



Published in final edited form as:

Hear Res. 2018 November ; 369: 90–102. doi:10.1016/j.heares.2018.03.024.

A perspective on brain-behavior relationships and effects of age and hearing using speech-in-noise stimuli

Curtis J. Billings^{a,b} and Brandon M. Madsen^a

^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, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, OR 97239, USA

Abstract

Understanding speech in background noise is often more difficult for individuals who are older and have hearing impairment than for younger, normal-hearing individuals. In fact, speech-understanding abilities among older individuals with hearing impairment varies greatly. Researchers have hypothesized that some of that variability can be explained by how the brain encodes speech signals in the presence of noise, and that brain measures may be useful for predicting behavioral performance in difficult-to-test patients. In a series of experiments, we have explored the effects of age and hearing impairment in both brain and behavioral domains with the goal of using brain measures to improve our understanding of speech-in-noise difficulties. The behavioral measures examined showed effect sizes for hearing impairment that were 6–10 dB larger than the effects of age when tested in steady-state noise, whereas electrophysiological age effects were similar in magnitude to those of hearing impairment. Both age and hearing status influence neural responses to speech as well as speech understanding in background noise. These effects can in turn be modulated by other factors, such as the characteristics of the background noise itself. Finally, the use of electrophysiology to predict performance on receptive speech-in-noise tasks holds promise, demonstrating root-mean-square prediction errors as small as 1–2 dB. An important next step in this field of inquiry is to sample the aging and hearing impairment variables continuously (rather than categorically) – across the whole lifespan and audiogram – to improve effect estimates.

Keywords

speech in noise; speech perception in noise; cortical auditory evoked potentials; aging; hearing impairment; hearing loss; brain and behavior; correlation; prediction

1. Introduction

Successful communication in difficult listening environments is dependent upon how well the auditory system encodes and extracts signals of interest from other competing acoustic information. Therefore, it is not surprising that older individuals and people with impaired auditory systems have increased difficulty understanding target speech in background noise. This point is especially important for auditory rehabilitation of older individuals with

hearing impairment. Kochkin (2002, 2010) found hearing-aid users who responded to a survey were most dissatisfied with listening in noisy situations; however, very little is done clinically to diagnose and treat listening-in-noise problems. Part of the problem may be the limited understanding of the mechanisms at work in the perceptual process. Understanding central-auditory-system contributions to signal-in-noise difficulties may help to improve performance in noise.

1.1. Noise is all around us, and the ability to understand speech in noise varies

Difficulties understanding speech in background noise have been linked to aging and hearing impairment, both as independent factors and when their effects are combined (Dubno et al., 1984; Gordon-Salant and Fitzgibbons, 1995; Humes and Roberts, 1990; Souza and Turner, 1994; Studebaker et al., 1999). The overlap between these factors is noteworthy, as hearing impairment increases dramatically with each decade of life: by 60 years of age approximately 50% of individuals experience hearing loss, with the prevalence increasing to 90% by 80 years of age (Agrawal et al., 2008; Cruickshanks et al., 1998). In addition, several other conditions exacerbate difficulties understanding signals in noise, such as traumatic brain injury, diabetes, multiple sclerosis, and Parkinson's disease (Folmer et al., 2014; Frisina et al., 2006; Gallun et al., 2012; Lewis et al., 2006). Understanding speech in background noise is crucial for day-to-day functioning in the modern world, given the ubiquitous nature of loud noise in everyday life. Figure 1 represents the signal-to-noise ratios (SNRs) in various listening situations that are encountered in everyday life as recorded across several studies (Hodgson et al., 1999; Markides et al., 1986; Pearsons et al., 1977; Plomp, 1977; Smeds et al., 2015; Teder, 1990). Although the use of different methodologies across these studies results in some SNR variability within each situation, generally these results strikingly depict the variety of common situations where one must contend with the presence of background noise. Furthermore, impaired understanding of speech in background noise may have far-reaching consequences, given that hearing dysfunction has been linked to depression, isolation, decreased functional status, and poorer quality of life, as well as decreased participation in social activities and increased stress levels that can impact personal relationships (Arlinger, 2003; Jones et al., 1984; Keller et al., 1999; Mulrow et al., 1990; Seniors Research Group, 1999).

1.2. Ability to understand speech in noise varies

A hallmark of perception-in-noise literature is the wide range of performance across individuals, despite similarities in the pure-tone audiogram or speech understanding in quiet (e.g., Dirks et al., 1982). Figure 2 illustrates this characteristic variability in a group of older, hearing-impaired individuals with similar ages and hearing thresholds, as measured in steady-state speech-spectrum noise (data from Billings et al., 2015). Performance is quantified employing a commonly-used metric for speech-in-noise perception: the SNR50, defined as the dB SNR threshold at which 50% speech-recognition performance is achieved. A portion of the variability in auditory rehabilitation success may be explained by central auditory coding differences between individuals (Billings et al., 2013, 2015). In the clinic, an audiologist may see two patients of similar age with similar hearing impairments who perform very differently in background noise (if speech-in-noise testing is performed at all); however, auditory rehabilitation techniques for these two individuals are likely quite similar.

1.3. The case for a physiological measure

A physiological measure of auditory neural coding may be helpful in explaining some of the performance variability and perhaps, with further development, serve as a substitute for speech-understanding testing in those who cannot be tested reliably. One technique that shows particular promise is electroencephalography (EEG): using electrodes placed on the scalp to record electrical potentials associated with synchronous neural activity.

There has been great interest in using EEG data to understand more about the process of speech comprehension in background noise. One approach is to examine waypoints along the auditory neural pathway (e.g., brainstem nuclei, auditory cortex) to determine how the auditory signal is being encoded, and to determine what neural information is available to the listener for the resulting perception of speech. The underlying assumption is that accurate understanding in background noise is dependent, in part, on the integrity of neural coding of the auditory stimulus. Measurements of cortical neural coding may help clarify the capacity of the auditory system to faithfully code signals in noise before higher-level processing is applied, resulting in perception. Representing the intermediate processing stages may be important, because as listening environments get more difficult, more top-down processes may be used to make up for missed cues; therefore, it would be helpful to have a reliable physiological measure that reflects pre-cognitive information about acoustic coding.

The main advantage of EEG over other techniques is that it has extremely high temporal resolution. Furthermore, EEG recording is both non-invasive and silent, which is especially advantageous for studying hearing. The hardware for recording EEG data also tends to be less expensive than most other techniques for studying brain function.

There are several limitations and caveats that come with EEG, however. Only synchronous activity from thousands or tens of thousands of neurons at once, all with similar spatial orientations, will produce potentials large enough to measure from the scalp. The actual “firing” of a neuron (action potential) happens too fast and the resulting voltage field decreases too rapidly to achieve this type of summation across neural units; therefore, the potentials that sum to produce the EEG signal do not come directly from the firing of neurons but rather from postsynaptic potentials that follow. Thus, the pyramidal neurons, with their long apical dendrites running parallel to one another and perpendicular to the cortical surface, are thought to be disproportionately represented in the EEG signal compared to other areas and cell types. Finally, EEG has poor spatial resolution, and tracing signals back to their generation sites in the brain poses difficult problems which are beyond the scope of this article to address in detail. In general, given these caveats, researchers must be cautious not to misinterpret EEG as reflecting “overall” brain activity, and must also be cautious in making claims about the specific underlying neural generators involved.

When EEG recordings are analyzed in time windows that are aligned to the presentation of an auditory stimulus, they are referred to as auditory-evoked potentials (AEPs). AEPs are the most popular category of electrophysiological measures that have been used in attempts at capturing the type of neural coding mentioned above, and they will constitute the focus of this paper. Like all measures in the evoked-potentials family, AEPs attempt to quantify the

nervous system's time-locked response to an external stimulus – in this case, an auditory stimulus. The time-locked response is significantly smaller in amplitude than the ongoing EEG activity, so it must be extracted through signal averaging over many stimulus presentations, and analysis is generally done on voltage waveforms in the time domain. The major landmarks of the waveform, positive and negative voltage peaks, are characterized in terms of their amplitude and latency. Evoked potentials are considered distinct from “induced” potentials, which are also related to discrete stimulus events but are not phase-locked to them, and which tend to be analyzed in the frequency domain. Together, evoked and induced potentials form the larger class known as event-related potentials (ERPs). These are in turn distinct from analysis based on the ongoing or spontaneous EEG activity.

AEPs are known for having low variability within subjects (Tremblay et al., 2003) and relatively high variability between subjects, which is one notable property that suggests they may be sensitive to the types of individual differences in encoding that may be important clinically. However, aside from the auditory brainstem response (ABR) – which, with its comparatively low intersubject variability within a clinical subpopulation, is considered the “gold standard” for hearing assessment in newborns – the potential clinical utility of AEPs remains largely unproven. Still, there are many clinically relevant applications being explored, and several of the most promising are briefly described below.

1.3.1. Difficult-to-test patients—AEPs may complement or replace behavioral measures in difficult-to-test individuals. Individuals with co-morbid health conditions such as traumatic brain injury (TBI), post-traumatic stress disorder (PTSD), and dementia encounter difficulties when completing behavioral audiologic testing. Specifically, Veterans with TBI often suffer from fatigue, those with PTSD suffer from anxiety, and those with dementia often cannot follow instructions. As a result, behavioral testing can be difficult to complete, and the results may be inaccurate. Thus, replacing or supplementing behavioral testing with a short and simple physiological measure that accurately predicts speech understanding in noise could fill a gap in current clinical assessments for individuals who are difficult to test behaviorally.

1.3.2. Diagnosis of supra-threshold hearing impairments—When speech understanding in noise is worse than what would be predicted from pure-tone audiometric results, AEPs may aid in more accurately describing performance and diagnosing problems. It is not uncommon to encounter patients who present with normal or borderline-normal pure-tone thresholds. Among U.S. Veterans seen in Veterans Affairs audiology clinics, the prevalence of those with normal audiometric thresholds is 10%, and a portion of these Veterans reports communication and speech-understanding difficulties disproportionate to their normal hearing thresholds (Billings et al., in press). Evidence suggests that the audiogram, which is based on tonal thresholds, is unable to reliably detect or describe supra-threshold auditory impairment such as perception in noise (Carhart and Tillman, 1970; Groen, 1969; Wilson, 2011). One potential cause of suprathreshold difficulties may be neural degradation (Kujawa and Liberman, 2009). Supplementing audiometric testing with a physiological measure could assist in identifying sources of neural degradation and thus improve diagnosis of these conditions. For example, envelope- and frequency-following

responses, a type of steady-state AEP, have been shown to be sensitive to speech-understanding difficulties (Anderson et al., 2010; Ruggles et al., 2011).

