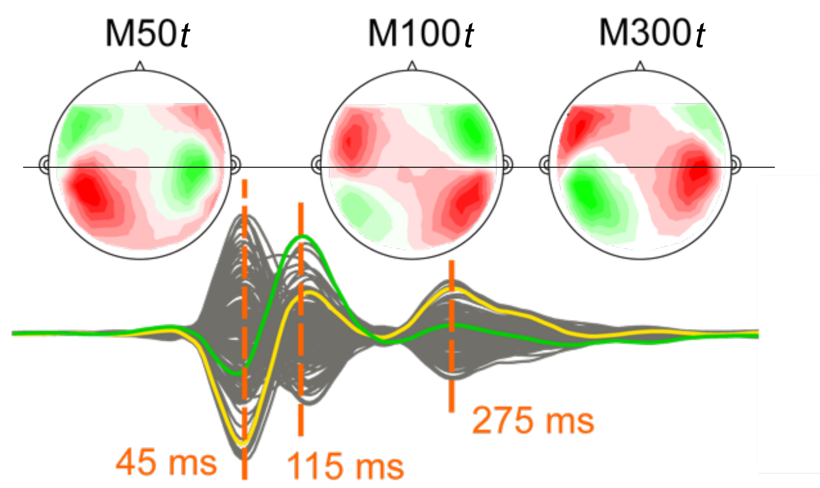
- Each stimulus was played 3 times. After every 1-minute duration stimulus, the subject was asked a **comprehension question** about the story.
- 4 normal hearing subjects participated in the experiment.
- 96% questions were answered correctly for 12-band vocoded speech at +6 dB SNR.
- 91% questions were answered correctly for the normal speech at 0 dB SNR.
- Subjectively rated speech intelligibility was 87% for both of the above two conditions.
- The intelligibility of 12-band speech is between 80% - 90%, as is measured using HINT sentences (Friesen et al. 2001).

### MEG Recording & Analysis

- 157 channel whole-head MEG system, sampled at 1 kHz, with a 60 Hz notch filter.
- The neural source of MEG response was localized using a bilateral equivalent current dipole model.
- The MEG response was filtered between 1 and 9 Hz.
- The temporal response function was estimated using boosting with 10-fold cross validation, based on a sub-cortical spectro-temporal representation of speech.

## Modeling the Cortical Response to Speech

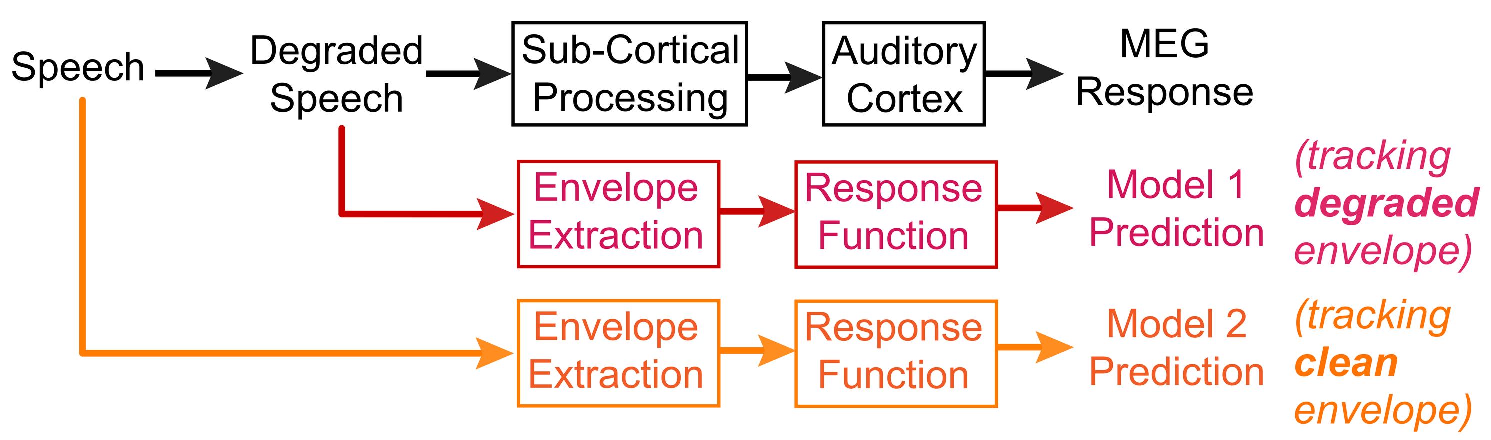
The MEG response tracks the envelope of normal clean speech, and can be modeled as a filtered version of the speech envelope. The fitted linear filter is referred to as the **temporal response function** (TRF).




