ABSTRACT

Title of Dissertation: DECODING AUDITORY BRAIN RESPONSES WITH MUTUAL INFORMATION AND THE EFFECTS OF AGING

Peng Zan, Doctor of Philosophy, 2019

Dissertation directed by: Professor Jonathan Z. Simon
Department of Electrical and Computer Engineering

The ability to segregate and understand speech in complex listening scenarios is an inherent property of the human brain. However, this ability deteriorates as the brain ages. The underlying age-related alteration of neural mechanisms is still unclear. Understanding the subcortical and cortical neural mechanisms of auditory processes might be critical in order to get a better understanding of how they degraded by age. Importantly, the likely non-linearity nature of these auditory processes may conceal important internal mechanisms that might not be captured with traditional linear methodology. This thesis develops a novel non-linear approach based on information theory and investigates the non-linear representation of speech in both the midbrain and the cortex. In this dissertation, midbrain and cortical activities from younger and

older listeners are noninvasively recorded with both clean speech (i.e. subjects listening to a single speaker) and with adverse listening conditions (i.e. two competing speakers). Additionally, the effect of informational masking is also investigated. Results from the mutual information analysis suggest an age-related deterioration of the response in the midbrain and a strong effect of the informational masking only in older adults. Conversely, the cortical analysis reveals an exaggerated response in older listeners. Interestingly, this exaggerated response is strongly correlated with behavioral measurements, such as speech-in-noise score and behavioral inhibitory control score. Further analysis also reveals that the exaggerated response in the aging cortex manifests only in the neural representation of the low-frequency speech envelope, while at higher frequencies (60-100 Hz) no differences were seen between younger and older listeners. However, the aging cortex demonstrates neural deficits, at such higher frequency, in suppression of the competing speech in challenging listening conditions, shown by an increasing trend of response level with increasing sound level of the competing speech. In summary, this dissertation develops a novel mutual information approach for analyzing neural recordings, and the results reveal new findings of age-related changes in auditory midbrain and cortical activities.

DECODING AUDITORY BRAIN RESPONSES WITH MUTUAL INFORMATION AND THE EFFECTS OF AGING

by

Peng Zan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Jonathan Z. Simon, Chair
Professor Steve Marcus
Professor Shihab Shamma
Associate Professor Daniel A. Butts
Associate Professor Samira Anderson

Dedication

To my parents, Zan Yuntong and Yao Xiaofang.

Acknowledgements

I would like to thank my supervisor Dr. Jonathan Simon for his guidance and support for more than 4 years. I am really grateful for him introducing me to the world of auditory neuroscience and helping me build up knowledge and interest in science and research. I would like to thank Dr. Marcus, Dr. Shamma, Dr. Butts and Dr. Anderson for serving on my dissertation committee.

I would like to thank Dr. Anderson for important feedbacks on my work, and Dr. Elliot Hong for helpful discussions.

I want to thank Alessandro Presacco and Christian Brodbeck for their insights into and help with data analysis. I am also grateful for Kai Lu and Wanyi Liu from the Shamma Lab for many inspiring discussions, and most importantly, for the friendship over the years of research.

I also want to thank my wife Yaoyuan Liang for the company and being a great source of encouragement and happiness over the years. Lastly, I would like to thank my parents for the endless love and support.

Table of Contents

List of Figures

Chapter 1. Introduction

One of the most remarkable features of the brain is its innate ability to track and process speech in adverse conditions, such as noisy environments. However, this skill tends to deteriorate with age, thus causing older adults to experience significant problems when having a conversation in challenging situations. (i.e. restaurants). These communications difficulties have a strong impact on our society, as the National Institute of Deafness and Other Communication Disorders (NIDCD) has estimated that one third of the U.S. population between age 65 and 74 have hearing problems, and nearly half of those older than 75 have difficulty hearing (NIDCD 2015). Furthermore, age-related hearing deficits contribute to increased risk of depression (Carabellese et al. 1993) and are associated with dementia (Herbst and Humphrey 1980).

Older listeners often report problems listening to speech in noise even when they have clinically normal hearing (Burke and Shafto 2008; Helfer and Freyman 2008). This test for hearing is measured by a tone detection task, where the threshold for pure tones at different levels and frequencies is measured. A subject may be considered to have clinically normal hearing if they have air conduction thresholds $\leq$ 25 dB hearing level from 125 to 4,000 Hz bilaterally (sometimes this threshold might be lowered to 20 dB; also, higher frequencies (i.e. 8,000 Hz) might be considered). Behavioral studies have found age-related deficits in auditory temporal resolution by showing larger gap-detection thresholds on tonal stimulus (Schneider et al. 1994), reduced sensitivity to amplitude modulations (Takahashi and Bacon 1992), and speech-

recognition dysfunction (Frisina and Frisina 1997) for older listeners. Age-related temporal processing deficits in behavioral studies are consistent with observations from neurophysiological studies: the aging midbrain shows a delayed and reduced response to speech syllables (Anderson et al. 2012; Clinard and Tremblay 2013) and clicks (Burkard and Sims 2002). Previous studies have also shown that for younger listeners, cortical responses can demonstrate segregated speech from either a competing speaker (Ding and Simon 2012a) or spectrally matched noise (Ding and Simon 2013), but, unexpectedly, more recent studies have shown that this cortical response is enlarged (exaggerated), not diminished, in older listeners (Lister et al. 2011; Presacco et al. 2016a, 2016b). The neural mechanisms underlying age-related auditory temporal processes deficits have also been investigated in animal studies: aging animals show decreased release of inhibitory neurotransmitters, such as gamma-aminobutyric acid (GABA), in dorsal cochlear nucleus (Caspary et al. 2005; Parthasarathy and Bartlett 2011; Schatteman et al. 2008; Wang et al. 2009), inferior colliculus (IC) (Caspary et al. 1995) and in the auditory cortex (Juarez-Salinas et al. 2010; de Villers-Sidani et al. 2010). In aging rats, altered neural inhibition and functional impairments in the cortex are mostly due to a regulated plasticity change (de Villers-Sidani et al. 2010). Animal studies are also in agreement with human findings in showing an exaggerated cortical response (Hughes et al. 2010).

Despite these studies on aging, it still remains an open question that how aging affects the amount of stimulus information represented in the aging midbrain and the

aging cortex in challenging listening conditions. In this dissertation, this question is studied by adoption of novel informational measures from information theory.

Chapter 2. Background

2.1 Magnetoencephalography (MEG)

Magnetoencephalography (MEG) is a noninvasive electrophysiological technique that records the magnetic field originating from the human brain. MEG signals are mainly generated by postsynaptic currents conducted by apical dendrites of pyramidal cells (Hämäläinen et al. 1993; Levänen 1998). The order of magnitude of the neuromagnetic signals is typically ~10-100 fT (1 fT = 10^{-15} T), which is roughly 10^9 times weaker than the magnetic field generated by the earth (Hämäläinen et al. 1993). The net neural currents in pyramidal neurons flow in the direction normal (perpendicular) to the local cortical surface. Such cortical currents that are also tangential to the skull generate a magnetic flux, perpendicular to the current, which passes through the scalp without distortion (Figure 2.1). Consequently, MEG is most sensitive to tangential cortical currents. However, the main factor affecting MEG sensitivity is actually likely to be source depth not current orientation, so tilted neural currents from sources near cortical surface are also detected by MEG (Hari and Salmelin 2012; Hillebrand and Barnes 2002).

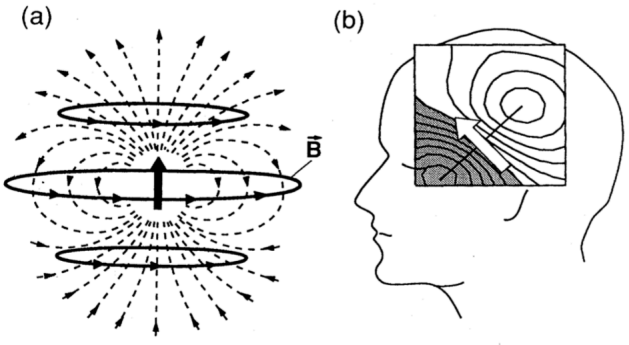

Figure 2.1. Magnetic field of a current dipole. (a) Current dipole (big arrow in the center) in a homogeneous conducting medium. Volume currents in dashed lines and magnetic-field lines **B** in solid lines are produced by the primary current. (b) Example of topographic map calculated from the MEG signals. The geometrically constructed equivalent current source dipole locates in the midway between the field extrema. Adapted from (Hämäläinen et al. 1993).

2.1.1 MEG instrumentation

MEG uses Superconducting Quantum Interference Devices (SQUID) as sensors for recording (Silver and Zimmerman 1965), which offers sufficient sensitivity for detecting neuromagnetic signals at a level of ~10 fT (Ryhänen et al. 1989), without needing any reference (necessary for measuring electrical potential) (Cohen 1972). Modern MEG systems contain helmet-shaped arrays of ~100-300 SQUID sensors, which are immersed into liquid helium to keep the temperature at 4 K (-269 ℃). The whole system is placed in a magnetically shielded room to reduce environmental magnetic interference (Hämäläinen et al. 1993).