1.3.3. Rehabilitation planning—Ideally, rehabilitation interventions should be tailored to target the underlying hearing deficit, and AEPs may help to pinpoint the nature of these deficits. For instance, regarding auditory training: if obligatory bottom-up neural coding (as indicated by brainstem and cortical AEPs) is occurring as expected but performance is poor, that may constitute evidence in favor of an auditory rehabilitation approach that targets top-down processing, such as through explicitly taught listening strategies or by employing synthetic methods of auditory training (i.e., with tasks focused on higher-order speech representations, such as sentences, stories, or conversations). In contrast, if AEPs show that exogenous neural coding is not occurring normally, that might support a bottom-up rehabilitation program that focuses on recognition of lower-order acoustic cues and employs analytic auditory training methods (i.e., with speech broken down into basic constituent parts: syllables, phonemes, or even isolated cues within a phoneme, such as vowel formant transitions) in order to improve access to the important acoustic cues (Kricos and McCarthy, 2007).

1.3.4. Monitoring effectiveness of auditory training—Finally, AEPs could be used to monitor success of an auditory training program. The time course of training effects is an important scientific question that may have direct clinical application. Specifically, it has been demonstrated that neural changes can sometimes be detected earlier than behavioral changes (e.g., Tremblay et al., 1998). Clinically, then, if a physiological test detected emerging neural changes that had not yet consolidated into improved behavioral performance, that could be considered as evidence for continuing with the current training regime. However, if physiological testing during training does not show neural changes, then the clinician might choose to pursue an alternate strategy. The combination of behavioral and electrophysiological information could lead to tailored rehabilitation specific to the needs of each individual.

1.4. Some other functional neuroimaging techniques

Though not the focus of the present paper, it is important to note that, aside from EEG, there are other functional neuroimaging techniques that may be able to help us understand human speech-in-noise listening more fully. Two of the most promising – magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) – are briefly reviewed below. For a deeper comparison between functional neuroimaging techniques, see Bunge and Kahn (2009).

1.4.1. Magnetoencephalography (MEG)—Just as the neural activity of the brain in response to auditory stimulation produces electrical potentials that can be detected by electrodes and recorded as AEPs, the electrical currents from that synchronous, aligned neural activity in turn induce magnetic fields that can be detected by magnetometers and recorded as AEFs (auditory-evoked fields). The technique of recording these magnetic fields is called magnetoencephalography (MEG).

MEG shares many of the strengths and weaknesses of EEG. Like scalp EEG, MEG has very high temporal resolution. Also like EEG, MEG is only capable of detecting *synchronous* neural activity that is relatively close to the cortical surface, and the signals of interest for AEFs are about 100 times smaller in amplitude than the ongoing MEG.

The main advantage of MEG over EEG is that magnetic fields pass freely through the scalp with virtually no distortion, whereas scalp EEG is highly affected by individual differences in head geometry, as the signal is “smeared” according to the volume, shape, and conductance properties of the meninges, cerebrospinal fluid, skull, and scalp.

On the other hand, MEG setups tend to be larger, rarer, more expensive, and less portable than their EEG counterparts. A key factor underlying these limitations is the fact that cortical AEF amplitudes are about 10 million times smaller than the “noise” of the ambient magnetic field in a typical urban environment, so MEG is only possible in specially designed rooms with robust magnetic shielding.

1.4.2. Functional magnetic resonance imaging (fMRI)—Functional magnetic resonance imaging (fMRI) is a very different type of neuroimaging technique from EEG and MEG. Rather than measuring the voltage changes associated with neural activity, the most common implementation of fMRI measures brain activity by monitoring changes in blood flow, relying on the fact that oxygenated blood flow to a given brain region increases in concert with heightened neural activity because of increased energy demands. Like EEG and MEG, the fMRI signal is noisy and must be extracted through signal averaging and statistical analyses.

The main advantage of fMRI over EEG and MEG is its ability to localize the activity it detects to specific regions in the brain, within 1–5 mm. This stands in marked contrast to EEG and MEG, which both present problems and complications when it comes to source localization. The tradeoff is that fMRI has a temporal resolution on the scale of seconds rather than milliseconds. This means that the order in which different regions are activated cannot be determined from fMRI data, and transient activations lasting only a few milliseconds might not be detected at all. One aspect of fMRI that complicates testing with auditory stimuli in particular is that the scanners themselves produce frequent “knocking” sounds that are quite loud. This can be somewhat mitigated by careful timing of stimuli relative to the magnetic pulses of the scanner (sometimes called “sparse” fMRI) or by presenting through specialized earphones.

1.5. Cortical auditory-evoked potentials (CAEPs)

All electrophysiological data presented in the remainder of this paper are from *cortical* AEPs (CAEPs). More specifically, the data concern the P1, N1, P2, and N2 peaks. This complex of voltage peaks is generated in response to audible stimulus onsets and relatively abrupt audible changes in stimulus properties such as amplitude, frequency, and spatial location. In YNH listeners, these peaks occur roughly 50–300 ms after the evoking stimulus. The most stable and robust of these peaks are N1 and P2, and they are where we focus most of our analyses in this paper. Traditionally, the most common single wave to be analyzed is N1 (~100 ms) for adults and adolescents or P1 for children under 10 years old (in children, P1 is

both larger and later than in adults); these are thought to represent the processing of stimuli at the level of the auditory cortex. Though attentiveness is not required to produce the N1 peak, its properties can be affected by attention in some circumstances. For additional background on CAEPs and their clinical applications, see Stapells (2002).

2. Experiments and outcome measures

2.1. Overview: A brain-and-behavior approach

Herein, we review data from our laboratory and attempt to quantify the effects of aging and hearing impairment on both behavior and CAEPs. We also present some new data that illustrate these effects as a function of background noise type. Finally, we review some of the attempts that have been made toward clarifying the ways in which CAEPs in noise may correlate with or predict behavioral speech-in-noise listening performance.

In a series of experiments, CAEPs and behavioral data were collected with the goal of investigating how electrophysiology might inform our understanding of speech perception in noise. The data summarized in this review paper are taken from three studies conducted in our laboratory, and they have all been at least partially presented elsewhere. The portions of the data we have chosen to present here were selected to illustrate the effects of age, hearing impairment, signal type, and noise type. Special emphasis has been placed on: (i) instances where CAEPs and behavioral measures intersect to reveal the differential effects of age and hearing impairment, and (ii) instances where CAEPs were used to predict behavior. All three experiments used subject groups to explore age and hearing impairment effects by testing younger normal-hearing (YNH), older normal-hearing (ONH), and older hearing-impaired (OHI) participants.

2.2. Experiment 1: Billings et al., 2015

Experiment 1, reported previously (Billings et al., 2015), included both CAEP and behavioral measures for the 15 participants in each subject group. Behavioral performance was tested by asking subjects to listen to and repeat sentences from the IEEE corpus (Institute of Electrical and Electronic Engineers, 1969). These sentences were presented in steady-state speech-spectrum noise at up to seven SNRs ranging from -10 to $+35$ dB. Responses were scored by the number of pre-designated key words subjects correctly repeated from the sentences. CAEPs were recorded in response to the naturally-produced syllable /ba/ presented at a subset of the SNRs used behaviorally. To quantify these responses, the following waveform features were examined: four peaks from the average-referenced waveform at the Cz electrode, located on top of the head (P1, N1, P2, N2); three peaks from the global field power (GFP) waveform, which represents the standard deviation across all electrodes at each time point (N1, P2, N2); and the full-wave-rectified area in the N1-P2-N2 latency range (50–550 ms post-stimulus) for both Cz and GFP. Each peak was quantified in terms of both amplitude and latency. Not all measures were included in all analyses; this will be further explained as each analysis is discussed below.

2.3. Experiment 2: Billings et al., 2016

Experiment 2, reported previously (Billings et al., 2016), employed behavioral measures *only* and had 20 participants in each group. Northwestern University Auditory Test Number 6 (NU-6; Tillman and Carhart, 1966) words were presented in the same steady-state speech-spectrum noise used in Experiment 1, with SNRs ranging from -10 to 35 dB. Responses to target words were scored in two ways: (i) percentage of words correct and (ii) percentage of phonemes correct. As with Experiment 1, multiple signal levels were tested, but only results obtained using the highest signal level (80 dB C) are included here.

2.4. Experiment 3: Maamor and Billings, 2017

Experiment 3 included both CAEP and behavioral measures for the 10 participants in each group. Noise type was varied between four-talker babble, steady-state speech-spectrum noise, and speech-spectrum noise with the amplitude-modulation envelope of a single talker. Only results obtained using four-talker babble are included in this review. The CAEP results have been reported previously (Maamor and Billings, 2017); however, the behavioral data and prediction models incorporated here are being reported for the first time. The behavioral measures presented here are the Quick Speech-in-Noise Test (QuickSIN; Killion et al., 2004) and the Words-in-Noise test (WIN; Wilson and Burks, 2005). Both tests use speech signals presented against a background of four-talker babble at varying SNRs. For QuickSIN, the target signals consist of sentences drawn from the IEEE corpus (six per trial); five “key words” are scored per sentence (the rest are ignored); and the level of the babble increases by 5 dB after each sentence while the signal level remains constant, producing SNRs from $+25$ to 0 dB. For WIN, the target signals are NU-6 words (35 per trial); each word is scored entirely correct or incorrect; and signal level decreases by 5 dB after each block of 5 words while the babble level remains constant, producing SNRs from $+30$ to 0 dB. CAEPs were recorded to /ba/ syllables at 65 dB SPL in four-talker babble at several SNRs. CAEPs were used to develop universal (non-group-specific) prediction models for behavioral performance.

2.5. SNR50 and the psychometric function

To quantify behavioral performance across SNRs with a single scalar metric, in all three experiments we chose to use the 50% speech-recognition-in-noise threshold, commonly known as the SNR50. An SNR50 was derived for each participant by fitting a logistic psychometric function to their performance data and calculating the SNR at which the function intersected the 50-percent-correct line. Group and individual results for IEEE sentences presented in steady-state speech-spectrum noise for each of the three test groups from Experiment 1 are shown in Figure 3 as an example of the nature of the behavioral data collected as well as general group effects and individual variability.

2.6. CAEP waveform measures: N1, P2, area

Figure 4A shows how SNR affects the waveform morphology of the response across the three participant groups in Experiment 1, and 4B illustrates the growth functions associated with N1, P2, and area measures recorded at the Cz electrode using a common-average reference. It is notable that in looking at P2 amplitude the YNH group emerges as having a

pattern that is distinct from both older groups, whereas for N1 amplitude it is the ONH group that stands out from the other two. One might interpret this as evidence that the OHI and YNH groups share similar characteristics. However, these surface-level similarities belie what we suspect are fundamentally different forces in play underneath.