The temporal response function has 3 major peaks at about 50 ms, 100 ms, and 300 ms, referred to as M50*t*, M100*t*, and M300*t*. The M100*t* has a magnetic field distribution different from the other two peaks.

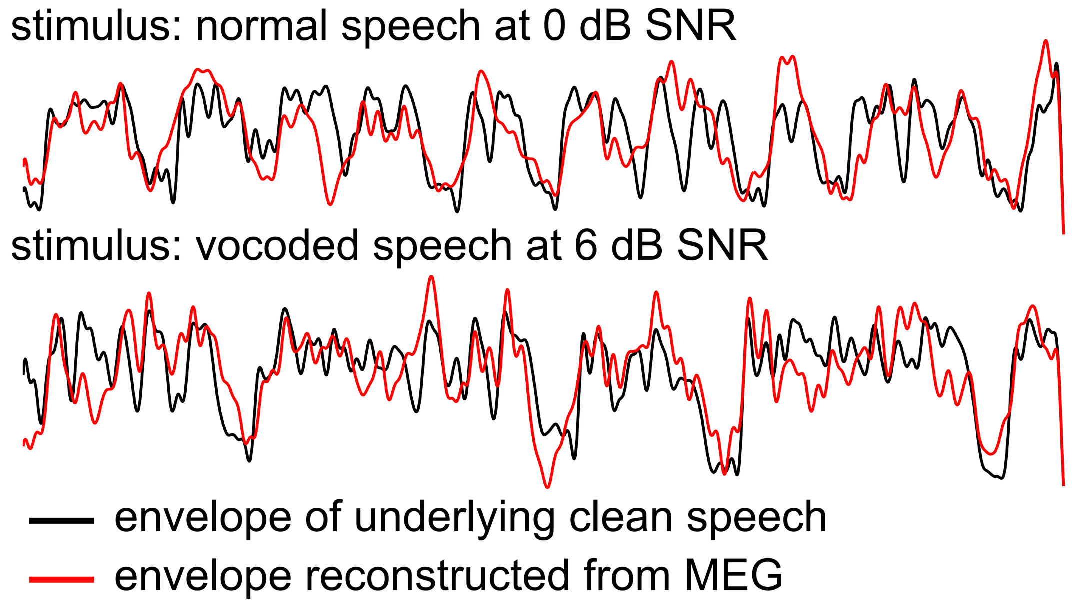
The neural sources of the 3 peaks are all localized to bilateral superior temporal gyrus.

### Two Models for Response to Degraded Speech




### Reconstructing speech envelope from MEG

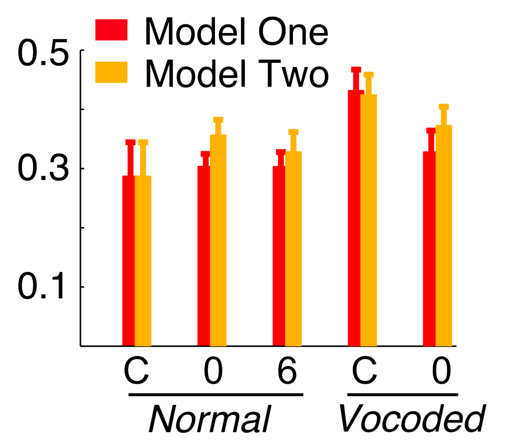
Using the same linear model as the temporal response function, we can **reconstruct** the speech envelope from the MEG response.




*grand averaged envelope reconstruction results (based on Model Two)*

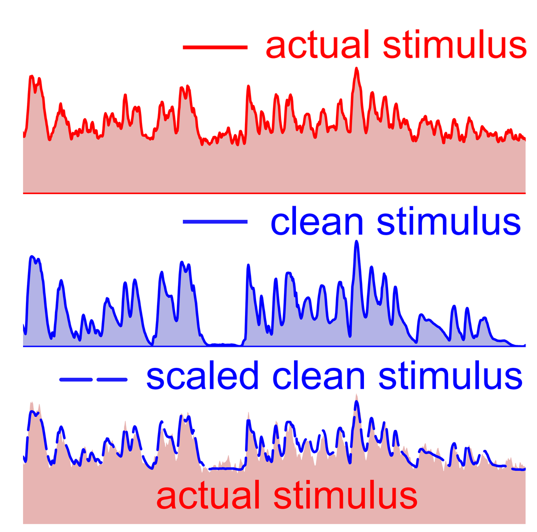
*(5 second duration, filtered between 1-9 Hz)*

*Correlation between reconstructed envelope and real envelope*




*C: clean 6: 6 dB 0: 0 dB*
*error bar: 1 s.e. over subjects*

### Discussion




The envelope of speech in 0-dB stationary noise is roughly a scaled version of the envelope of the clean speech.