The MEG system used in this dissertation is a whole-head 160-channel Kanazawa Institute of Technology (KIT) system, among which 157 are data channels and 3 are reference channels. The data channels are gradiometers that contain pairs of coils and are sensitive to local magnetic field in a certain direction. The reference channels are magnetometers built away from the head to record environmental magnetic field. The system operates at a typical sampling frequency of 1 kHz. This

high temporal resolution makes it suitable to study auditory activity with fast fluctuations in time. Since the magnetic field is not volume conductive, it passes the scalp without distortion. This magnetic property, coupled with effective noise shielding, make MEG feasible for neural source localization.

2.1.2 MEG source localization

Knowing the conductivity of the brain tissue and the neural current generators, the magnetic field outside the brain can be derived by Maxwell's equations and the continuity equation. Based on a quasistatic approximation (Hämäläinen et al. 1993), the magnetic field intensity can be approximated by a linear combination of those created by the neural generators. Several models have been proposed to invert the generative relationship between neural current sources and the MEG observations. One model, the equivalent current dipole model, assumes the source current is localized in one restricted area and finds the best fit that explains the MEG sensor measurements (Williamson and Kaufman 1981). The second model, with a minimal assumption regarding the *a priori* distribution of potential sources, captures the distribution of sensitivity for every magnetometer to neural currents distributed in the brain. In Chapter 5, the second model is used to localize high-gamma time-locked response. There, a source space with 5124 current sources is defined using a boundary element model (BEM). The *lead field* matrix (Mosher et al. 1999) that maps neural currents to MEG sensor measurements is computed based on the electromagnetic conductivity of the brain tissue. Then neural source currents can be estimated by minimum norm estimation (MNE) (Hämäläinen and Ilmoniemi 1994). In Chapter 5, an averaged brain

surface is adopted from FreeSurfer, and co-registration (Figure 2.2) is performed to adjust the size and position of the averaged head to approximately fit the individual's digitized head (Fischl 2012).

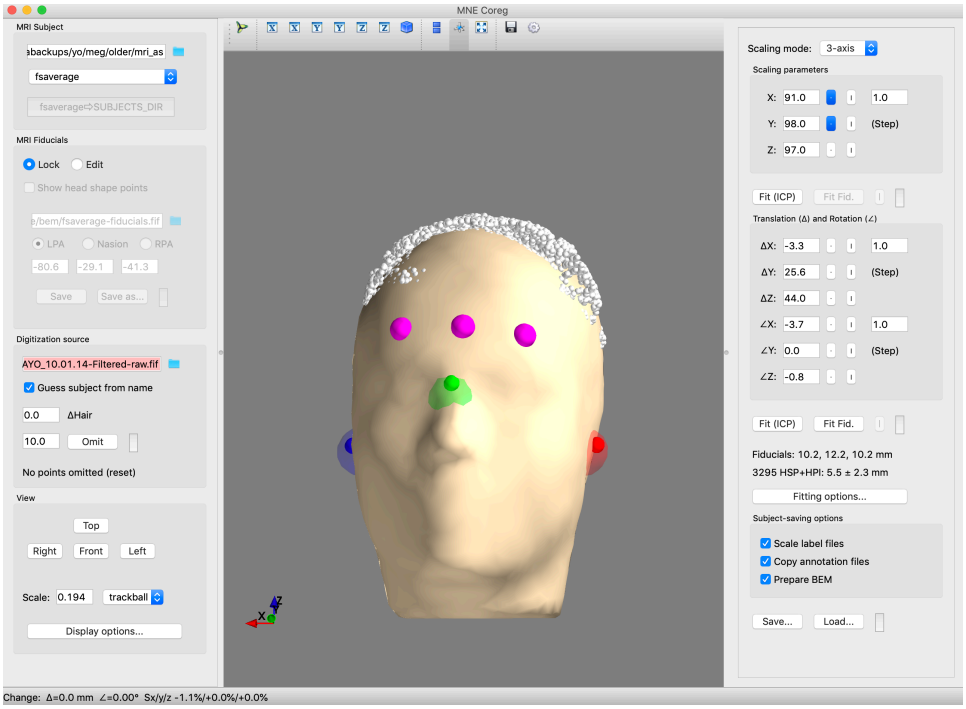

Figure 2.2. MNE co-registration. Both head size and position can be adjusted to fit with the outline shape formed by digitization points.

After alignment of head with the MEG system, the model can finally be formulated, and here is described using the notation of Babadi et al. (2014). Given $\boldsymbol{y}_t \coloneqq [y_{1,t}, y_{2,t}, \dots, y_{N,t}]'$ as the MEG measurement at time *t*, where *N* is the total number of channels, the multi-dimensional observation time series in the interval of [0, T] is $\boldsymbol{Y} \coloneqq [\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_T]$. Similarly, given the neural source currents at time *t* and source *i* to be $x_{i,t}$, the amplitudes of all dipoles at time *t* is $\boldsymbol{x}_t \coloneqq [x_{1,t}, x_{2,t}, \dots, x_{M,t}]'$

and the multi-dimensional neural source currents is $\boldsymbol{X} \coloneqq [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T]$. The MNE model assumes $\boldsymbol{Y} = \boldsymbol{G}\boldsymbol{X} + \boldsymbol{V}$, where $\boldsymbol{G} \in \mathbb{R}^{N \times M}$ is the lead field matrix, and $\boldsymbol{V} \coloneqq [\boldsymbol{v_1}, \boldsymbol{v}_2, \dots, \boldsymbol{v}_T] \in \mathbb{R}^{N \times T}$ is the observation noise matrix. The problem of computing lead field matrix is called *the forward problem*, which takes into account the information of source space set up by BEM model and its conductivity. Given $\boldsymbol{G}$ and $\boldsymbol{Y}$, the estimation of $\boldsymbol{X}$, also known as *the inverse problem*, can be solved by finding the solution to the *l2*-norm optimization problem,

$$\widehat{\boldsymbol{X}}^{MNE}(\boldsymbol{Y}, \boldsymbol{G}, \boldsymbol{C}, \boldsymbol{Q}, \lambda) \coloneqq \arg min_x \sum_{t=1}^{T} \{\|\boldsymbol{y}_t - \boldsymbol{G}\boldsymbol{x}_t\|^2{}_{C^{-1}} + \lambda^2 \|\boldsymbol{x}_t\|_{Q^{-1}}{}^2\},$$

where λ is a scaling factor, $\boldsymbol{C} \in \mathbb{R}^{N \times N}$ is a spatial covariance matrix of sensor space and $\boldsymbol{Q} \in \mathbb{R}^{M \times M}$ a spatial prior covariance matrix which is used to penalize the energy of the estimated sources (here it is taken to be the identity matrix). Given that the neural sources greatly outnumber the sensors, the inverse problem is ill-posed. Therefore, the regulation term restricts solution by limiting the power of sources. The solution can be derived in a closed form:

$$\widehat{\boldsymbol{X}}^{MNE}(\boldsymbol{Y}, \boldsymbol{G}, \boldsymbol{C}, \boldsymbol{Q}, \lambda) = \boldsymbol{Q}\boldsymbol{G}'(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}' + \lambda^2 \boldsymbol{C})^{-1}\boldsymbol{Y}.$$

The MNE-Python toolbox allows co-registration and implements BEM and MNE to solve for forward and inverse solutions (Gramfort et al. 2013, 2014).

2.2 Electroencephalography (EEG)

Electroencephalography (EEG) shares a common physical principle with MEG and provides another non-invasive technique for the measurement of brain activity with temporal resolution comparable to, or even higher than, MEG. Cortical EEG signals

are also mainly generated by postsynaptic currents conducted by dendrites of pyramidal neurons. Different than MEG, EEG measures the electric field potentials induced by neural currents through electrodes (varying from a few to a few hundred in number) fixed to the scalp. Due to the orthogonal orientations of magnetic and electric fields induced by a current dipole, EEG is more sensitive to radial currents, and more readily detects signals from the deep sources of the brain (Hämäläinen et al. 1993; Hari and Salmelin 2012). The recorded voltage oscillates with a range of tens of micro-volts (Bear et al. 2006).

According the Biot-Savart law, the magnetic field is proportional to the inverse of the squared distance from a current source. Additionally, a radial current produces no magnetic field outside a concentrically homogeneous volume conductor (Sarvas 1987). Therefore, far-field responses known to originate from the brainstem may not be detected by MEG but EEG. In this dissertation, two-channel EEG is used to record neural activity from midbrain, an upper brainstem structure.

2.3 Auditory processing

2.3.1 Midbrain frequency following response

Scalp-EEG recorded neural responses (believed to be dominated by brainstem) include the auditory brainstem response (ABR) which is phase-locked to clicks, and the frequency-following response (FFR) which is phase-locked to a periodic sound stimulus such as tones (Galbraith et al. 2000, 2003; Smith et al. 1975), syllables (Anderson et al. 2012), vowels and words (Galbraith et al. 1998, 1997). ABR peaks I

and II with latencies of 1.3-3.5 ms are generated by the auditory nerve (Møller et al. 1988). In contrast to the ABR, the FFR has a generator in rostral brainstem or midbrain (Galbraith 1994). The neurons that generate FFR can synchronize to frequencies from below 100 Hz up to 1500 Hz (Gardi et al. 1979; Stillman et al. 1978). FFR has been shown to be useful in studying temporal processing deficits (Anderson et al. 2012; Presacco et al. 2016a, 2016b). The effects of aging on midbrain FFR are shown in Chapter 3.