3. Effects of age and hearing loss

The relative effects of age and hearing impairment on speech understanding in noise are important to understand given their co-occurrence. However, because age and hearing loss are often correlated, it is difficult to fully control for confounds that may be present when trying to isolate these effects. The approach that many researchers have taken is to recruit groups of subjects with and without hearing loss who are older and younger so as to represent three or four discrete groups. Using this method, we have attempted to characterize the effects of age and hearing loss for both brain and behavioral measures. For the purposes of this paper we define the age effect as the difference in the measurements between ONH and YNH groups (ONH-YNH), and we define the hearing loss effect as the difference between OHI and ONH (OHI-ONH).

3.1. Behavioral effects

Figure 5 represents the behavioral effects of age (Figure 5A) and hearing loss (Figure 5B) for different signal types, noise types, and scoring methods. Within each panel, a dashed line separates estimates in steady-state noise (left) from estimates in four-talker babble (right). It should be noted that results from all three experiments are presented in this figure: results for IEEE sentences in steady-state noise are from Experiment 1; results for NU-6 words in steady-state noise, scored by word and by phoneme, are from Experiment 2; and all the four-talker babble results are from Experiment 3. Increased age (ONH vs. YNH) was associated with a 2- to 4-dB increase in SNR50, depending on the signal and noise that was used. In babble noise hearing impairment was associated with a 1- to 3-dB increase in SNR50, closely resembling the effect of age. In steady-state noise, however, hearing impairment was associated with much larger effects: between about 10–12 dB. In steady-state noise, the hearing-impairment effect was 6–10 dB larger than the age effect, reflecting the difficulty experienced by older-hearing impaired individuals. It is interesting that the hearing-impairment effect is more comparable with the age effects when testing occurred in babble noise. Initially, one might think that the four-talker babble was a difficult noise for all groups and effectively limited the spread that is often seen within older hearing-impaired individuals. Although the spread was limited for the OHI group in babble, it was because many OHI participants actually performed better in babble than in steady-state noise. This result does not follow the established understanding of how babble would increase difficulty.

It may be that the differences in measurement tool contributed to the small effect in babble relative to steady-state noise. All babble SNR50s (Experiment 3) were derived using clinical tests (QuickSIN and WIN), whereas steady-state SNR50s (Experiments 1 and 2) were derived using custom tests developed in our laboratory. The custom tests drew from the same corpora as the QuickSIN and WIN (IEEE sentences and NU-6 words, respectively), but each used randomly selected tokens from the entire corpus, not just the relatively small subset

used by its clinical counterpart. It seems feasible that a sparser sampling from a larger set of tokens could result in a greater spread in results than a denser sampling of a smaller set. Furthermore, for the custom tests a given token had an equal chance of appearing at any SNR, whereas the clinical tests always present a given token in the same SNR each time it occurs. Another hypothesis one might raise to explain the results is the fact that the clinical tests estimate SNR50 directly from the total score using the Spearman-Kärber method, while the custom tests extract SNR50 from a fitted logistic function. However, we currently believe that the estimation method (curve-fitting vs. Spearman-Kärber) is unlikely to have an appreciable effect, because the two methods are known to produce very similar estimates of 50% thresholds (Armitage and Allen, 1950). Finally, an overarching problem with both of these explanations is that no equivalent noise-type discrepancy is observed for the age-effect data.

Another way to characterize the effects of age and hearing impairment is to examine differences between groups in measured performance at a fixed SNR. Referring back to the psychometric function (Figure 3), this would be equivalent to drawing a vertical comparison line instead of a horizontal one. Figure 6 shows results by group and SNR and illustrates the importance of relatively small changes in SNR. Older individuals can experience extreme difficulty even at SNRs that can typically be found in common conversational environments, such as at a cocktail party or in a restaurant, represented by 0 and 4 dB, respectively. In these difficult situations, average speech understanding accuracy remains below 50% correct for the older, hearing-impaired group.

3.2. CAEP effects

Data from Experiment 1 (steady-state noise) and Experiment 3 (four-talker babble) are combined here to summarize the effects of age and hearing impairment on CAEPs. Figure 7 shows effects of age (Figure 7A) and hearing impairment (Figure 7B) on N1 and P2 latencies and amplitudes as well as full-wave-rectified area measures, which integrate absolute amplitude over a wide range of response latencies (50–550 ms for Experiment 1 and 30–350 ms for Experiment 3). Values for peak latencies, peak amplitudes, and rectified areas were averaged within each group, and inter-group contrasts were used to estimate the effect sizes for age and hearing impairment.

It is not clear why behavioral effects are more systematic than electrophysiological effects. One possibility is potentially increased inter-subject variability on average for CAEPs relative to behavior. One might think that reduced variability could result because behavior is filtered through active cognitive processes with similar intent across participants, whereas pre-attentive AEPs give only a glimpse of the signal coding and not what the participant does with that information. Cortical AEPs are largely resistant to the top-down control that is present in behavior, which likely contributes to the discrepancies between electrophysiological and behavioral effects. In animal models, aging and hearing impairment have been shown to result in a net down-regulation of neural inhibition throughout the central auditory system, from the cochlear nucleus all the way up through the primary auditory cortex (PAC; Caspary et al., 2008). It is therefore plausible that the counterintuitive

pattern of larger AEP amplitudes seen in older and hearing-impaired human participants may also be related to reduced neural inhibition.

4. Associations between brain and behavior

The relationship between cortical AEPs and behavior has been explored for many decades. Historically, cortical AEPs have been used for the clinical purpose of estimating behavioral thresholds in difficult-to-test adult populations. However, this practice has since fallen out of favor clinically and been almost entirely supplanted by use of the ABR. This occurred mainly because CAEPs often do not reach maturity until the late teenage years, whereas the ABR is present even before birth and can therefore be used to test patients of all ages, including newborn infants. Nonetheless, cortical AEPs continue to be used for some applications in adults and have been used successfully to estimate the pure-tone thresholds, usually to within 10 dB, with the cortical AEP threshold generally being elevated relative to the behavioral threshold (Ikeda et al., 2010; Prasher et al., 1993; Rapin et al., 1970; Yeung and Wong, 2007). Sometimes, however, the CAEP threshold may be more than 20 dB above the behavioral threshold, or even absent entirely (e.g., Glista et al., 2012; Van Maanen et al., 2005). On its own, an absent evoked response is non-diagnostic, because lack of a response to a given stimulus may be due to a number of factors unrelated to the brain's response to that stimulus (e.g., an increase in biological or environmental noise unrelated to the stimulus could obscure the response).

For signal-in-noise testing, the existing literature combining CAEPs and behavior is often limited to small-sample, group-design studies resulting in somewhat restricted generalizability, with results specific to the characteristics of the tested subgroup. Several studies have explored the relationship between behavior and CAEPs, particularly in the domain of speech-in-noise testing. Some researchers have used the Hearing in Noise Test (HINT) and found statistically significant ($\alpha = .05$) weak-to-moderate correlations with N2 amplitude in 32 children ($r = -0.383$; Anderson et al., 2010) and N1 amplitude in 22 adults ($r = 0.520$; Parbery-Clark et al., 2011). Campbell et al. (2013) and Bidelman et al. (2016) both used the QuickSIN as their speech-in-noise test and found a statistically significant ($\alpha = .05$) moderate-to-strong correlations with P2 latency ($r = 0.494$) and N1 amplitude ($r = 0.63$) in 17 and 12 adults, respectively. Using a behavioral measure that combines accuracy and reaction time, a weak-to-moderate correlation with N1 and P2 amplitude was found ($r = 0.37$; Bidelman et al., 2014). Our past studies reveal similar values with best correlations between IEEE sentences and N1, P2, P3 latency/amplitude ranging between $r = 0.53$ and $r = 0.80$ (Bennett et al., 2012; Billings et al., 2013; Billings et al., 2015). The correlation coefficients that have been found for brain-behavior relationships using cortical AEPs are generally very similar to those found using brainstem AEPs (Anderson et al., 2011; Bidelman et al., 2014; Parbery-Clark et al., 2011). Using an oddball paradigm, Bertoli and colleagues (2005) examined the degree to which the SPIN-generated SNR50 was correlated with the latency and amplitude from three CAEP peaks (MMN, N2b, and P3b) recorded in a noisy background from 30 individuals and found only a moderately strong correlation with N2 latency to be statistically significant ($\alpha = .05$). These studies highlight some of the difficulties associated with brain-behavior correlations: (a) these studies are often limited to relatively small samples; (b) a single behavioral measure (often the SNR50) is correlated

with any number of CAEP measures; and (c) it is usually not clear how many total correlations were truly calculated to find the “significant” few.

One of the problematic aspects of relating brain measures to behavioral performance is deciding what physiological measures to use. In contrast to behavior, where the SNR50 stands as the single most well-established measure used to characterize the psychometric function, there are many potential electrophysiological measures that might be used to characterize the neural response, and few guidelines on how to select between them. In a study from our laboratory, even beginning with a restricted set of outcome measures ultimately led to making many comparisons, given the large number of test conditions (Billings et al., 2013). Because of the obvious problems with type I error due to multiple testing, we took a modeling approach to predict the behavioral SNR using a weighted average of all the electrophysiological outcome measures, with the individual weights set to produce the best possible predictions. The resulting weighting factors could help us relate the behavioral and electrophysiological data in a more thorough and systematic way.

4.1. Using functional neuroimaging to predict speech understanding in noise

4.1.1. Using electrophysiological methods: CAEPs—Speech-in-noise studies demonstrate that two important factors—SNR and background noise type—interact with age and hearing impairment to affect performance. When signals are presented in continuous background noise, perception is relatively unaffected in normal-hearing individuals until SNR reaches about 0 dB; in contrast, individuals with hearing impairment can require SNRs of 4 to 12 dB to maintain comparable performance with normal-hearing individuals (Crandell and Smaldino, 1995; Killion, 1997; Moore, 1997). The critical point of about 0 dB SNR for normal-hearing individuals is reflected electrophysiologically as well as behaviorally: Billings et al. (2009) found that the most drastic changes in latency and amplitude occurred at SNRs 0 dB. These similarities suggest that certain electrophysiological measures may prove to be good predictors of speech-in-noise understanding. As mentioned previously, a physiological predictor of speech understanding in noise may be especially useful in difficult-to-test individuals.

We created several different prediction models using data from Experiments 1 and 3 to investigate the feasibility of using electrophysiology to predict speech understanding in noise. More specifically, we have attempted to use peak amplitudes, peak latencies, and area measures to predict the SNR50 that was determined using word and sentence tests. Partial least squares (PLS) regression was used to develop the prediction models. PLS regression has the advantage of making judicious use of many correlated predictors, in this case the CAEP measures, by determining weighted linear combinations of independent variables that are optimized to best predict the dependent variable(s). The accuracy of these predictions was quantified in terms of root-mean-square prediction error (RMSPE), with prediction error being defined as the difference between the predicted SNR50 and the measured SNR50 for each subject.