The envelope of the degraded speech and the envelope of the underlying clean speech can be reconstructed with similar precision.

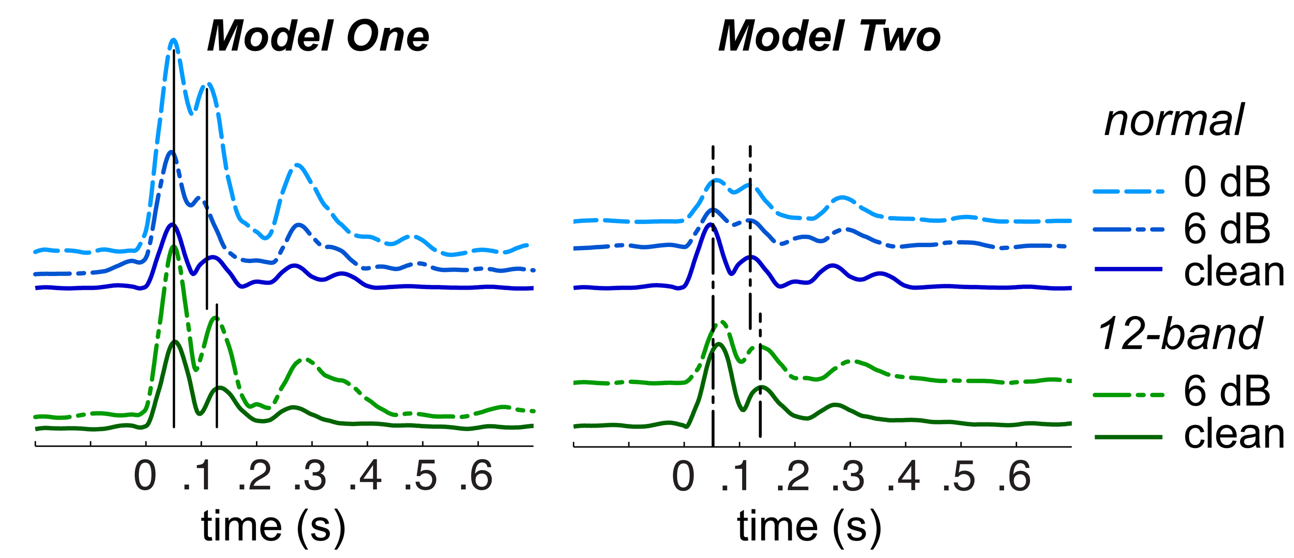
Stationary noise mainly changes the dynamic range of speech envelope rather than the shape.

✓ **The cortical neural tracking of the temporal modulations of speech is not degraded by vocoding or stationary background noise.**

## Temporal Response Function

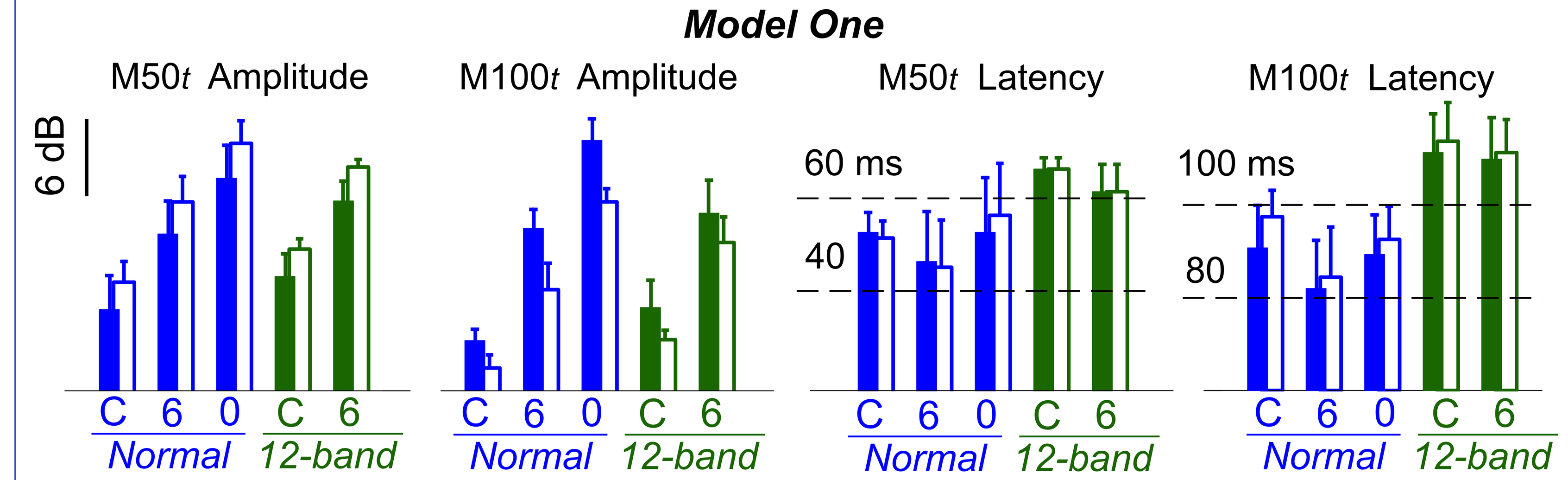
The temporal response function resembles the MEG response to sound onsets but reflects the **continuous neural tracking of temporal modulations**.