2.3.2 Cortical low-frequency modulation

Cortical slow temporal modulations arising from low-frequency modulation of speech have frequencies below 16 Hz (Chi et al. 2005; Rosen et al. 1992). Slow modulations in the delta band (1-3 Hz) reflect phrasal boundaries and suprasegmental prosodic linguistic features (Gandour et al. 2003; Rosen et al. 1992), and frequencies of the theta band (~3-7 Hz) correspond to the average length of a syllable, ~150-300 ms (Greenberg et al. 1996; Poeppel 2003). The cortical low-frequency temporal response function (TRF) to modulations of continuous speech, at rates of 1-12 Hz, is sparse in time, and has response peaks at ~50 ms and ~100 ms. (Ding and Simon 2012a, 2013) The later peak demonstrates top-down attentional gain control, and is strengthened in response to the attended speech. By reconstructing the speech envelope from neural responses, Pressaco et al. (2016a, 2016b) showed an exaggerated responses for older listeners, suggesting an imbalance of neural excitatory and inhibitory mechanisms.

In Chapter 4, cortical low-frequency modulations in the frequency range 1-8 Hz are investigated, and results show age-related exaggeration. Furthermore, the responses in older listeners correlate to behavioral measures of speech-in-noise performance and behavioral inhibitory control.

2.3.3 Cortical high-gamma response

Auditory high-gamma responses are defined as event-related changes in spectral power in the frequency range of 60-150 Hz (Cervenka et al. 2011). Demonstrated by electrocorticographic (ECoG) studies, cortical high-gamma responses (60-150 Hz) occur approximately 75-120 ms after stimulus presentation onset in response to phonemes (Crone et al. 2001), tones (Edwards et al. 2005) and click trains (Brugge et al. 2009; Howard et al. 2000). A recent MEG study shows clear cortical contributions to FFR, which falls into the frequency range of high-gamma, in response to tones with a latency of 48-60 ms (Coffey et al. 2016a). High-gamma responses are associated with multiple functions of auditory processing, including sound discrimination (Crone et al. 2001; Edwards et al. 2005; Fishman et al. 2004), phonological processing (Chang et al. 2010; Steinschneider et al. 2011), auditory selective attention (Herrmann and Knight 2001; Ray et al. 2008), auditory verbal memory (Herrmann et al. 2004; Kaiser et al. 2003) and auditory comprehension (Towle et al. 2008). In Chapter 5, cortical high-gamma responses are studied by analyzing MEG recordings, which demonstrate age-related deficits for older listeners.

2.4 MEG signal processing

2.4.1 Denoising signal by time-shifted principle component analysis (TSPCA)

Though recorded in heavily shielded rooms, MEG signals are affected by environmental noise, such as magnetic field generated by power line with frequencies at 60 Hz and its harmonics. Physiological sources such as heartbeat, eye-blink and muscle activity also induce noise in MEG recordings (de Cheveigné and Simon 2008a). Time-shifted principle component analysis (TSPCA) is a denoising algorithm used to remove environmental noise (de Cheveigne et al., 2007). The algorithm utilizes time-shifted signals from reference channels to regress out the environmental noise from signals recorded by data channels.

2.4.2 Extracting auditory component by denoising source separation (DSS)

Due to the presence of environmental noise, physiological noise and background brain activity, the auditory responses may be hidden in noise. Denoising source separation (DSS) is a blind source separation algorithm that can be used to extract auditory component(s) from noisy MEG signals. The response patterns, repeatable over trials, can be extracted by DSS (de Cheveigné and Simon, 2008; Sarela and Valpola, 2005). The algorithm computes a set of spatial filters to project MEG signal from sensor space to a virtual sensor space where signals are ranked based on descending reproducibility.

In this dissertation, a bias function is defined by filtering signals into a frequency band of interest (1-8 Hz in Chapter 4 and 60-100 Hz in Chapter 5; details follow in the corresponding chapters) and averaging over epochs to compute stimulus-evoked response. The detailed algorithm is described in (de Cheveigné and Simon 2008b).

2.4.3 Temporal response function (TRF)

The temporal response function (TRF) models cortical modulation of speech envelope as a finite impulse response (FIR) filter, and is characterized by sparse peaks around 50 ms and 100 ms (Ding et al. 2014; Ding and Simon 2012a, 2013). The TRF can be estimated by solving an *l2*-norm optimization problem of minimizing squared error of the estimated response, often regularized by sparsity of the linear kernel. The boosting algorithm may be used for its estimation (David et al., 2007; Friedman et al., 2000).

2.5 Information theory in auditory research

2.5.1 Mutual information applications and interpretations in auditory research

Established by Claude E. Shannon (Shannon, 1948), information theory laid the foundation of communication systems and the information era. Entropy and mutual information are two fundamental concepts that provide mathematical representations of information (Cover and Thomas 1991). Rieke et al. were among the first researchers to apply information theory into auditory research, using mutual information to reveal the transmission rate in the low-frequency fibers of the auditory periphery in bullfrog

(Rieke et al. 1995). Mutual information has also been used to characterize the amount of information encoded in spike trains in auditory neurons in animal studies (Brenner et al. 2000). Slee et al. (2005) investigated whether a simplified model for responses of nucleus laminaris neurons in chick embryos can capture all the stimulus features relevant for spiking. They compared the mutual information between the stimulus and response and the mutual information between stimulus and response generated by the reduced models; one model is better than the other if the former gives a mutual information value closer to the full mutual information than the latter. Chase and Young (2005) artificially modified different acoustic cues such as interaural level difference, interaural time difference and spectral notches and compute mutual information between responses of inferior colliculus (IC) neurons in cats and stimuli with different combinations of cues to study information interaction between these acoustic features, and their results suggest that IC neurons integrate information from multiple input streams. Other studies implementing mutual information include comparing between IC responses to artificial sound and natural sound in cats (Escabí et al. 2003), characterizing selectivity of neurons, in the midbrain, primary forebrain and secondary forebrain areas, to the natural sound ensembles, in zebra finch (Hsu et al. 2004), and various studies based on mutual information encoded in spiking trains in neurons in the auditory cortex in cats (Furukawa and Middlebrooks 2002; Middlebrooks et al. 1994; Stecker et al. 2005). A good review covering implementations of information theory in auditory research, especially for the decoding of spiking patterns, can be found in Nelken and Chechik (2007).

Subsequent research has applied information theory to MEG responses from human auditory processing of continuous speech, examining the stimulus information encoded in the phase of the responses in frequency sub-bands of delta (1-3 Hz), $theta_{low}$ (3-5 Hz) and $theta_{high}$ (5-7 Hz) (Cogan and Poeppel 2011). Their results suggest that each frequency sub-band processes independent information. In this study, the Hilbert phase (Bedrosian 1963) of the response in each sub-band is binned into 4 bins, and a combination of two frequency sub-bands form a 16-bin histogram. Mutual information is then estimated based on the assumption of a uniform distribution for the stimulus and the probability distribution formed by 16-bin histogram for the phase of the response. The authors interpret the mutual information as the average amount of information that a single response provides about the stimulus. In Chapter 3 of the dissertation, the same approach is followed but for midbrain responses recorded by EEG. There the mutual information between stimulus and response amplitude (not just phase) is also estimated. The mutual information can then be interpreted as the average amount of information that a single amplitude or phase response provides about the stimulus. In Chapter 4, the mutual information approach is modified to reveal phase-locked temporal information, i.e., to create a mutual information analog of the TRF. Here, by binning both speech envelope (1-8 Hz) amplitude and response amplitude and creating a 64-bin histogram, temporal mutual information can be estimated by shifting the response by different time lags. Following the same interpretation, the mutual information quantity, at a specific time lag reference to stimulus onset, can be interpreted as the amount of information contained in a single response, at the latency, about the speech envelope. Chapter 5 uses similar approach as Chapter 4 but studies

the amount of information contained in a single response in high-gamma band (60-100 Hz) about stimulus waveform in the same frequency band, and the effects of aging.

2.5.2 Mutual information estimation

While mutual information does naturally apply to continuous random variables, when used in data analysis, in practice, the continuous values are binned, so here the stimulus and response are quantized into be discrete random variables. Mutual information between two random variables, X and Y is defined by the following equation,

$$I(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right),$$

where $p(x,y)$ is the joint probability distribution function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively. Based on definition of entropy,

$$H(X) = -\sum_{x\in X} p(x)\log p(x),$$

mutual information is equivalent to

$$I(X;Y) = H(Y) - H(Y|X),$$

where $H(Y)$ is the entropy of response Y, $H(Y|X)$ is the conditional entropy of response Y given stimulus X. Detailed computation can refer to methods sections of Chapters 3 and 4.

2.5.3 Stimulus information contained in phase of the response

The phase of response contains information of linguistic components such as syllable, word and prosody (Cogan and Poeppel 2011). Other studies have shown that phase is important in speech discrimination (Luo and Poeppel 2007) and auditory learning (Luo et al. 2013). In Chapter 3, the aging midbrain demonstrates significantly lower phase-locking values than younger, especially in the steady-state response region (corresponding to the response to the vowel), suggesting that loss of reliable response phase is important in aging. Therefore, in Chapter 3, the mutual information between stimulus and Hilbert phase of response is estimated following methods reported by Cogan and Poeppel (2011).