Table 1 represents the prediction errors for Experiment 1 (upper portion) and Experiment 3 (lower portion). Our first attempts at developing the predictions model were based solely on young, normal-hearing individuals, as we were interested in determining the accuracy of

predictions under “ideal” conditions—i.e., when speech understanding was good and AEP morphology was robust. Finding prediction errors smaller than 1 dB led us to apply the YNH-trained model to ONH and OHI individuals. The YNH model performed well for ONH (RMSPEs around 2 dB) but poorly for OHI (RMSPEs > 16 dB). Subsequently, an OHI-trained model was developed and resulted in smaller, but still relatively large, RMSPEs of around 6 dB. Given reasonable predictions for the normal-hearing groups, in Experiment 3 we attempted to improve predictions by modifying the noise type and using a babble background noise that would be more representative speech-on-speech listening situations. Instead of models based on a specific participant group, one comprehensive prediction model was created by combining all 30 participants across the three groups into a single data pool; and instead of lengthy custom-designed speech-in-noise tests that required separate function-fitting after the fact to derive SNR50s, we tried using shorter clinical tests that estimate SNR50 directly from the number of correct responses (WIN and QuickSIN). The resulting RMSPEs (1–3 dB across the board) were reduced for the OHI group and remained relatively stable for YNH and ONH groups.

4.1.2. Using other neuroimaging methods—It is worth noting that other neuroimaging-based methods – ones that do not rely on latency, amplitude, or area analysis of waveforms – have also been investigated for possible predictive value. One such approach is based on identifying locations in the brain where changes in activity are correlated with changes in behavioral performance. A second approach is to look at cortical entrainment in the delta-band (1–4 Hz) and theta-band (4–8 Hz) ranges in response to slow fluctuations in the speech envelope. The first approach requires high spatial resolution such as that provided by fMRI, while the second approach depends on high temporal resolution, as can be supplied by MEG or EEG. A review of the literature suggests that there may indeed be potential to develop methods of predicting receptive speech-in-noise performance based on these approaches. However, the story is not at all straightforward, and numerous challenges exist.

First, seemingly slight differences in the task, stimulus, subject population, or method of analysis can result in rather different activation loci being identified – or even the reversal of an effect direction within the same location. For example, Vaden et al. (2013, 2015) found that cingulo-opercular network activity immediately *before* a word-recognition trial was predictive of *better* performance, yet activity in this region *during* the trial was predictive of *poorer* performance. They also found that occipitotemporal regions were predictive of performance for older adults but not for younger ones. It can be especially difficult to draw conclusions in comparing across studies when more than one of these factors changes at a time. For instance, using positron emission tomography Scott et al. (2004) found that cortical processing of multiple speakers in unmodulated noise occurred in the dorsolateral temporal lobe, while Salvi et al. (2002) found that when subjects listened monaurally to sentences in multitalker babble and repeated the last word of each sentence, activation occurred in the right anterior lobe of the cerebellum and the right medial frontal gyrus. With activation loci apparently being so sensitive to particular task variables, it is unclear whether results on any particular test will be generalizable to overall speech-in-noise performance.

Similar issues present themselves in the temporal and spectral domains as well. While the bulk of neuroimaging studies on cortical phase-locking support the idea that more robust entrainment to the speech envelope in the delta band is predictive of increased behavioral performance (Ding et al, 2014; Ding and Simon, 2013; Peelle et al., 2013; Zoefel et al., 2017) while theta-band entrainment is not (Ding et al., 2014; Peelle et al., 2013), these studies focus on younger adults listening in stationary noise – or in quiet with degraded speech signal (as through noise-vocoding). However, in modulated noise, for a subject group with equal numbers of normal-hearing and hearing-impaired participants, Millman et al. (2017) found that larger amplitudes of envelope-locked cortical responses were associated with *poorer* performance, even after controlling for age and hearing loss.

Second, when brain scans are performed concurrently with behavioral testing, there is no way to discern if the same results would have been obtained from the stimuli alone in a passive condition – a property that would be important if aiming for clinical utility with hard-to-test patients. Indeed, per Binder et al. (2004), the cingulo-opercular network that Vaden et al. (2013, 2015) found to be predictive of performance is not associated with the processing of acoustic cues, but with preparation for behavioral task performance in general. Interestingly, areas that *are* widely thought to be involved in processing of acoustic cues and segregation of auditory streams – most notably in the left superior temporal gyrus (STG; Deike et al., 2004; Hwang et al., 2007) – were *not* found to be predictive of performance in the Vaden et al. studies.

Finally, a note on terminology: While the word “prediction” and its variants are used in these MEG and fMRI studies, they generally look at overall trends and correlations between variables in a single data set, without using a validation set or cross-validation methods, and without looking at how close the individual predictions are on average. Furthermore, these studies tend to collect behavioral and neuroimaging data simultaneously on the same stimuli, which allows for very little to be said about generalization to overall speech-in-noise listening performance.

5. Future directions

The existing literature combining electrophysiology and behavior is mainly limited to small-sample, group-design studies and due to the inter-subject variability on AEPs even within same group, they lack statistical power. The results of these studies are important but are somewhat restricted in their generalizability, with results specific to the characteristics of the tested subgroup or stimulus. For example, middle-aged individuals are often not included in group designs studying age. Figure 8 shows the sampled age and hearing impairment participant space for Experiment 1 with the amplitude of the auditory evoked response represented by color brightness. It is clear from this figure that a large portion of the population were not represented in the study (e.g., individuals from 35 to 59 years of age). To establish generalizable effect sizes of age and hearing impairment and to develop useful AEP-based prediction models will require larger and more diverse samples, so that age and hearing thresholds can be treated as continuous covariates rather than being used to separate subjects into discrete (and somewhat arbitrary) groups. With a sample that more fully spans the variable space, the effects of age and hearing impairment will be better characterized,

and prediction models will likely be more generalizable to clinical populations. Such a robust experimental design will maximize our ability to clarify the relationship between electrophysiology and behavior, and support the development of a robust electrophysiology-based prediction model.

Another area of improvement may be in determining how to best characterize the electrophysiological waveform. Currently, peak latency and peak amplitude are most often used to represent the evoked waveform. Perhaps taking the entire waveform into account rather than extracting certain peaks would enable better predictions. Certainly, reducing the need for choosing a peak could improve the utility and ease of a prediction tool. Using area measures was our attempt at a more automated method of characterizing the waveform, but there are also more advanced methods being explored that have been made possible by modern computing power.

Finally, when creating prediction models, overfitting to the data can be a problem. As a model gets more complex, it will likely fit the dataset it was developed on better and better; however, with increasing complexity the model becomes less and less generalizable to other independent datasets (Hastie et al., 2009). Some degree of over-fitting is likely when the model was created using the same sample that is being tested. A better approach is to divide a dataset into two independent samples: a model development sample and a model validation sample. By applying a developed model to an independent validation sample, overfitting can be detected (and then minimized) and generalization can be tested.

Acknowledgments

This work was supported by the United States (U.S.) National Institutes of Health (NIDCD-DC10914 and DC15240) and by the U.S. Department of Veterans Affairs (51K2RX000714 and 1101RX002139). The contents do not represent the views of the U.S. Department of Veterans Affairs or the U.S. government. Special thanks to Garnett McMillan, Von Praska, and Erin Robling for their contributions to figures and tables within this manuscript.

Appendix

Appendix A.

Details for everyday situations presented in Figure 1.

Group Name	Specific Situation	SNR (dB)	Description	Study
Inside				
	Urban inside	9	Conversation in urban household noise	Pearsons et al, 1977
	Suburban inside	14	Conversation in suburban household noise	Pearsons et al, 1977
	Office	13	12×14 carpeted office	Teder, 1990
	Conference room	9	Carpeted conference room	Teder, 1990
Radio/TV				
	Music	11.1	Conversation in a music background	Smeds et al, 2015
	Radio/TV	7.6	Conversation in radio or TV noise	Smeds et al, 2015
Classroom				
	Classroom	7.9	University	Hodgson et al, 1999

Group Name	Specific Situation	SNR (dB)	Description	Study
	Classroom	22	High school, located near moderately busy street	Pearsons et al, 1977
	Classroom	21	High school, located under landing path for LAX	Pearsons et al, 1977
	Classroom	11.5	12 classrooms with ventilation	Markides et al., 1986
	Classroom	-4.5	12 classrooms with student-activity	Markides et al., 1986
Hospital				
	Hospital	11	Patient room	Pearsons et al, 1977
	Hospital	4	Nurses station	Pearsons et al, 1977
Outdoors				
	Urban outside	4	Urban backyard or patio area facing street	Pearsons et al, 1977
	Suburban outside	7	Suburban backyard or patio area facing street	Pearsons et al, 1977
	Outdoors	12.5	Conversing amongst traffic, birds singing, etc.	Smeds et al, 2015
	Outdoors	4	Suburban patio party	Teder, 1990
	Outdoors	6	Lakeshore, moderate wind	Teder, 1990
Restaurant				
	Bar	1	Neighborhood bar, Friday evening	Teder, 1990
	Restaurant bar	7	Restaurant bar, 1/3 full	Teder, 1990
	Small restaurant	5	Small restaurant, 1/2 full	Teder, 1990
	Small restaurant	4.5	Small restaurant, full	Teder, 1990
	Hilton bar	5.5	Hotel lobby bar with piano music	Teder, 1990
Dept. Store				
	Dept. Store	4	Department stores	Pearsons et al, 1977
	Dept. Store	3.8	Conversation at checkout or while shopping	Smeds et al, 2015
Automobile				
	Car	1	1986 Chevrolet Nova at 55 mph, asphalt	Teder, 1990
	Truck	2	1986 Dodge Ram at 60 mph, concrete	Teder, 1990
	Truck	1.5	1986 Dodge Ram at 40 mph, concrete	Teder, 1990
	Car	2	1988 Pontiac Bonneville at 65 mph, asphalt	Teder, 1990
	Car	6	Conversation in moving car	Smeds et al, 2015
Cocktail Party				
	Cocktail party	-2	In halls with a sound-reflecting ceiling	Plomp, 1977
	Cocktail party	1	Only taking into account horizontally-radiated sound	Plomp, 1977
	Babble	4.6	Conversation in multi-talker babble noise	Smeds et al, 2015
Public Trans.				
	Public trans.	-8	Bay Area Rapid Transit system	Pearsons et al, 1977
	Aircraft	-11	Different planes cruising at normal speed/altitude	Pearsons et al, 1977
	727 jet	5	In the cabin while taxiing	Teder, 1990

Group Name	Specific Situation	SNR (dB)	Description	Study
	727 jet	10	In the cabin while descending	Teder, 1990
	Public trans.	4.4	Conversation on train, in the station hall, or close to the platform	Smeds et al, 2015

Appendix B.