**Temporal Response Function** *(RMS over MEG sensors, grand average)*



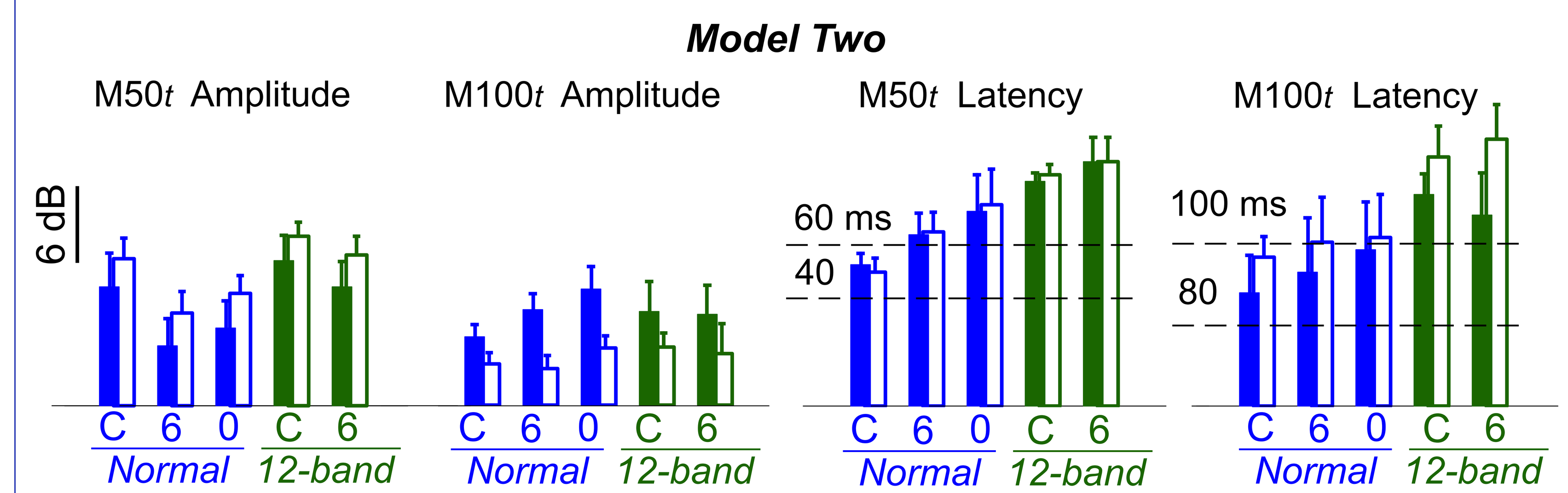

*Every condition shows the M50t, M100t, and M300t peaks. (filtered between 1-9 Hz)*

### Amplitude and Latency of M50t / M100t




*C: clean 6: 6 dB 0: 0 dB*
*Filled bars: Right Hemisphere;*
*Error Bar: 1 S.E. over subjects*
*Hollow Bars: Left Hemisphere*

✓ **The gain of neural response increases with decreasing SNR.**
✓ **The latency of neural response is longer for vocoded speech.**




✓ **Gain modulation of the cortical response effectively compensates for the reduced dynamic range of noisy speech.**

## Conclusions

✓ **The cortical representation of the temporal envelope of speech is largely independent of the spectral resolution of stimulus, even in the presence of a moderate amount of stationary noise.**

✓ **The cortical neural response adapts to the dynamic range of the intensity of noisy speech, which facilitates the robust temporal processing of speech in noise.**

✓ **It is possible that the cortical temporal processing of speech is delayed but otherwise normal in the better-performing cochlear implant listeners.**

## Reference

L.M. Friesen, R.V. Shannon, D. Baskent & X. Wang, J. Acoust. Soc. Am. (2001)
A. de Cheveigné & J.Z. Simon, *J. Neurosci. Methods* (2008)
S.V. David, N. Mesgarani & S.A. Shamma, Network: *Comput. Neural Syst.* **18** (2007)
X. Yang, K. Wang & S.A. Shamma, *IEEE Trans. Info. Theory* (1992)

**Acknowledgements**
Supported has been provided by R01-DC-008342.