Published in final edited form as:

Ear Hear. 2019 ; 40(4): 1035–1038. doi:10.1097/AUD.0000000000000674.

Compensatory and serial processing models for relating electrophysiology, speech understanding, and cognition

Curtis J. Billings^{a,b}, Garnett P. McMillan^a, Marilyn F. Dille^{a,b}, and Dawn Konrad-Martin^{a,b}

^aNational Center for Rehabilitative Auditory Research, Veterans Affairs Portland Health Care System, 3710 SW US Veterans Hospital Road, P5-NCRAR, Portland, OR 97239, USA

^bDepartment of Otolaryngology—Head & Neck Surgery, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, OR 97239, USA

Abstract

OBJECTIVE: To develop a framework for investigating the roles of neural coding and cognition in speech perception.

DESIGN: N1 and P3 auditory evoked potentials, QuickSIN speech understanding scores, and the digit symbol coding cognitive test results were used to test the accuracy of either a compensatory processing model or serial processing model.

RESULTS: The current dataset demonstrated that neither the compensatory nor the serial processing model were well supported. An additive processing model may best represent the relationships in these data.

CONCLUSIONS: With the outcome measures used in this study, it is apparent that an additive processing model, where exogenous neural coding and higher order cognition contribute independently, best describes the effects of neural coding and cognition on speech perception. Further testing with additional outcome measures and a larger number of subjects is needed to confirm and further clarify the relationships between these processing domains.

Keywords

Speech in noise; perception in noise; signal-to-noise ratio; auditory evoked potentials; event-related potentials; brain and behavior; cognition

1. Introduction

The relationship between neural coding and behavior is of great interest to both researchers and clinicians. Within human auditory sciences, the influences of neural coding and cognitive abilities on speech understanding and the subtleties of their interactions are especially important, particularly when perception difficulties in background noise are considered, because understanding-in-noise deficits may be caused by sensory changes (e.g., reduced hearing sensitivity, frequency resolution, temporal resolution, etc.) and/or cognitive deficits (e.g., slowed processing, diminished working-memory capacity, attentional control, etc.).

Some neural coding and cognitive measures have been shown to correlate with, and even predict, speech-in-noise understanding. For example, Billings et al. (2015) found strong relationships between speech perception and cortical auditory evoked potentials, Vaughan et al. (2008) found strong relationships between speech perception and cognitive tests, and Polich et al. (1983) demonstrated correlations between evoked potentials and the digit span cognitive test. Such strong correlations between some of these variables have been shown even in normal hearing populations (Billings et al., 2013; 2015). It may be that incorporating cognitive measures into a model that relates neural coding to speech-in-noise perception will improve our understanding of perception in noise difficulties. Therefore, we consider two simple models relating outcome measures from different points along the processing-to-perception continuum, namely neural coding, cognition, and speech understanding. Many researchers have attempted to understand how these and other variables may interact to affect speech-perception-in-noise abilities; for example, Anderson et al. (2013) found compelling evidence for brainstem-level neural encoding and cognition to explain significant portions of the variance in speech understanding in noise; Furthermore, top-down cognitive processing has been implicated to varying degrees in compensatory-type models of perception (Anderson et al., 2013; Erb & Obleser, 2013; Henry et al., 2017; Wong et al., 2009, 2010). Here we present our initial approach to identify mechanistic models that quantitatively predict the relationships among neural coding, cognition, and speech understanding.

We propose two contrasting models of speech understanding: (1) the compensatory model, and (2) the serial processing model. These models assume that both better neural coding and better cognition result in improved speech understanding. The difference between the models is the nature of the interaction between neural coding and cognition. In the compensatory model (Figure 1a), cognitive processes, such as working memory, make up for distorted and poor neural coding through filling in gaps in coded information; whereas, when neural coding is good the compensation benefit from good cognition is smaller as top-down processes are not needed to the same degree to extract meaning. The compensatory model corresponds to a negative interaction effect between neural coding and cognitive function such that the slope is shallower for people with good cognition relative to those with poor cognition. In contrast, the serial model (Figure 1b) would predict that poor neural coding limits the benefits that good cognition can provide such that less speech perception improvement would result when poor encoding is occurring compared to when good encoding is present. In other words, the effect of cognition on speech perception is greatest when the signal is best represented neurally because cognitive function accentuates the existing neural code. This benefit is less apparent with poorer neural coding because of the limited accurate coding information. The serial processing model corresponds to a positive interaction effect between neural coding and cognitive function (i.e., the slope increases from the poor to good cognitive groups).