Study methodology details for those measurements that are characterized in Figure 1.

Study	Filtering (dB)	Distance	Location of mic(s)	Signal	Noise
Hodgson et al, 1999	A-weighted	1, 2, 4, 8, 16 meters from instructor	Unlisted	Male and female instructors	Classroom noise (i.e., student talking, coughing, chair movement) and noise from ventilation system. Noise from outside the classroom was removed
Markides et al., 1986	A-weighted, fast	2 meters	Center of classroom	Two groups of teachers: schools for the deaf, PHU	Short duration noise (i.e., footsteps inside and outside classroom, banging of doors and desk lids), Non-stationary long duration noise (i.e., primarily chatter of the children), Quasi-stationary noise (i.e., machinery, cars, aircraft)
Pearsons et al, 1977	A-weighted, fast	Corrected to 1 meter	Classroom: two mics placed at different distances from the teacher, and a lavalier microphone worn by the teacher. All other situations: utilized a tape recorder and a single microphone located at the listener's ear	Varied by situation	Varied by situation
Plomp, 1977	Unlisted	1 meter	Unlisted	Talker distanced 1 meter away	Sound-pressure level of one source (average level for randomly oriented talkers)
Smeds et al, 2015	A-weighted	Varied depending on environment	Two microphones: one located next to left ear and one next to right ear	Data collected from hearing aids. If selected segments were judged to contain target speech and were long enough (minimum 2 sec or an entire sentence), they were kept for further analysis. Speech was defined as "target speech" when the informant was directly	Data collected from hearing aids. Background noise categories were sorted based on the median A-weighted noise levels at the better ear

Study	Filtering (dB)	Distance	Location of mic(s)	Signal	Noise
Teder, 1990	A-weighted, slow	~ 1 foot from observer's body	Sound level meter held in front of researcher's chest, angled 45 degrees toward the front	involved in a conversation with the person speaking In most cases, the speech levels recorded were those of non-hearing impaired partner	Varied by situation

References

- Agrawal Y, Platz E, Niparko J, 2008 Prevalence of hearing loss and differences by demographic characteristics among US adults. *Arch. Intern. Med* 168, 1522–1530. [PubMed: 18663164]
- Anderson S, Parbery-Clark A, Yi HG, Kraus N, 2011 A neural basis of speech-in-noise perception in older adults. *Ear Hear* 32, 750–757. [PubMed: 21730859]
- Anderson S, Skoe E, Chandrasekaran B, Kraus N, 2010 Neural timing is linked to speech perception in noise. *J. Neurosci* 30, 4922–4926. [PubMed: 20371812]
- Arlinger S, 2003 Negative consequences of untreated hearing loss: a review. *Int. J. Audiol* 42, 2S17–2S21. [PubMed: 12918624]
- Armitage P, Allen I, 1950 Methods of estimating the LD 50 in quantal response data. *J. Hygiene* 48, 298–322. [PubMed: 20475815]
- Bennett KO, Billings CJ, Molis MR, Leek MR, 2012 Neural encoding and perception of speech signals in informational masking. *Ear Hear* 33, 231–238. [PubMed: 22367094]
- Bertoli S, Smurzynski J, Probst R, 2005 Effects of age, age-related hearing loss, and contralateral cafeteria noise on the discrimination of small frequency changes: Psychoacoustic and electrophysiological measures. *J. Assoc. Res. Otolaryngol* 6, 207–222. [PubMed: 16027962]
- Bidelman GM, Howell M, 2016 Functional changes in inter- and intra-hemispheric cortical processing underlying degraded speech perception. *NeuroImage* 124, 581–590. [PubMed: 26386346]
- Bidelman GM, Villafuerte JW, Moreno S, Alain C, 2014 Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiol. Aging* 35, 2526–2540. [PubMed: 24908166]
- Billings CJ, Dillard LK, Hoskins ZB, Penman TM, Reavis KM, in press A large-scale examination of Veterans with normal pure-tone hearing thresholds within the Department of Veterans Affairs. *J. Am. Acad. Audiol*
- Billings CJ, Penman TM, Ellis EM, Baltzell LS, McMillan GP, 2016 Phoneme and word scoring in speech-in-noise audiometry. *Am. J. Audiol* 25, 75–83. [PubMed: 26989823]
- Billings CJ, Penman TM, McMillan GP, Ellis E, 2015 Electrophysiology and perception of speech in noise in older listeners: Effects of hearing impairment and age. *Ear Hear* 36, 710–722. [PubMed: 26502191]
- Billings CJ, McMillan GP, Penman TM, Gille S, 2013 Predicting perception in noise using cortical auditory evoked potentials. *J. Assoc. Res. Otolaryngol* 14, 891–903. [PubMed: 24030818]
- Billings CJ, Tremblay KL, Stecker GC, Tolin WM, 2009 Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hear. Res* 254, 15–24. [PubMed: 19364526]
- Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD, 2004 Neural correlates of sensory and decision processes in auditory object identification. *Nat. Neurosci* 7, 295–301. [PubMed: 14966525]
- Bunge SA, Kahn I, 2009 Cognition: an overview of neuroimaging techniques. In: Binder MD, Hirokawa N, Windhorst U (Eds.), *Encyclopedia of Neuroscience* (pp. 1063–1067). Berlin, Germany: Springer.
- Campbell J, Sharma A, 2013 Compensatory changes in cortical resource allocation in adults with hearing loss. *Front. Syst. Neurosci* 7, 1–9. [PubMed: 23420631]

- Carhart R, Tillman TW, 1970 Interaction of competing speech signals with hearing losses. *Arch. Otolaryngol* 91, 273–279. [PubMed: 5414080]
- Caspary DM, Ling L, Turner JG, et al., 2008 Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *J. Exp. Biol* 211, 1781–1791. [PubMed: 18490394]
- Crandell C, Smaldino J, 1995 An update of classroom acoustics for children with hearing impairment. *Volta Rev* 1, 4–12.
- Cruikshanks KJ, Wiley TL, Tweed TS, Klein BE, Klein R, Mares-Perlman JA, Nondahl DM, 1998 Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. The epidemiology of hearing loss study. *Am. J. Epidemiol*, 148, 879–886. [PubMed: 9801018]
- Deike S, Gaschler-Markefski B, Brechmann A, Scheich H, 2004 Auditory stream segregation relying on timbre involves left auditory cortex. *Neuroreport* 15, 1511–1514. [PubMed: 15194885]
- Ding N, Chatterjee M, Simon JZ, 2014 Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage* 88, 41–46. [PubMed: 24188816]
- Ding N, Simon JZ, 2013 Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci* 33, 5728–5735. [PubMed: 23536086]
- Dirks DD, Morgan DE, Dubno JR 1982 A procedure for quantifying the effects of noise on speech recognition. *J. Speech Hear. Disord* 47, 114–123. [PubMed: 7176586]
- Dubno JR, Dirks DD, Morgan DE, 1984 Effects of age and mild hearing loss on speech recognition in noise. *J. Acoust. Soc. Am* 76, 87–96. [PubMed: 6747116]
- Folmer RL, Vachhani J, Theodoroff SM, Chung K, Nutt J, 2014 Auditory processing abilities in patients with Parkinson disease. *Neurology* 82, P3.062.
- Frisina ST, Mapes F, Kim S, Frisina DR, Frisina RD, 2006 Characterization of hearing loss in aged type II diabetics. *Hear. Res* 211, 103–113. [PubMed: 16309862]
- Gallun FJ, Diedesch AC, Kubli LR, Walden TC, Folmer RL, Lewis MS, McDermott DJ, Fausti SA, Leek MR, 2012 Performance on tests of central auditory processing by individuals exposed to high-intensity blasts. *J. Rehabil. Res. Dev* 49, 1005–1025. [PubMed: 23341276]
- Glista D, Easwar V, Purcell DW, Scollie S, 2012 A pilot study on cortical auditory evoked potentials in children: Aided CAEPs reflect improved high frequency audibility with frequency compression hearing aid technology. *Int. J. Otolaryngol* 2012, 12 pages, doi:10.1155/2012/982894.
- Gordon-Salant S, Fitzgibbons PJ, 1995 Comparing recognition of distorted speech using an equivalent signal-to-noise ratio index. *J. Speech Lang. Hear. Res* 38, 706–713.
- Groen JJ, 1969 Social hearing handicap: Its measurement by speech audiometry in noise. *Int. J. Audiol* 8, 182–183.
- Hodgson M, Rempel R, Kennedy S, 1999 Measurement and prediction of typical speech and background-noise levels in university classrooms during lectures. *J. Acoustical Soc. Am* 105, 226–233.
- Humes LE, Roberts L, 1990 Speech-recognition difficulties of the hearing-impaired elderly: The contributions of audibility. *J. Speech Lang. Hearing Res* 33, 726–735.
- Hwang J, Li C, Wu. C, Chen J, Liu T, 2007 Aging effects on the activation of the auditory cortex during binaural speech listening in white noise: an fMRI study. *Audiol. Neurotol* 12, 285–294.
- Ikeda K, Hayashi A, Matsuda O, Sekiguchi T, 2010 An ignoring task improves validity of cortical evoked response audiometry. *Neuroreport* 21, 709–715. [PubMed: 20505550]
- Institute of Electrical and Electronic Engineers, 1969 IEEE Recommended Practice for Speech Quality Measures New York: IEEE.
- Jones DA, Victor CR, Vetter NJ, 1984 Hearing difficulty and its psychological implications for the elderly. *J. Epidemiol. Community Health* 38, 75–78. [PubMed: 6231351]
- Keller BK, Morton JL, Thomas VS, Potter JF, 1999 The effect of visual and hearing impairments on functional status. *J. Am. Geriatr. Soc* 47, 1319–1325. [PubMed: 10573440]
- Killion MC, 1997 SNR loss: I can hear what people say, but I can't understand them. *Hear. Rev* 4, 8–14.
- Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, Banerjee S 2004 Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am*, 116: 2395–2405. [PubMed: 15532670]