2.5.4 Temporal mutual information function (TMIF)

The temporal mutual information function (TMIF) is a mutual information analog of the TRF (Ding and Simon, 2012), but characterizes the time-locked response by a non-linear measurement. For low-frequency (1-8 Hz) response, a typical TRF contains peaks at latencies of about 50 ms (M50) and 100 ms (M100), which suggests that TMIF may also have higher level of responses at about 50 ms and 100 ms (Chapter 4). For high-gamma (60-100 Hz) response in Chapter 5, TMIF demonstrates a single peak around 50 ms. In this dissertation, the TMIF of the first 6 DSS components are computed, but only the TMIF for the first DSS component is presented. The details of TMIF estimation can refer to Chapter 4.

2.5.5 TMIF in source space

MEG sensor-space data can be projected into neural source space by source localization using MNE (2.2.3), and TRFs can be estimated for each neural source (Brodbeck et al., 2018a). By determining the TRF for the speech envelope, it has been shown that the age-related exaggerated response to speech originates from an early response in higher auditory cortex (Brodbeck et al., 2018b). In Chapter 5, the TMIF is estimated for all neural sources, and response significance is tested in auditory cortex.

2.5.6 Different choices of probability distributions

In Chapter 3, the stimulus X is the amplitude at each time point. The probability distribution of x is unknown, but here assumed to be uniformly distributed across time points ($p(x) = \frac{1}{T}$, a constant, where T is the sample size of X, so each bin contains roughly the same number of stimulus value instances) for two reasons. First, when the actual stimulus distribution is unknown, this assumption minimizes estimation bias (Nelken and Chechik 2007). Second, while there is not yet evidence for any particular distribution (e.g., Gaussian or Laplacian), the assumption of uniform distribution was employed for stimulus amplitude by Cogan and Poeppel (2011) with encouraging results. In Chapter 4 and Chapter 5, in order to examine the temporal information time-locked to stimulus, the stimulus is also binned. The stimulus is not assumed to be uniform because otherwise the estimated TMIF would be approximately flat in time.

2.5.7 Mutual information vs. linear measures

In this dissertation, mutual information analysis results have revealed new findings that are otherwise hidden by linear measures. Chapter 3 shows a significant effect of the informational masking only for older adults, a results that was not found in a previous that used the same data set (Presacco et al. 2016b). In Chapter 4, the neural response, represented by non-linear measure of TMIF, shows significant exaggerations in early (~50 ms), middle (~100 ms) and late (~200 ms) responses for older listeners. However, linear representations show only exaggeration with a latency of ~50 ms (Brodbeck et al. 2018a). Additionally, this non-linear measure shows strong prediction power regarding behavioral scores, such as Flanker inhibitory control and speech intelligibility, while the linear approach showed limited prediction power to Flanker inhibitory control (significant only when average across all conditions) and no correlation to speech intelligibility (Presacco et al. 2016b). Chapter 5 generalizes TMIF to high-gamma frequency band. The results show deficits of selective attention in challenging listening conditions in older listeners, a result that has not yet been seen by using alternative algorithms.

Chapter 3. Mutual information analysis in the auditory midbrain and the effects of aging

Note: this chapter was recently published as: Zan, P., A. Presacco, S. Anderson and J. Z. Simon (2019) Mutual Information Analysis of Neural Representations of Speech in Noise in the Aging Midbrain, J Neurophysiol. https://doi.org/10.1152/jn.00270.2019

3.1 Introduction

Understanding speech in the presence of background noise becomes more challenging as humans age. Older listeners often report problems in listening to speech in noise even with clinically normal hearing sensitivity (Burke and Shafto 2008; Helfer and Freyman 2008). Behavioral studies have revealed age-related temporal processing deficits in a number of auditory tasks, such as pitch discrimination (Fitzgibbons and Gordon-Salantt 1996), gap-in-noise detection (Fitzgibbons and Gordon-Salant 2001) and recognition of speech in noise (Frisina and Frisina 1997; Gordon-Salant et al. 2006; He et al. 2008; Schneider and Hamstra 1999). These results suggest a temporal processing degradation in the auditory pathway, consistent with observed age-related changes in response latency and strength in midbrain (Anderson et al. 2012; Burkard and Sims 2002; Clinard and Tremblay 2013) and cortical evoked responses (Lister et al., 2011; Presacco et al., 2016a, 2016b).

The neural mechanism underlying age-related temporal auditory process deficits has also been investigated in animal studies: decreased release of inhibitory neurotransmitters, such as gamma-aminobutyric acid (GABA), in dorsal cochlear

nucleus (Caspary et al. 2005; Parthasarathy and Bartlett 2011; Schatteman et al. 2008; Wang et al. 2009), inferior colliculus (IC) (Caspary et al. 1995) and auditory cortex (Juarez-Salinas et al. 2010; de Villers-Sidani et al. 2010) have been found in aging mammals. Because the spectro-temporal fine structure of speech is encoded by synchronous neural firing in midbrain, and the accurate processing of rapid fluctuations depends partly on neural inhibitory mechanisms, the representation of speech there also may deteriorate as a result of greater variability of neural firing (Walton et al. 1998; Yang et al. 1992) or loss of neural inhibition (Caspary et al. 2005, 2006; Walton et al. 1998). The midbrain frequency-following response (FFR), which tracks periodic components of speech or other sounds, may be detrimentally affected by the resulting neural jitter. In older listeners, jitter may be more prevalent than in younger listeners, as reflected by a decreased inter-trial response consistency (Anderson et al. 2012), or, as we hypothesize here, by increased entropy and decreased mutual information as defined in the context of information theory (Cover and Thomas 1991; Shannon 1948).

Mutual information, in particular, can be interpreted as a reduction in auditory response variability due to the presentation of a stimulus (Nelken and Chechik, 2007). It has been used to estimate transmission rates in the low-frequency fibers of the auditory periphery in bullfrog (Rieke et al. 1995), and applied to magnetoencephalography (MEG) auditory responses to continuous speech (Cogan and Poeppel 2011). Auditory information transmitted from midbrain to auditory cortex has been observed to show greater redundancy in older listeners compared to younger listeners (Bidelman et al. 2014). However, given that older listeners have a weaker

midbrain response than younger listeners (Presacco et al. 2016a, 2016b), it remains an open question whether the aging midbrain itself processes more information or less information than younger listeners.

Here in this study, mutual information is calculated between stimulus waveform and the amplitude and phase of response. For stimulus X, band-passed stimulus waveform is used, and the distribution is assumed to be uniform in time points. For response amplitude, band-passed response of 2,000 trials is used, two consecutive trials with opposite polarities are averaged together. The probability distribution of response Y, P(Y), is estimated by binning all samples from 1,000 trials of response with averaged polarities. The conditional distribution of P(Y|X) is estimated by 1,000 samples from Y at time point *t*, i.e., $p(y|x_t)$.

The current study is a mutual informational analysis of auditory midbrain FFR. A more traditional analysis (evoked response) of this dataset has already been published (Presacco et al., 2016a, 2016b). The goals of this new analysis are: 1) to describe these new and innovative methods in detail, 2) to demonstrate rich examples of their use, and 3) to demonstrate that the results are quite often stronger in statistical power than the more traditional methods. First it is shown that the new analysis replicates the most basic earlier findings, that older listeners' midbrain FFR responses contain less auditory signal information about speech stimuli than younger listeners', at the fundamental frequency (F_0) of the FFR. Then the method is generalized to analysis at the harmonic frequencies, showing that speech information contained there

is similarly degraded with age (and falls off more quickly in frequency), consistent with earlier findings (Anderson et al. 2012). Finally, the results also show that when the speech stimuli are degraded by the addition of a competing talker, the stimulus information contained in the midbrain FFR is more sensitive to informational masking (competing speech in a familiar vs. unfamiliar language) in older listeners than in younger listeners.

3.2 Materials and methods

3.2.1 Subjects

The dataset used in this study has previously been described (Presacco et al., 2016a, 2016b). Seventeen younger listeners (3 men) between 18 and 27 years old (mean $\pm$ SD: 22.23 $\pm$ 2.27) and fifteen older listeners (5 men) between 61 and 73 years old (mean $\pm$ SD: 65.06 $\pm$ 2.30), recruited from the Maryland, Washington D.C. and Virginia areas, participated in the experiment. All subjects had clinically-normal hearing with air-conduction thresholds no greater than 25 dB hearing level (HL) from 125 to 4,000 Hz bilaterally and no interaural asymmetry. All of them were native English speakers and were free of neurological or middle-ear disorders, and none of them spoke or understood the Dutch language. All participants were paid for their participation, and each of them gave written informed consent before the experiment. The experimental protocol and all procedures were reviewed and approved by the Institutional Review Board of the University of Maryland.