2. Methods

2.1. Participants

Thirty-three individuals (31 male, 2 female) were tested as non-diabetic control participants in a larger diabetes-related research project (for recruitment details, see Konrad-Martin et al., 2015). Recruitment occurred at a Veterans Affairs Hospital, contributing to the imbalance in male and female participants. The mean pure-tone average (500, 1000, and 2000 Hz) for participants was 13.3 dB (min = 0.625, max = 27.5). Mean age was 47.7 years (min = 25, max = 66). The broad age range was recruited to allow age controls for the larger diabetes study. The Mini-Mental Status Exam (Folstein et al., 1975) was completed and participants demonstrated scores greater than or equal to 25.

2.2. Outcome Measures

For electrophysiological testing, syllables /ba/ and /da/ used previously (Bennett et al., 2012) were shortened to 100 ms (10-ms Blackman offset) and presented binaurally using ER3-A inserts in an oddball paradigm at 80:20, respectively. Signals were presented at 70 dB SPL in (1) quiet, with no noise present; (2) 10-dB SNR, with 60 dB of noise in the background; (3) 0-dB SNR, with 70 dB of background noise. Background noise was continuous four-talker babble also used previously. Stimuli were presented in two (quiet and 10-dB SNR) or three runs (0-dB SNR) of 250 presentations with an offset-to-onset interval of 1100 ms. Participants were seated comfortably in a sound-treated booth, fixating on a picture on the wall and instructed to press a button as quickly as possible to the deviant stimulus.

Fifteen channels of a QuickCap were referenced to a nose electrode with a forehead ground. Online filter settings of 0–100 Hz and a sampling rate of 1000 Hz were used (Neuroscan, SynampsRT). Blink correction was applied offline (see Bennett et al., 2012), and then sweeps with activity $\pm 70 \mu\text{V}$ were rejected. Average waveforms were then bandpass filtered from 1–30 Hz. Two separate waveforms were generated: a standard response to /ba/ from electrode Cz (for N1), and a deviant waveform to /da/ from electrode Pz (for P3). Only hits (deviant stimulus presented and participant pressed the button) and correct rejections (standard stimulus was presented and participant did not press the button) were included in the averages (misses and false alarms were excluded from averaging). Peaks were initially determined automatically based on latency range (N1: 70–140 ms and P3: 250–450 ms). Two judges then independently verified peaks using comparison electrodes (to demonstrate replicability and peak inversion) but focusing on instances where the algorithm may have forced a peak that should be considered noise, such as when the automatic peak was assigned to an upward or downward sloping portion of the waveform (i.e., non-peak) at the edge of the latency range. This occurred for 10% of peaks and was usually for the 0-dB SNR condition. In these instances, latencies and amplitudes were excluded from the data.

The QuickSIN (Etymotic Research, Inc.) was used as the behavioral speech understanding measure and represents the estimated “SNR loss,” which represents the participant’s deviation from the normal 50% performance (i.e., SNR thresholds). The Digit-Symbol Coding test (DSC), a subtest of the Wechsler Adult Intelligence Scale-III (Wechsler 1997), assesses non-auditory complex processing speed effects and executive function, and it was

used as the cognitive outcome measure in this study. The DSC scores were split in half at the mode resulting in “Poorer” (n=16) and “Better” (n=17) DSC groups.

2.3. Analysis

A Bayesian multilevel model was used to jointly model the DSC score, N1/P3 latency, and the QuickSIN score to evaluate the interaction effects of cognition and neural coding on speech understanding. A posterior probability on the interaction effect of greater than 0 (i.e., a positive interaction) supported an increased benefit of cognition with better neural coding, consistent with the serial processing model. Whereas, a posterior probability on the interaction effect that was less than 0 (i.e., a negative interaction) supported the compensatory model (see appendix for model details).