- Kochkin S, 2010 MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing. *Hear. J* 63, 19–32.
- Kochkin S, 2002 MarkeTrak VI: Consumers rate improvements sought in hearing instruments. *Hear. J* 9, 18–22.
- Kricos P, McCarthy P, 2007 From ear to there: A historical perspective on auditory training. *Semin. Hear* 28, 89–98.
- Kujawa S, Liberman MC, 2009 Adding insult to injury: Cochlear nerve degeneration after “temporary” noise-induced hearing loss. *J. Neurosci* 29, 14077–14085. [PubMed: 19906956]
- Lewis MS, Lilly DJ, Hutter M, Bourdette DN, Saunders J, Fausti SA, 2006 Some effects of multiple sclerosis on speech perception in noise: Preliminary findings. *J. Rehabil. Res. Dev* 43, 91–98. [PubMed: 16847775]
- Maamor N, Billings CJ, 2017 Cortical signal-in-noise coding varies by noise type, signal-to-noise ratio, age, and hearing status. *Neurosci. Letters* 636, 258–264.
- Markides A, 1986 Speech levels and speech-to-noise ratios. *British Journal of Audiology* 20, 115–120. [PubMed: 3719159]
- Millman RE, Mattys SL, Gouws AD, Prendergast G, 2017 Magnified neural envelope coding predicts deficits in speech perception in noise. *J. Neurosci* 37, 7727–7736. [PubMed: 28694336]
- Moore BCJ, 1997 An introduction to the psychology of hearing San Diego: Academic Press.
- Mulrow CD, Aguilar C, Endicott JE, Tuley MR, Velez R, Charlip WS, Rhodes MC, Hill JA, DeNiro LA, 1990 Quality-of-life changes and hearing impairment. *Annals Internal Med* 113, 188–194.
- Parbery-Clark A, Marmel F, Bair J, Kraus N, 2011 What subcortical-cortical relationships tell us about processing speech in noise. *Eur. J. Neurosci* 33, 549–557. [PubMed: 21255123]
- Pearsons KS, Bennett RL, Fidell S, 1977 Speech levels in various noise environments: Environmental health effects research series. Springfield: National Technical Information Service
- Peelle JE, Gross J, Davis MH, 2013 Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex* 23, 1378–1387. [PubMed: 22610394]
- Plomp R, 1977 Acoustical aspects of cocktail parties. *Acoustica* 38, 186–191.
- Prasher D, Mula M, Luxon L, 1993 Cortical evoked potential criteria in the objective assessment of auditory threshold: A comparison of noise-induced hearing loss with Meniere’s disease. *J. Laryngol. Otol* 107, 780–786. [PubMed: 8228590]
- Rapin I, Ruben RJ, Lyttle M, 1970 Diagnosis of hearing loss in infants using auditory evoked responses. *Laryngoscope* 80, 712–722. [PubMed: 5452510]
- Ruggles D, Bharadwaj H, Shinn-Cunningham BG, 2011 Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proc. Natl. Acad. Sci* 108, 15516–15521. [PubMed: 21844339]
- Salvi RJ, Lockwood AH, Frisina RD, Coad ML, Wack DS, Frisina DR, 2002 PET imaging of the normal human auditory system: responses to speech in quiet and in background noise. *Hear. Res* 170, 96–106. [PubMed: 12208544]
- Scott SK, Rosen S, Wickham L, Wise RJS, 2004 A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *J. Acoust. Soc. Am* 115, 813–821. [PubMed: 15000192]
- Seniors Research Group, 1999 The consequences of untreated hearing loss in older persons Washington, DC: National Council on the Aging.
- Smeds K, Wolters F, Rung M, 2015 Estimation of signal-to-noise ratios in realistic sound scenarios. *J. Am. Acad. Audiol* 26, 183–196. [PubMed: 25690777]
- Souza PE, Turner CW, 1994 Masking of speech in young and elderly listeners with hearing loss. *J. Speech Lang. Hearing Res* 37, 655–661.
- Stapells DR, 2002 Cortical event-related potentials to auditory stimuli. In: Katz J (Ed.), *Handbook of Clinical Audiology* (pp. 378–406). Baltimore, MD: Lippincott Williams & Wilkins.
- Studebaker GA, Sherbecoe RL, McDaniel DM, Gwaltney CA, 1999 Monosyllabic word recognition at higher-than-normal speech and noise levels. *J. Acoustical Soc. Am* 105, 2431–2444.
- Teder H, 1990 Noise and speech levels in noisy environments. *Hear. Instrum* 41, 32–33.

- Tillman TW, Carhart R, 1966 An Expanded Test for Speech Discrimination Utilizing CNC Monosyllabic Words. Northwestern University Auditory Test No. 6. Technical Report SAM-TR-66-55 Brooks AFB, TX: USAF School of Aerospace Medicine.
- Tremblay KL, Friesen L, Martin BA, Wright R, 2003 Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear. Hear* 24, 225–232. [PubMed: 12799544]
- Tremblay K, Kraus N, McGee T, 1998 The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *Neuroreport* 9, 3557–3560. [PubMed: 9858359]
- Vaden KI, Kuchinsky SE, Ahlstrom JB, Dubno JR, Eckert MA, 2015 Cortical activity predicts which older adults recognize speech in noise and when. *J. Neurosci* 35, 3929–3937. [PubMed: 25740521]
- Vaden KI, Kuchinsky SE, Cute SL, Ahlstrom JB, Dubno JR, Eckert MA, 2013 The cingulo-opercular network provides word-recognition benefit. *J. Neurosci* 33, 18979–18986. [PubMed: 24285902]
- Van Maanen A, Stapells DR, 2005 Comparison of multiple auditory steady-state responses (80 versus 40 Hz) and slow cortical potentials for threshold estimation in hearing-impaired adults. *Int. J. Audiol* 44, 613–624. [PubMed: 16379489]
- Wilson RH, 2011 Clinical experience with the words-in-noise test on 3,430 veterans: Comparisons with pure-tone thresholds and word recognition in quiet. *J. Am. Acad. Audiol* 22, 405–423. [PubMed: 21993048]
- Wilson RH, Burks CA 2005 The use of 35 words to evaluate hearing loss in terms of signal-to-babble ratio: A clinic protocol. *J. Rehabil. Res. Dev* 42, 839–852. [PubMed: 16680621]
- Yeung KN, Wong LL, 2007 Prediction of hearing thresholds: Comparison of cortical evoked response audiometry and auditory steady state response audiometry techniques. *Int. J. Audiol* 46, 17–25. [PubMed: 17365051]
- Zoefel B, Archer-Boyd A, Davis MH Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech

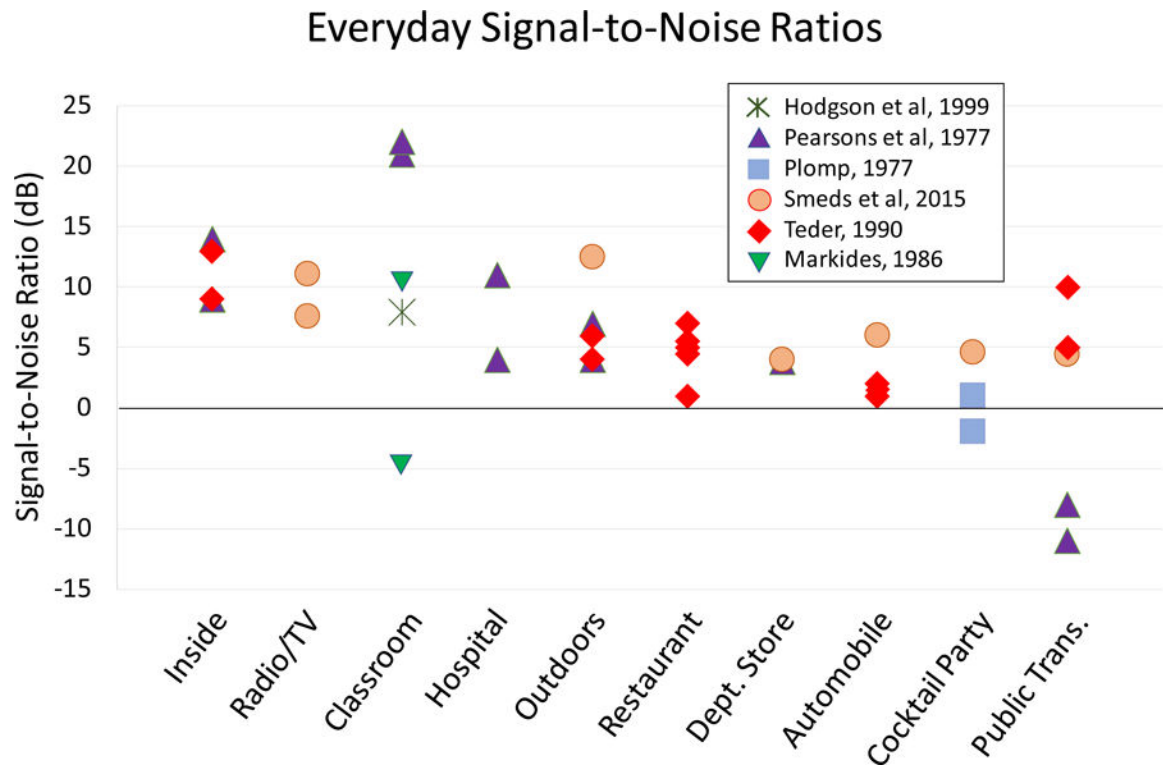


Figure 1. Signal-to-noise ratios (SNRs) measured in everyday listening situations across different studies. Specific environments from the reviewed studies have been grouped into broader categories and arranged according to mean or median SNR (depending on the study), with the highest (most favorable) on the left and the lowest (least favorable) on the right. See Appendices A and B for additional details about the recording procedures and listening situations from each study.

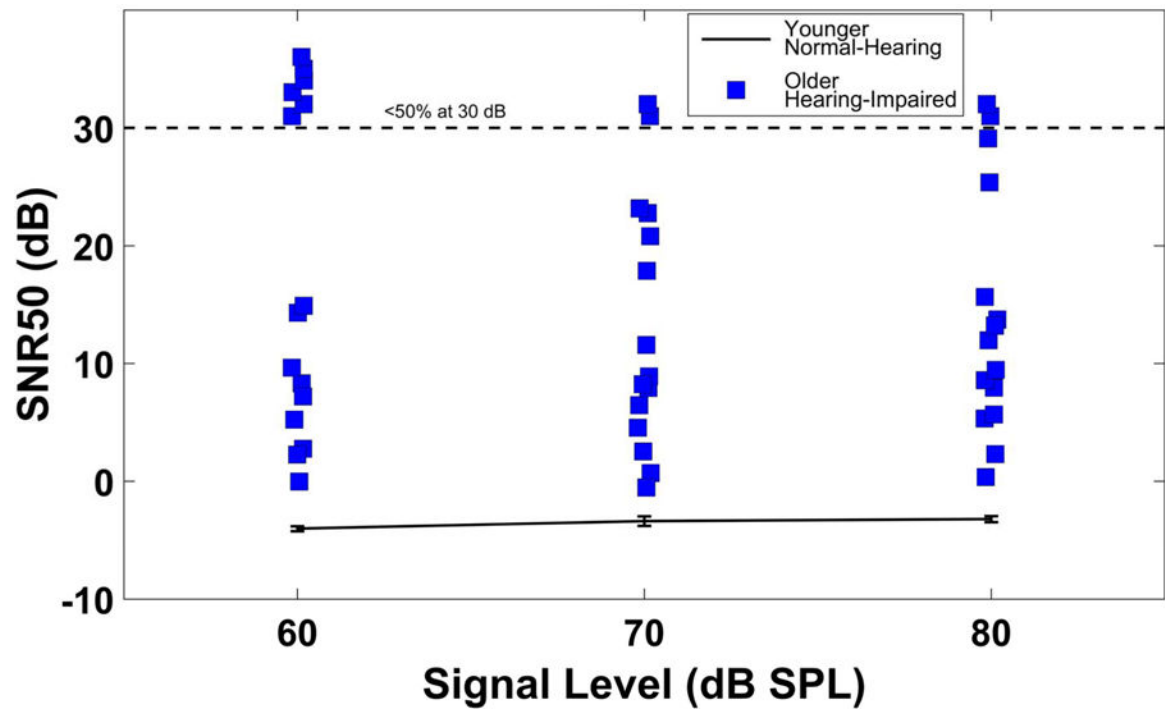


Figure 2. Variability in speech understanding in noise across listeners. SNR50s are plotted for different signal levels for a young, normal-hearing group (error bars show ± 1 standard error of the mean) and for older hearing-impaired individuals. Points above the dashed line represent those individuals who remained below 50% performance at the maximum test SNR of 30 dB.