3.2.2 Stimuli and EEG recording

The stimulus was a single speech syllable, a 170-ms /da/ (Anderson et al. 2012), synthesized at a 20-kHz sampling rate with a Klatt-based synthesizer (Klatt 1980) with a 100-Hz F_0. The syllable was chosen because it comprises both transient and steady-state components, the stop consonant /d/ is rich in phonetic information, and its perception is sensitive to background noise (Miller and Nicely 1955). Its waveform and spectrum are shown in Figure 3.1. The speech syllable was presented diotically at 75 dB SPL with a repetition rate of 4 Hz. Stimuli were presented with alternating polarities to allow cancellation of potential stimulus artifact by summing the responses to each pair (Aiken and Picton 2008a). The stimulus was presented to subjects both in quiet and in noise. For the noise conditions, a story narrated by a female competing speaker in either English or Dutch was used as a masker (a 1-minute duration segment, continuously looped). The English story was an excerpt from A Christmas Carol by Charles Dickens (http://www.audiobooktreasury.com/a-christmas-carol-by-charles-dickens-free-audio-book/), and the Dutch story was Aljaska en de Canada-spoorweg by Anonymous (http://www.loyalbooks.com/book/Aljaska-en-de-Canada-spoorweg). For each of the two masker types, four signal-to-noise ratio (SNR) levels, +3, 0, -3, and -6 dB SNR, were created by using the logarithm of the ratio between root-mean-squared values of syllable /da/ and the long-duration masking speech. All stimuli were presented by insert earphones (ER1, Etymotic Research, Elk Grove Village, IL) via Xonar Essence One (ASUS, Taipei, Taiwan) using Presentation software (Neurobehavioral Systems, Berkeley, CA). FFRs were recorded at a sampling frequency of 16,384 Hz using the ActiABR-200 acquisition system (BioSemi B.V.,

Amsterdam, Netherlands) with a standard vertical montage of five electrodes (Cz active, forehead ground common mode sense/driven right leg electrodes, earlobe references), and the recorded signal was filtered online by a band-pass filter with a cutoff band of 100 Hz to 3,000 Hz. During the 2-hour recording session, subjects sat in a recliner and watched a silent captioned movie of their choice to facilitate a relaxed but wakeful state. For each of the nine conditions (1 quiet + 2 masker languages × 4 SNRs), at least 2,300 trials of response (to repetitions of syllable /da/) were recorded.

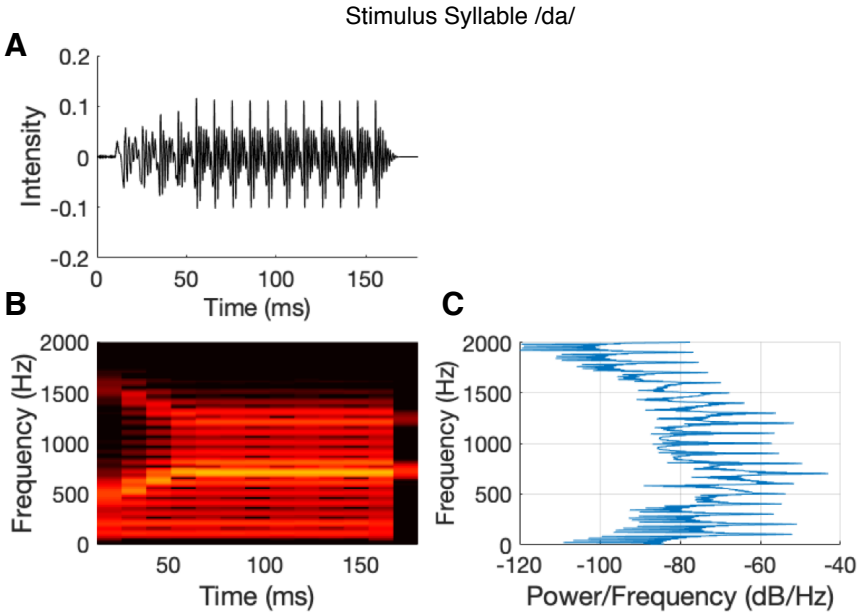

Figure 3.1. A: Stimulus waveform, B: spectrogram, and C: power spectral density of 170-ms syllable /da/. The locations of the horizontal peaks in C indicate that the syllable has a fundamental frequency of 100 Hz with harmonic peaks at its multiples (Anderson et al. 2012, 2013)

3.2.3 Data analysis

Encoding response amplitude

The EEG recordings were first converted into MATLAB format with the function *pop_biosig* from EEGLab (Delorme and Makeig 2004), and all remaining analyses performed in MATLAB (version 2017b; Mathworks, Natick MA). The EEG recordings were band-pass filtered offline, to remove low-frequency neural oscillations, from 70 Hz to 2000 Hz with a linear-phase FIR filter with low-pass transition band of 65-70 Hz and high-pass transition of 2000-2100 Hz. Filter delays were compensated by processing the data in both forward and backward directions, using the Matlab function *filtfilt* (Mathworks, Natick MA). The response of each trial was analyzed in the time window -47 ms to 170 ms with respect to stimulus onset. Within this window, the response of each trial was band-pass filtered with linear-phase FIR filters of order 200, designed using least-square error minimization, into frequency bands centered at harmonics of 100 Hz, i.e., 100, 200, …, 600 Hz, to investigate the midbrain representations of harmonics. Harmonics at or above 700 Hz, the first formant of the steady-state portion of the stimulus, were excluded from analysis. Sweeps with amplitudes larger than ± 30 μV were excluded, allowing 2000 artifact-free sweeps to be used. To eliminate any possible electrical feedthrough artifacts, a 10-ms temporal response function centered at 0 ms with reference to the stimulus onset time was estimated per trial, and its contribution was subtracted from the response (Maddox and Lee 2018). Additionally, since two consecutive sweeps were always presented with opposite polarities, their responses were averaged into one effective sweep, leading to 1000 such pair-averaged sweeps per subject and per condition that were then used for

the analysis; the results for the same sweeps, with artifacts removed but not averaged (2000 per condition) are presented in the Appendix. For each of the two analysis regions, the response waveforms were extracted from each sweep for every subject, for each of the nine conditions and 6 frequency bands.

Under each condition, for each subject and frequency band, a response matrix was obtained of size 1000 trials × T samples where T is the sample length of observation window. In addition to the entire response window 0-170 ms, the responses were also partitioned into two regions based on the acoustic properties of the syllable /da/, i.e., the transition (15-65 ms) and steady-state (64-170 ms) for analysis of masker type influence on the response at 100 Hz. Here $T = 2853$ samples for the entire response region, $T = 804$ samples for the transition region, and $T = 2049$ samples for the steady-state region. The response amplitudes at each sample were subdivided into N bins with the boundaries of the bins chosen so that approximately equal numbers of samples were assigned in each bin; each sample was then associated with its bin index (from 1 to N). The boundaries were chosen individually based on each subject's response. Different values of $N \in \{4, 8, 16, 32, 64, 128\}$ were evaluated to verify a lack of any interaction with age ($F_{(5,180)} = 0.46, p = 0.809$ and $F_{(5,180)} = 0.18, p = 0.970$ by ANOVA test on interaction of $age \times bin\ number$ for amplitude and phase information, respectively). A final choice of $N = 32$ bins was selected as an optimal trade-off between increased resolution between bins and decreased samples per bin due to limited samples (too few bins or too few samples per bin both lead to estimation bias).

The choice of 32 bins gave more than 30 samples/bin, on average, to estimate the conditional probability distribution.

Encoding response phase

For every sweep in each region, the phase for each frequency band was computed by first applying the Hilbert transform to the band-passed signal and then computing the phase of the resultant complex (analytic) signal, i.e.,

$$H\{x(t)\} = IFT\{-i\, sgn(f) FT\{x(t)\})\}, \tag{3.1}$$

where FT is the Fourier Transform, f is the frequency basis of the Fourier Transform, $sgn(f)$ is the algebraic sign of f, and IFT is the inverse Fourier Transform. Then

$$\theta(t) = \angle(x(t) + iH\{x(t)\}). \tag{3.2}$$

The phase-locking value (PLV) of the response in any single band can be computed as

$$PLV(t) = \frac{1}{M}\left|\sum_{j=1}^{M} e^{i\theta_j(t)}\right|, \tag{3.3}$$

where $\theta_j(t)$ is the phase of j^{th} trial at sample time *t*, and *M* is the number of trials.

The set of phase responses $\theta_j(t)$ obtained for each frequency band, were also subdivided in to $N = 32$ bins, analogously to encoding the amplitude response; here the phase samples were divided into bins of width $\frac{2\pi}{32} = \frac{\pi}{16}$ with each sample encoded by its bin index (from 1 to *N*).