3. Results

3.1. Signal-to-Noise Ratio Effects Electrophysiology

The effects of SNR on neural coding are apparent in Figure 2 showing the grand mean response waveforms for responses to the deviant (Figure2a) and standard (Figure2b) stimuli. Generally, amplitudes decrease and latencies increase as SNR decreases from Quiet to 10 dB to 0 dB. Variability of the responses across individuals is shown as shaded ranges of plus/minus one standard error of the mean.

3.2. Relationship between coding, cognition, and perception

Model estimates of the QuickSIN, DSC, and two different neural coding measures were completed (N1 and P3). Figure 3 presents the raw data in the same layout as the schematic models in Figure 1 (note the reverse scaling of the axes). Evoked potentials were averaged across the SNRs because of their parallel change across SNR within participants. Data in Figure 3 support the underlying assumptions that better neural coding (at least in the case of N1) and better cognition are both associated with better speech understanding. However, the data fail to strongly support either the compensatory or serial processing models as evidenced by the parallel nature of the lines. This is confirmed by the Bayesian statistical model fit to the full dataset: the posterior probability of a negative interaction effect between cognition and obligatory coding is 0.46 and cognitive coding is 0.49. Conversely, the posterior probability of a *positive* interaction effect between cognition and obligatory coding is 0.54 (i.e., 1 minus 0.46) and cognitive coding is 0.51 (i.e., 1 minus 0.49). These correspond to nearly 1:1 odds supporting either model; therefore, there is little evidence in this sample to prefer either model.

4. Discussion

Our initial models relating neural coding, speech perception, and cognition were proposed and tested by modeling N1/P3 latency, QuickSIN, and DSC outcome measures. Relationships between some of these measures have been explored previously. Billings et al. (2015) demonstrated strong significant relationships between speech perception and N1 and P2 evoked potentials, and Polich et al. (1983) illustrated correlates between digit span testing and P3 latency. Age-associated changes in attention and working memory have been

reported in the literature, and age-related changes in these factors have been associated with reduced performance on speech recognition tests (Hulme 1989; Rouleau and Belleville 1996; Wingfield 1999). In studies of Veterans, we have shown cognitive tests to be strongly associated with performance on compressed sentence recognition scores even when age and hearing loss were controlled (Vaughan et al, 2008). Previous studies also suggest that cognitive resources may be used to improve speech understanding in difficult listening situations and that memory may assist in this compensatory effect (Cabeza and Dennis, 2007; Luce and Pisoni, 1998; Moradi et al., 2014; Ronnberg et al., 2013). It is reasonable to think that both cognition and neural coding are significant contributors to speech perception in noise.

In the current study neither the compensatory nor the serial processing model fit the data. Instead an additive model may be more appropriate. In such a model, neural coding and cognition would have independent effects, rather than interactive effects, on speech perception. This study represents an effort to develop a mechanistic framework for how these variables may relate and may be limited by the appropriateness of these measures. These results illustrate the potential usefulness of jointly modeling these three modalities (i.e., speech perception, neural coding, and cognition).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the United States (U.S.) Department of Veterans Affairs (RR&D Service, C74554R) and the U.S. National Institutes of Health (NIDCD-DC15240). The contents do not represent the views of the U.S. Department of Veterans Affairs or the U.S. government. The authors wish to thank Jane Gordon, Dan McDermott, and Tina Penman for their efforts with data collection and processing.