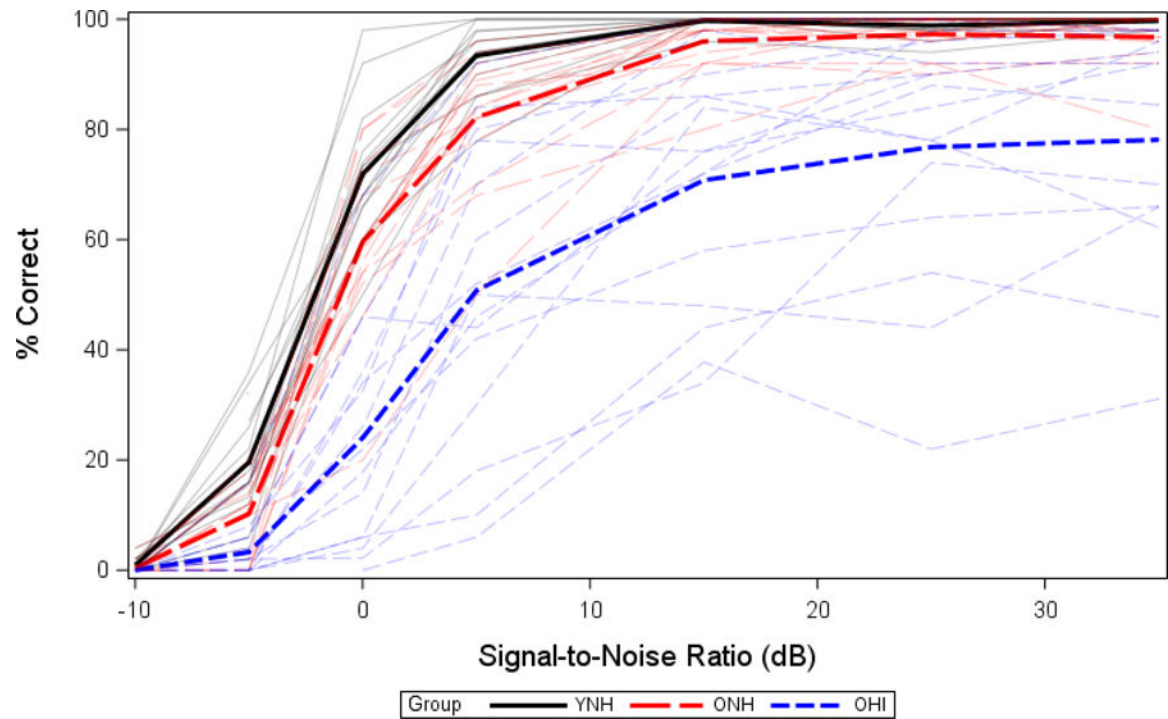


Figure 3. Psychometric functions for IEEE sentence understanding. Younger normal-hearing (solid black lines), older normal-hearing (long-dashed red lines), and older hearing-impaired (short-dashed blue lines) functions are shown both as a group average (thick) and for individuals (thin). (Figure modified from Billings et al., 2015).

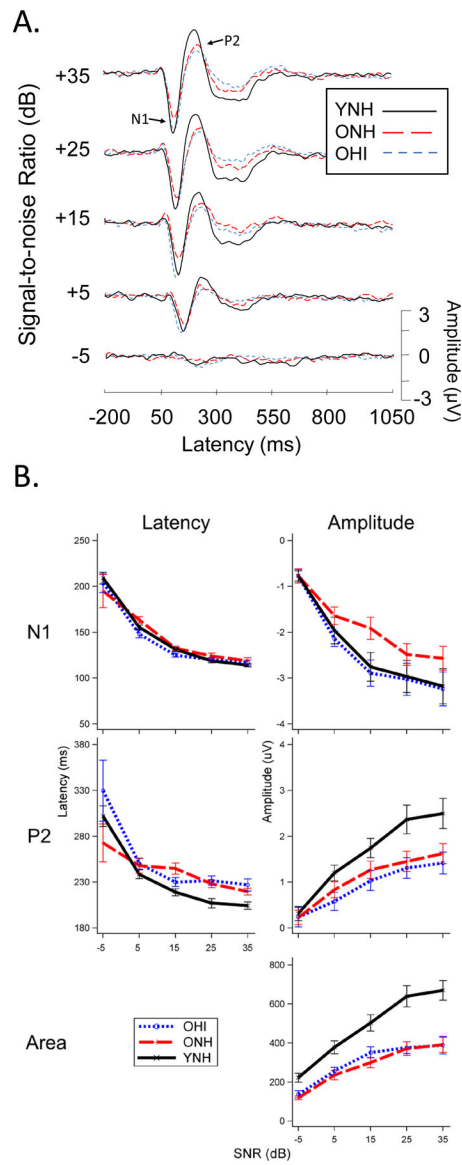


Figure 4.

Experiment 1: CAEP responses at the Cz electrode to speech token /ba/ in steady-state speech-spectrum noise as a function of SNR for three participant groups. A. Waveform morphology is affected by SNR (generally latencies increase and amplitudes decrease as SNR gets worse) and by age and hearing status. B. SNR growth functions for N1, P2, and area CAEP measures, reveal systematic effects of SNR and group. (Figure modified from Billings et al., 2015).

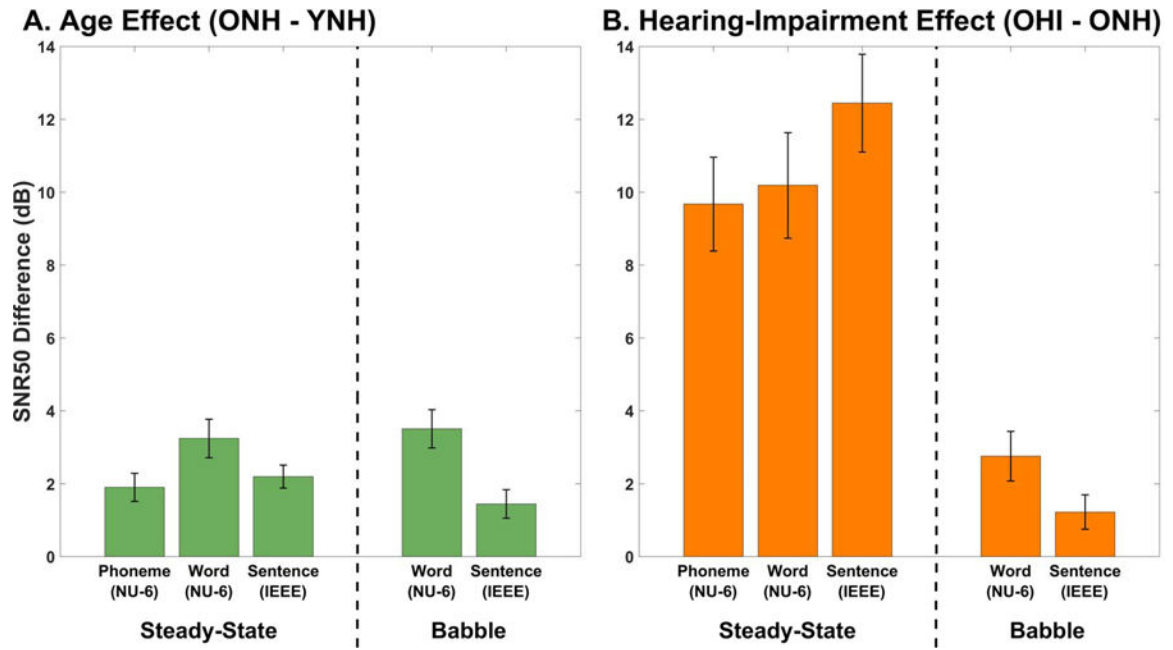


Figure 5.

Age and hearing impairment effects on speech understanding, using data from all three experiments. Error bars represent the pooled mean standard error for the two sets of measurements used to calculate each difference value. A. Age effects on SNR50 values are shown for steady-state speech spectrum noise and four-talker babble using word (scored by word and phoneme) and sentence tests. B. Hearing-impairment effects on SNR50 values are shown as a function of signal and noise type. Hearing-impairment effects are 6–10 dB larger than the age effect for steady-state noise, whereas the effects are similar when four-talker babble is used.