Mutual information

Under each condition, for each subject and each frequency band, the mutual information between stimulus and amplitude, and mutual information between stimulus and phase were estimated based on those integer-encoded responses. The response probability distribution was estimated as above (bin index for each of the *T* samples over 1000 trials). The conditional distribution of $P(Y|X)$ was drawn from response samples at the same latency from 1000 trials. The mutual information can then be estimated by the entropy of the response, whether amplitude or phase, minus the conditional entropy of the response given the (uniformly distributed) stimulus:

$$I(X;Y) = H(Y) - H(Y|X), \tag{3.4}$$

where *X* represents the stimulus distribution, and *Y* is the response distribution, whether amplitude or phase. $H(Y)$ is the entropy of the response,

$$H(Y) = -\sum_{y} p(y)\, log\, p(y), \tag{3.5}$$

where *p*(*y*) is the probability of observing the response value *y*. $H(Y|X)$ is the entropy of the response conditioned by the stimulus *X* and is given by:

$$H(Y|X) = \sum_{x} p(x)\, H(Y|X = x) \tag{3.6}$$

where

$$H(Y|X = x) = -\sum_{y} p(Y = y|x)\, log\, p\,(Y = y|x). \tag{3.7}$$

The stimulus X is the amplitude or phase at each time point. The probability distribution of x is unknown, but here assumed to be uniform ($p(x) = \frac{1}{T}$, a constant, so each bin contains roughly the same number of stimulus value instances) for two reasons. First, when the actual stimulus distribution is unknown, this assumption minimizes estimation bias (Nelken and Chechik 2007). Second, while there is not yet evidence for any particular distribution (e.g., Gaussian or Laplacian), the assumption of uniform distribution was employed for stimulus amplitude by Cogan and Poeppel (2011) with encouraging results. Then, equation (3.3) becomes

$$H(Y|X) = \frac{1}{T}\sum_{t=1}^{T} H(Y|X = x_t), \tag{3.8}$$

where x_t is the amplitude or phase bin at sample t.

To illustrate, consider an analysis of the quiet condition over the steady-state region, which encompasses the time window from 64 ms to 189 ms with respect to stimulus onset, i.e., 2,049 samples, giving $T = 2049$ and $p(x_t) = \frac{1}{2049}$ for every value of t.

The distribution of the response, $P(Y)$, is estimated for each subject with all bin-index-encoded samples in each of the 1,000 trials. The conditional distribution of Y given x_t, $P(Y|x_t)$, is estimated with 1,000 samples from trials at time point t. Then the conditional entropy is given by

$$
\begin{aligned}
& H(Y|X) \\
& = -\sum_{t=1}^{T}\sum_{i=1}^{N} p(X = x_t)p(Y = i|X = x_t)\log p(Y = i|X = x_t) \\
& = -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N} p(Y = i|X = x_t)\log p(Y = i|X = x_t),
\end{aligned} \tag{3.9}
$$

where $i \in \{1, 2, \dots, N\}$ is the bin number, and N is the number of bins. The mutual information is therefore,

$$
\begin{aligned}
& I(X;Y) \\
& = -\sum_{i=1}^{N} p(Y = i)\log p(Y = i) \\
& + \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N} p(Y = i|X = x_t)\log p(Y = i|X = x_t).
\end{aligned} \tag{3.10}
$$

Statistics

To examine the effects of aging, frequency, masker type and SNR level, multiple *t*-tests with correction were performed, separately for both amplitude and phase information. To facilitate analyzing the information at fundamental frequency, linear models were constructed to test effects from interactions between aging and other factors, namely masker type and SNR level, with mathematical form $I \sim age \times masker\ type + age \times SNR$. Tests were performed for both amplitude and phase, and for different temporal regions, separately. To test masker type influence within group, the mutual information difference between Dutch and English maskers for each subject was modeled as $I_{Dutch} - I_{English} \sim SNR$, and the positivity of intercept was

tested for both amplitude and phase, and for different temporal regions, separately. The results were justified by *t*-tests on the intercept of linearly fitted regression lines for each subject, and similar analysis for PLV.

Linear models with only fixed effects were analyzed in R (R. Core Team 2017) using the function *lm*, which reports the model significance using an *F*-test on the constructed model vs. the null model with only the intercept, and the significance of influence from fixed-effect factors with separate *t*-tests on the slope of each factor. The assumption of Homoscedasticity of the linear models was examined by global validation of linear model assumptions using toolbox *gvlma* (Peña and Slate 2006) in R. Responses at harmonic frequencies were analyzed using *t*-tests. False discovery rate correction (FDR) (Benjamini and Hochberg 1995), to correct for multiple comparisons, was applied when appropriate.

Where appropriate, *t*-tests for significance are supplemented with effect size (Cohen's *d*) and its 95% confidence interval (CI). When the CI excludes zero, this is alternate evidence that the result is statistically significant (i.e., the effect size is significantly greater than zero at an α level of 0.05). Note, however, that the effect size analysis is not compensated for multiple comparisons even when the *p*-value is.

The effective high-frequency cutoff for any frequency-decreasing statistical measure is defined to be the frequency at which the measure is not significantly higher than the noise floor (pure estimation bias). The noise floor is estimated using the same

mutual information method as used elsewhere, but instead using responses to quiet intervals between stimuli.

3.3 Results

This section reports results from the mutual information analysis of pair-averaged-polarity responses; the analogous analysis based on single sweeps is reported in the Appendix. Because the algorithm takes into account variations across trials, pair-averaging provides less variation and thus higher mutual information. Except for this overall scaling of mutual information, the results are typically comparable.

3.3.1 Information in FFR amplitude

Amplitude information at 100 Hz

For the amplitude response at 100 Hz, to examine masker type and SNR interactions with both age groups, the linear model, $I \sim age \times masker\ type + age \times SNR$ is tested. It is significant ($F_{(5,250)} = 4.99, p < 0.001$ for the entire region; $F_{(5,248)} = 2.93, p = 0.014$ for the transition region; $F_{(5,249)} = 6.11, p < 0.001$ for the steady-state region). Outliers that would otherwise cause the homoscedasticity requirement to be violated are excluded (2 samples from the transition region and 1 sample from the steady-state region, respectively). Results show no significant interactions between *age* and *masker type* ($t_{(250)} = 0.53, p = 0.587$ for the entire region; $t_{(248)} = 0.15, p = 0.884$ for the transition region; $t_{(249)} = 0.29, p = 0.773$ for the steady-state region), or between *age* and *SNR* ($t_{(250)} = 0.79, p = 0.428$ for the entire region; $t_{(248)} = 0.46, p = 0.645$ for the transition region; $t_{(249)} =$

$0.87, p = 0.386$ for the steady-state region). A linear model with no interactions was then constructed and tested, i.e., $I \sim age + masker\ type + SNR$. The model itself is significant ($F_{(3,252)} = 8.05, p < 0.001$, $F_{(3,250)} = 4.84, p = 0.003$, $F_{(3,251)} = 9.96, p < 0.001$ for the entire region, the transition and steady-state regions, respectively). Comparisons between the models show that younger listeners' responses contain significantly more information than older listeners' responses in the whole and steady-state regions ($t_{(252)} = 4.24, p < 0.001$ and $t_{(251)} = 4.99, p < 0.001$ respectively), and that information increases as SNR increases ($t_{(252)} = 2.37, p = 0.018$ for the entire region; $t_{(250)} = 2.86, p = 0.005$ for the transition region; $t_{(251)} = 2.15, p = 0.033$ for the steady-state region).

Since the stimulus has a fundamental frequency of 100 Hz and the phase-locking of FFR is more robust in low frequencies than in high frequencies (Zhu et al. 2013), the 100-Hz FFR may contain significantly more information than its harmonics. To rule out the possibility that significant contributions to mutual information derive from averaging the opposite polarities, the same mutual information analysis is performed on single trials, where similar results are observed (see Appendix). Figure 3.2A displays the mutual information as a function of SNR level. Older listeners not only have a noticeably lower amount of information than younger listeners, but also extract more speech information when the masker is Dutch than for English. To eliminate within-subject variance, a linear regression line of information-by-SNR is fitted for each subject and its y-intercept and slope were analyzed, with results

illustrated in Figure 3.2. A one-tailed *t*-test (younger > older) on the y-intercept shows a significantly larger amount of information in younger than older listeners for the English masker ($t_{(30)} = 1.71, p = 0.048;\ d = 0.75, 95\%\ CI =\ [0.032, 1.469]$). The difference is not significant for Dutch ($t_{(30)} = 1.41, p = 0.102;\ d = 0.51, 95\%\ CI = [-0.195, 1.216]$) (but as will be seen below, it does become significant for higher harmonic frequencies). Both age groups demonstrate decreasing information with worsening SNR: a one-tailed *t*-test on the negativity of the regression slope shows information loss for all cases except for older listeners with the Dutch masker ($t_{(16)} = 3.42, p = 0.002$ and $t_{(16)} = 2.54, p = 0.013$ for younger listeners for English and Dutch maskers, respectively, and $t_{(14)} = 2.32, p = 0.027;\ d = 0.60, 95\%\ CI = [2.55 \times 10^{-5}, +\infty]$ and $t_{(14)} = 2.35, p = 0.059; d = 0.61,\ 95\%\ CI = [1.92 \times 10^{-5}, +\infty]$ for older listeners). No significant difference is seen between the slopes across age groups (one-tailed *t*-test: $t_{(30)} = 1.28, p = 0.106;\ d = 0.55, 95\%\ CI = [-0.155, 1.260]$ for the English masker; $t_{(30)} = 1.20, p = 0.120;\ d = 0.73, 95\%\ CI = [0.018, 1.452]$ for the Dutch masker, though the effect size CI is consistent with significance in the last case.