References

- Anderson S, White-Schwoch T, Parbery-Clark A, & Kraus N (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing Research*, 300, 18–32. [PubMed: 23541911]
- Bennett KO, Billings CJ, Molis MR, et al. (2012). Neural encoding and perception of speech signals in information masking. *Ear Hear*, 32, 1–8.
- Billings CJ, McMillan GP, Penman TM, Gille S (2013). Predicting perception in noise using cortical auditory evoked potentials. *J Assoc Res Otolaryngol*, 14(6): 891–903. [PubMed: 24030818]
- Billings CJ, Penman TM, McMillan GP, Ellis E (2015). Electrophysiology and perception of speech in noise in older listeners: effects of hearing impairment & age. *Ear Hearing*, 36(6): 710–722. [PubMed: 26502191]
- Cabeza R, & Dennis NA (2007). Neuroimaging of healthy cognitive aging In Craik FIM & Salthouse TA (Eds.), *The handbook of aging and cognition*. Psychology Press.
- Erb J, & Obleser J (2013). Upregulation of cognitive control networks in older adults' speech comprehension. [Original Research]. *Frontiers in Systems Neuroscience*, 7.
- Folstein MF, Folstein SE, McHugh PR (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*, 12, 189–198. [PubMed: 1202204]
- Henry MJ, Herrmann B, Kunke D, & Obleser J (2017). Aging affects the balance of neural entrainment and top-down neural modulation in the listening brain. *Nat Commun*, 8, 15801. [PubMed: 28654081]

- Hulme C, Tordoff V (1989). Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology*.
- Konrad-Martin D, Billings CJ, McMillan GP, McDermott D, Gordon J, Austin D, Dille MF (2015). Diabetes-associated changes in cortical auditory evoked potentials in relation to normal aging. *Ear Hearing*. 37(3): 1–15.
- Luce PA, Pisoni BD (1998). Recognizing spoken words: the neighborhood activation model. *Ear and Hearing*. Vol.19 p.1–36. [PubMed: 9504270]
- Polich J, Howard L, Starr A (1983). “P300 latency correlates with digit span.” *Psychology*. Vol.20 p665–9.
- Moradi S, Lidestam B, Saremi A, Ronnberg J (2014). Gated auditory speech perception: effects of listening conditions and cognitive capacity. *Front Psychol*. 5:531. [PubMed: 24926274]
- Ronnberg R, Lunner T, Zekveld A, Sorqvist P, Danielsson H, Lyxell B et al. (2013). The ease of language understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*. 7:31. [PubMed: 23874273]
- Rouleau N, Belleville S (1996). Irrelevant speech effect in aging: an assessment of inhibitory processes in working memory. *The Journals of Gerontology. Series B*, Vol.51 p.356–63
- Vaughan N, Storzbach D, Furukawa I (2008). Investigation of potential cognitive tests for use with older adults in audiology clinics. *J Am Acad Audiol*. 19(7):533–41 [PubMed: 19248729]
- Wingfield A, Ducharme J (1999). Effects of age and passage difficulty on listening-rate preferences for time-altered speech. *The Journals of Gerontology. Series B*, Vol.54 p.199
- Wong PCM, Ettlinger M, Sheppard JP, Gunasekera GM, & Dhar S (2010). Neuroanatomical characteristics and speech perception in noise in older adults. *Ear & Hearing*, 31(4), 471–479. [PubMed: 20588117]
- Wong PCM, Jin JX, Gunasekera GM, Abel R, Lee ER, & Dhar S (2009). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia*, 47(3), 693–703. [PubMed: 19124032]

Figure 1. Schematic representing theoretical expectations for compensatory and serial models. Hypothesized relationships between neural coding, cognition, and speech perception measures are shown. In a compensatory model, cognitive function might compensate for poor neural coding; whereas, in a serial model sufficient neural coding would be necessary to allow cognition to then improve performance.

Figure 2. Grand mean ($n=33$) time waveforms and response data to deviant signals (left) and waveforms to standard signals (right). SNR conditions for Quiet (blue, solid), 10 dB (red dashed), and 0 dB (green, dash-dot) illustrate effects of SNR on both N2b-P3 waves (left) and N1-P2 waves (right).

Figure 3.

Raw data for 33 participants relating electrophysiology (N1 latency shown on left; P3 latency shown on the right), speech perception (QuickSIN score), and cognition (DSC score). Lines of best fit are somewhat parallel demonstrating the mismatch with the compensatory processing and serial processing models. Perhaps a third additive processing model may be more appropriate to characterize these data.