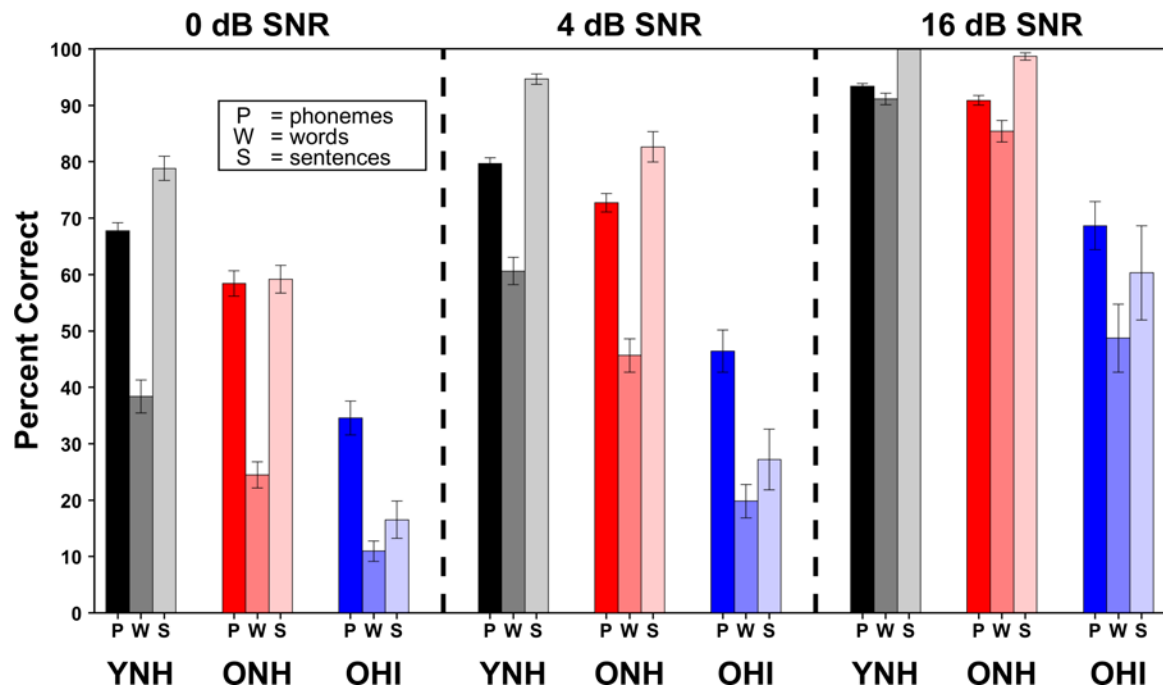


Figure 6. Speech understanding accuracy at selected SNRs, derived from individual fitted logistic functions, grouped by signal type and participant group. All three SNRs are within the range that commonly occurs in everyday listening situations. Age and hearing-impairment effects in terms of percent correct illustrate the difficulty that may be experienced, especially by older hearing-impaired listeners. Error bars represent the standard error of the mean.

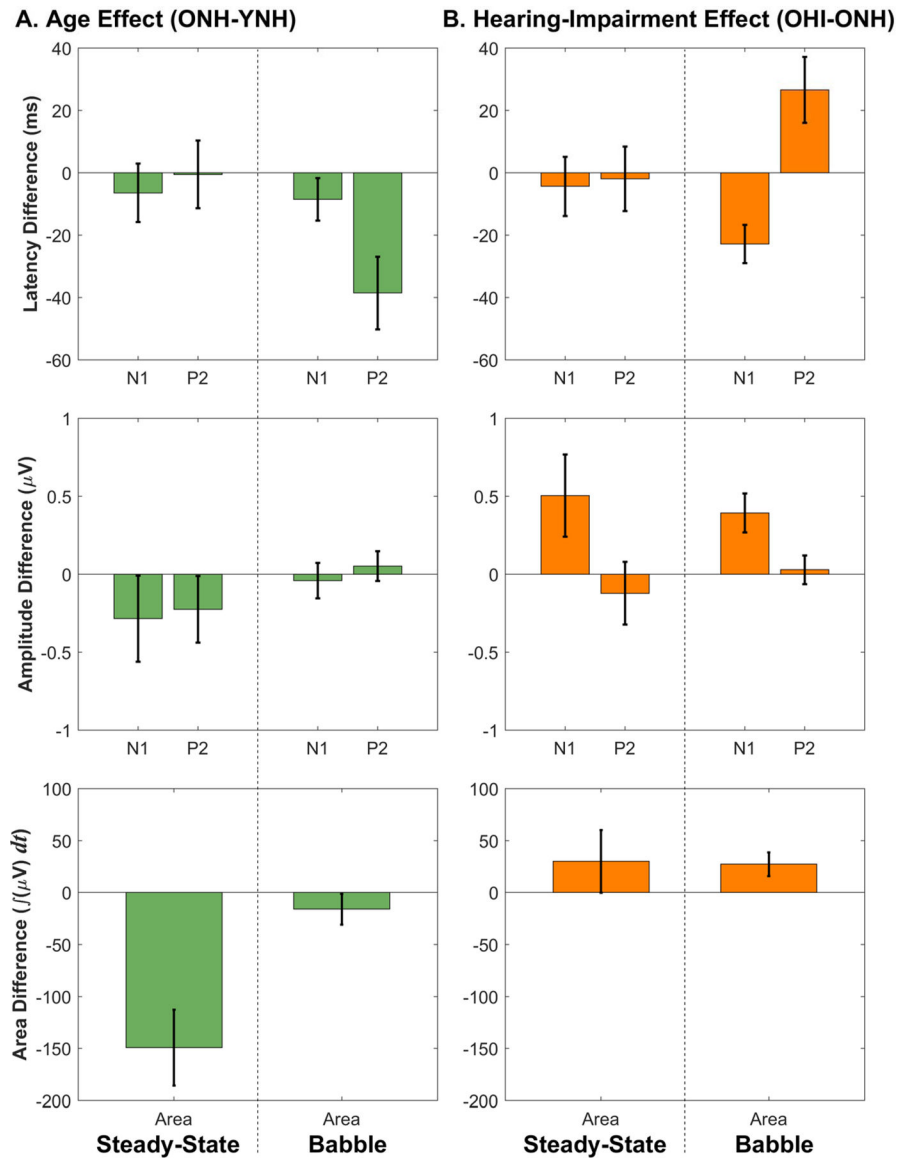


Figure 7. Age and hearing impairment effects on cortical AEPs. Error bars represent the pooled mean standard error for the two sets of measurements used to calculate each difference value. A. Age effects on N1, P2, and area are shown for the speech-syllable stimulus /ba/ presented in steady-state speech-spectrum noise (data from Experiment 1, collapsed across -5, 5, and 15 dB SNRs) and in four-talker babble (data from Experiment 3, collapsed across -3, 3, and 9 dB SNRs). Increased age is mostly associated with decreases in latency, amplitude, and area. Later P2 latencies for YNH participants are also seen. B. Hearing impairment appears to be associated with larger areas and peak amplitudes, shorter N1 latency, and longer P2 latency.

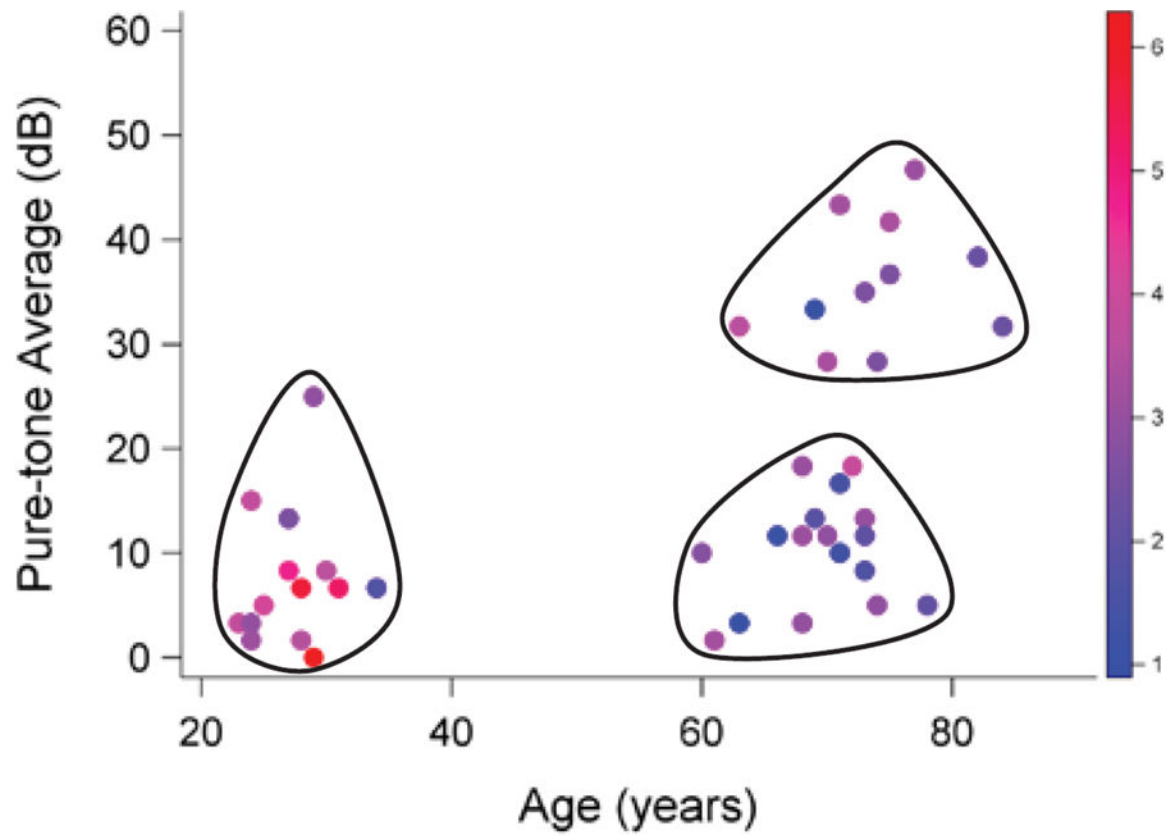


Figure 8.

Age and pure-tone average (PTA) hearing thresholds for Experiment 1 participants. Each dot represents one of the individuals tested in the three participant groups: YNH, ONH, and OHI. Shading of each dot represents the magnitude of the auditory evoked response for each individual. Large portions of the PTA-age space are not represented (e.g., individuals between 35 and 59 years of age). Continuous (non-grouped) sampling of individuals with a wide range of ages and hearing thresholds would improve our estimates of aging and hearing impairment effects and expand the generalizability of a prediction model.

TABLE 1.

Electrophysiology-based predictions of speech-understanding SNR50s for Experiment 1 (steady-state noise; EPs & IEEE sentences at 80 dB C) and Experiment 3 (babble noise; EPs at 65 dB SPL; QuickSIN at 70 dB HL, WIN at 84 dB HL)

Predicted Variable	Model Development			Model Accuracy		
	Training Group	EP Measures	EP SNR (dB)	Test Group	RMPSE (dB)	
Experiment 1[*]						
IEEE SNR50	YNH	5 peak + 2 area	5	YNH	0.7	
			5	ONH	1.9	
			5	OHI	16.7	
		2 area	5	YNH	0.7	
			5	ONH	2.7	
			5	OHI	16.5	
	OHI	5 peak + 2 area	5	OHI	7.8	
			2 area	OHI	6.9	
	Experiment 3[†]					
	QuickSIN SNR50	All subjects	2 peak + 2 area	9	YNH	1.1
				9	ONH	1.2
				9	OHI	1.2
9				All subjects	1.2	
All subjects		2 peak + 2 area	9	YNH	2.9	
			9	ONH	2.1	
			9	OHI	2.3	
			9	All subjects	2.4	

* Billings et al., 2015

† Maamor & Billings, 2016