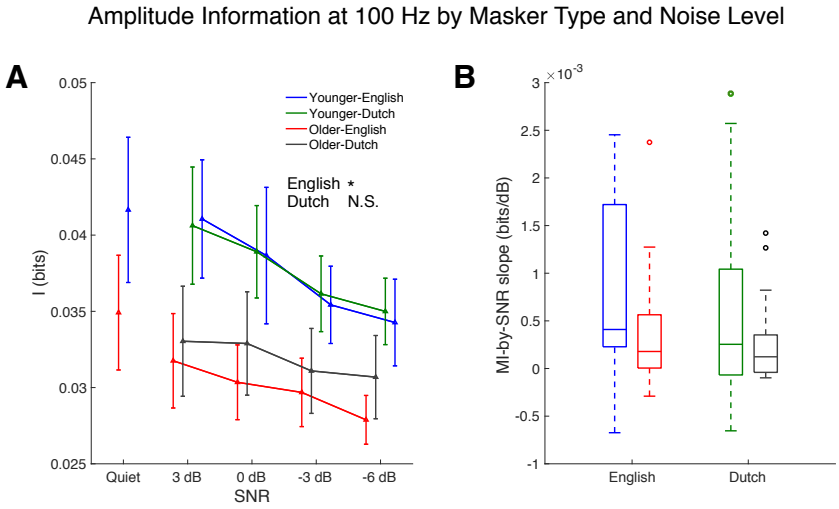

Figure 3.2. Mutual information between stimulus and response amplitude as a function of noise level for each age group and masker condition (masker language). A: Mutual information at the fundamental frequency as a function of noise level (quiet condition and 4 SNR levels) with blue and green for younger listeners (English and Dutch maskers, respectively), and red and gray for older listeners (English and Dutch maskers, respectively). The response in younger listeners conveys noticeably more information than the response in older listeners for the English masker condition, but the difference for Dutch is not significant at 100 Hz. Older listeners show consistently higher mutual information for the Dutch masker than for the English (the younger listeners show no consistent difference), but the difference is not significant at 100 Hz. B: The MI-by-SNR slopes of the previous plots show decreasing trends as SNR worsens, regardless of masker type, for both age groups. Younger listeners show a steeper decrease than older listeners but the difference is not significant at 100 Hz response. Error bars indicate one standard error of mean (SEM). ($* p < 0.05$)

Amplitude information in harmonics of 100 Hz

To analyze aging-associated informational loss for the harmonics (200 to 600 Hz), similar tests are performed on mutual information in responses of these frequencies (analysis stops before 700 Hz, which represents the first formant of the steady-state portion of the stimulus). In each harmonic, a linear regression line of mutual information as a function of SNR is fitted for each subject under each masker type. First the y-intercept of the fitted line at 3 dB is analyzed for group differences (see Figure 3.3).

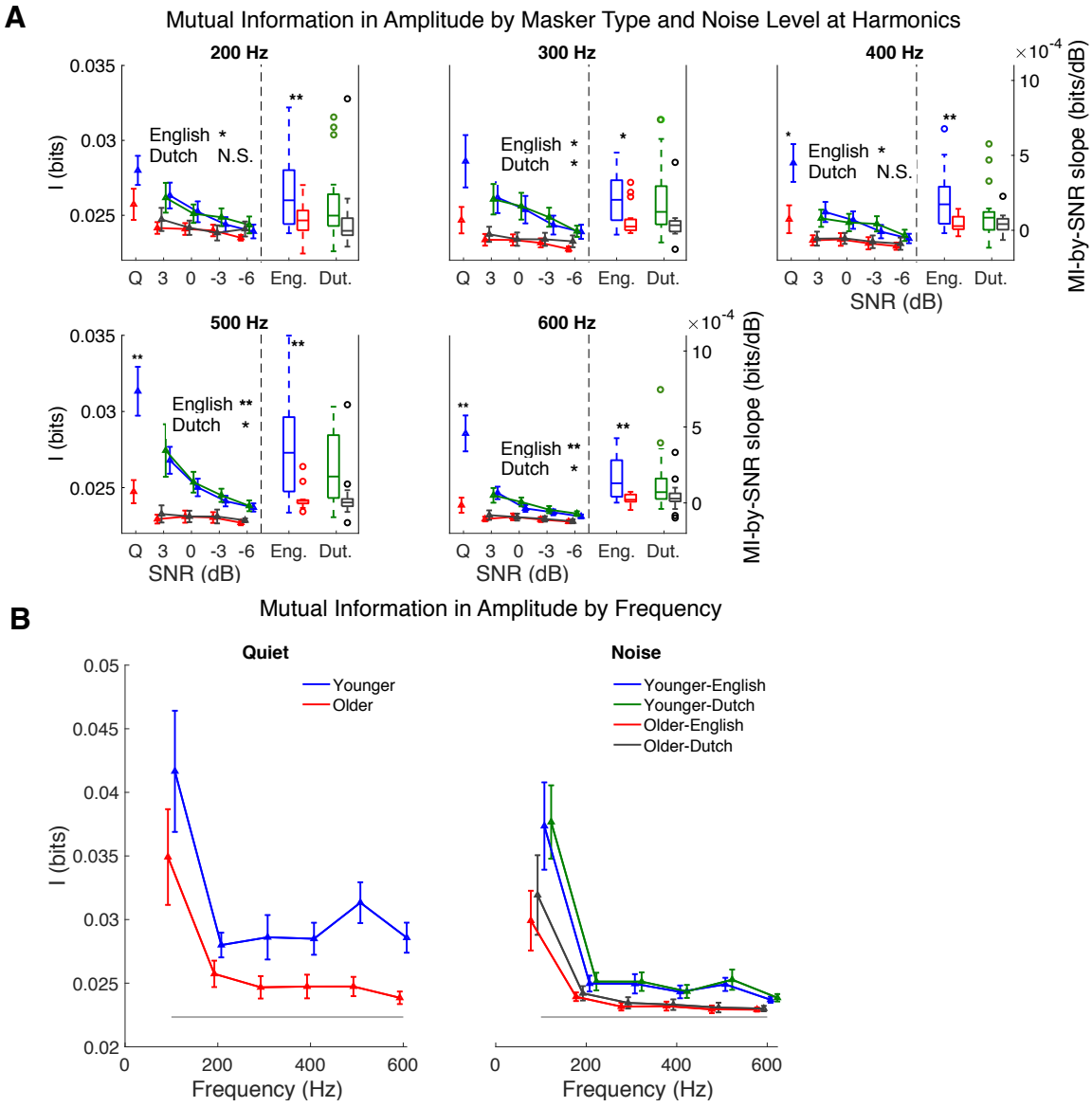

Figure 3.3. A: Mutual information for amplitude across frequency bands from 200 Hz to 600 Hz (separate subplot for each band). Within each subplot, the left panel shows the mutual information as a function of SNR, separately for age group and masker type. For the quiet condition, any asterisks above the error bars indicate the significance levels of group differences; text and any asterisks above the plots demonstrate significance levels of group differences in the corresponding masker

types. Only younger listeners convey a significant amount of information in the higher harmonics. In the right panels, the bar plots depict the linearly fitted decreasing slopes (of the plots shown in the left panel) for the different age groups and masker types. In most bands, the mutual information decreases at a faster rate in younger listeners than in older. B: Overall, both in quiet (left) and averaged over SNR levels (right), mutual information decreases with increasing frequency (except for a single increase at 500 Hz for younger listeners). For older listeners, the decreasing trend in mutual information levels off at 300 Hz, which is lower than the frequency (>600 Hz) at which amplitude information levels off in younger listeners. The lower gray line represents the noise floor. Error bars indicate one SEM. ($* \, p < 0.05, ** \, p < 0.01$)

One-tailed (younger > older) *t*-tests (with FDR correction) and effect size analysis on the y-intercept (corresponding to 3 dB SNR) of the line fit across all SNR levels suggest that the aging midbrain contains significantly less information than the younger midbrain in all frequencies from 100 to 600 Hz in the English masker condition. For *p*-values near 0.05 (see Table 1), effect size analysis is further applied. For the English masker condition, the 100-Hz condition shows consistent significance from both tests ($t_{(30)} = 1.714, p = 0.048;\ d = 0.75, 95\%\ CI = [0.032, 1.469]$), and similarly for the Dutch masker condition at 300 Hz ($t_{(30)} = 2.05, p = 0.049;\ d = 1.236, 95\%\ CI = [0.478, 1.993]$), 500 Hz ($t_{(30)} = 2.27, p = 0.047;\ d = 0.787, 95\%\ CI = [0.0663, 1.507]$) and 600 Hz ($t_{(30)} = 2.26, p = 0.047; d = 1.053, 95\%\ CI = [0.312, 1.794]$) (see also Figure 3.3A). In the English masker

condition, one-tailed *t*-tests on fitted regression line slopes of younger listeners compared to older listeners show significantly steeper slopes for younger listeners compared to older listeners at frequencies from 200 to 600 Hz (all *p*-values are smaller than 0.05). All *p*-values of multiple comparisons are corrected. Overall, higher harmonics contain significant information only for younger listeners, and the difference in information between the two age groups becomes more statistically significant as the observed frequency increases, which is consistent with the linear model analysis, where age × frequency interaction is significant.

Table 3.1. Amplitude information: one-tailed *t*-test (younger > older) results applied to the fitted y-intercepts (3 dB values) and slopes from the linear regression analysis of mutual information (for response amplitude) as a function of SNR, for each harmonic. *p*-values are corrected for multiple comparisons by FDR correction. Boldfaced entries indicate the corresponding tests are statistically significant.

Harmonic (Hz)	Quiet (Y>O)		English masker (Y>O)				Dutch masker (Y>O)			
			y-intercept		slope		y-intercept		slope	
	t(30)	*p*	*t*(30)	*p*	*t*(30)	*p*	*t*(30)	*p*	*t*(30)	*p*
100	1.056	0.150	**1.714**	**0.048**	1.275	0.106	1.405	0.102	1.199	0.120
200	1.542	0.080	**1.965**	**0.035**	**2.737**	**0.008**	1.223	0.115	1.262	0.120
300	1.871	0.053	**2.242**	**0.024**	**2.390**	**0.014**	**2.051**	**0.049**	2.019	0.108
400	**2.271**	**0.030**	**2.261**	**0.024**	**2.835**	**0.008**	1.767	0.066	1.502	0.108
500	**3.449**	**0.003**	**3.671**	**0.003**	**3.677**	**0.002**	**2.268**	**0.047**	1.830	0.108
600	**3.412**	**0.003**	**3.340**	**0.003**	**3.565**	**0.002**	**2.259**	**0.047**	1.629	0.108

Amplitude information frequency limits

As seen in Figure 3.3B, the stimulus information contained in the response amplitude decreases with frequency for both age groups. The frequency-decreasing measure used here is the amplitude information's y-intercept at 3dB of the fitted MI-by-SNR regression line. The frequency bands below 700 Hz are analyzed, separately for different masker types. The measure at 600 Hz for older listeners is not statistically distinguishable from the noise floor ($t_{(14)} = 1.72, p = 0.107$ by one-sample *t*-test). For younger listeners, the measure is significantly higher than the noise floor at all frequencies ($t_{(30)} = 3.34, p = 0.002$ for English masker; $t_{(30)} = 2.26, p = 0.016$ for Dutch masker (younger > older), both at 600 Hz where the lowest information is observed), i.e., the information for younger listeners has not yet reached floor by 600 Hz. In contrast, the cutoff frequency for older listeners is 300 Hz: the information measure at 300 Hz is not significantly greater than that at 600 Hz ($t_{(14)} = 1.32, p = 0.130$ under the English masker; $t_{(14)} = 1.65, p = 0.095$ under the Dutch masker). Therefore, results suggest a lower frequency limit of in amplitude information of 300 Hz for older listeners than that of beyond 600 Hz for younger listeners.

Effect of masker type on amplitude information

As seen in Figure 3.2B, older listeners demonstrate a slower fall-off in amplitude information as a function of SNR when the noise masker is Dutch than for English. To test for any potential amplitude information benefit from the Dutch masker over the English masker, the difference in information between the Dutch and English maskers is calculated for each subject in all SNR levels (for both transition and steady-

state regions), and a linear model of $I_{Dutch} - I_{English} \sim SNR$ shows a significantly positive intercept for older listeners in the transition region ($t_{(57)} = 2.35, p < 0.001$ with 2 samples omitted) but not in the steady-state region ($t_{(56)} = 1.38, p = 0.173$ with one sample omitted). Younger listeners, however, do not show a significant positive intercept in either transition ($t_{(65)} = 1.90, p = 0.061$ with one sample omitted) or steady-state region ($t_{(66)} = -0.60, p = 0.549$). Samples were omitted from the tests to satisfy the homoscedasticity requirement. A regression line was fitted as a function of SNR to reduce within-subject variance. Using a one-tailed *t*-test on the y-intercept (effective mutual information benefit at 3 dB SNR) of the regression line against zero, the mutual information benefit from the Dutch masker over the English masker is significantly higher for older listeners in the transition region ($t_{(14)} = 2.35, p = 0.017$), but not the steady-state region ($t_{(14)} = 1.67, p = 0.058$). No significant benefit is found for younger listeners in either region ($t_{(16)} = 1.17, p = 0.130$ and $t_{(16)} = 0.51, p = 0.307$ for transition and steady-state region, respectively). The regression slope is not significantly positive or negative for either group ($p > 0.05$ by two-tailed *t*-tests), as seen in the bar plots in the right panels of figure 3.4C and 3.4D.

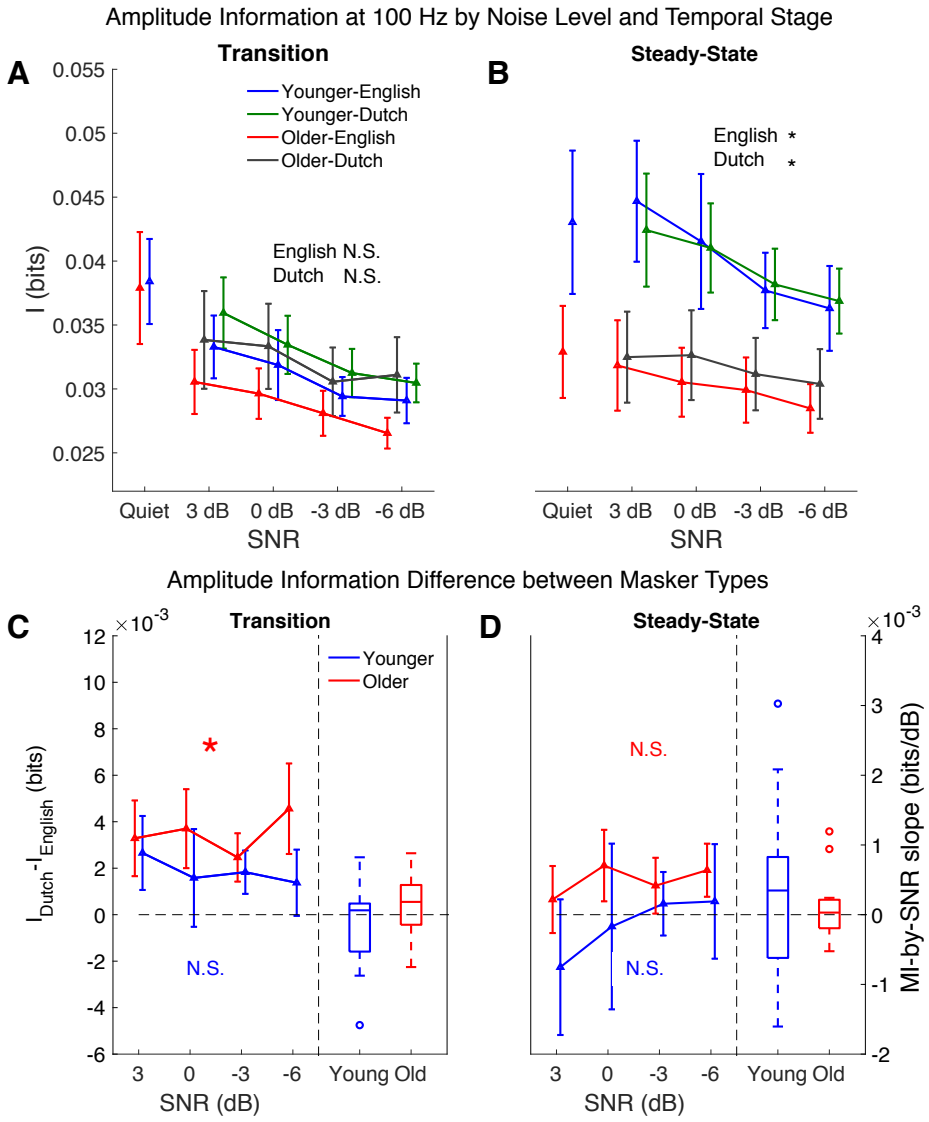

Figure 3.4. Mutual information of amplitude response by masker type and response region for younger listeners in blue (English) and green (Dutch) and older in red (English) and gray (Dutch). A and B demonstrate the mutual information as a function of SNR in the transition and steady-stage regions, respectively. In the steady-state region, group differences are significant for both masker types, indicated by asterisks. C and D illustrate the mutual information difference between masker types (denoted $I_{Dutch} - I_{English}$) in the transition region and steady-state region,

respectively. In each plot, the left panel displays information as a function of SNR, and the right panel displays a bar plot showing the slopes of the linear fits. The y-intercepts (corresponding to the fit at 3 dB SNR) are tested against 0 bits. Older listeners show significant benefit from the Dutch masker over English (denoted by asterisk), but only in the transition region. Error bars in all plots indicate SEM. ($* \; p < 0.05$)

3.3.2 Phase-locking value

Phase-locking value (PLV) is a traditional measure of inter-trial coherence for a narrow-band response. Figure 3.5 shows the grand average of PLV at 100 Hz by age and masker condition. Older listeners have lower phase-locking values than the younger listeners ($t_{(30)} = 2.62, p = 0.007$ for one-tailed *t*-test) on the averaged phase-locking values across time and SNR levels. By one-tailed *t*-tests ($PLV_{Dutch} - PLV_{English} > 0$), older listeners have significantly higher PLV under Dutch masking than English ($t_{(14)} = 2.74, p = 0.008$ for transition region; $t_{(14)} = 1.80, p = 0.047$ for steady-state region), while younger listeners' PLV is not significantly affected by informational masking ($t_{(16)} = 1.67, p = 0.058$ for transition region; $t_{(16)} = 0.05, p = 0.479$ for steady-state region).

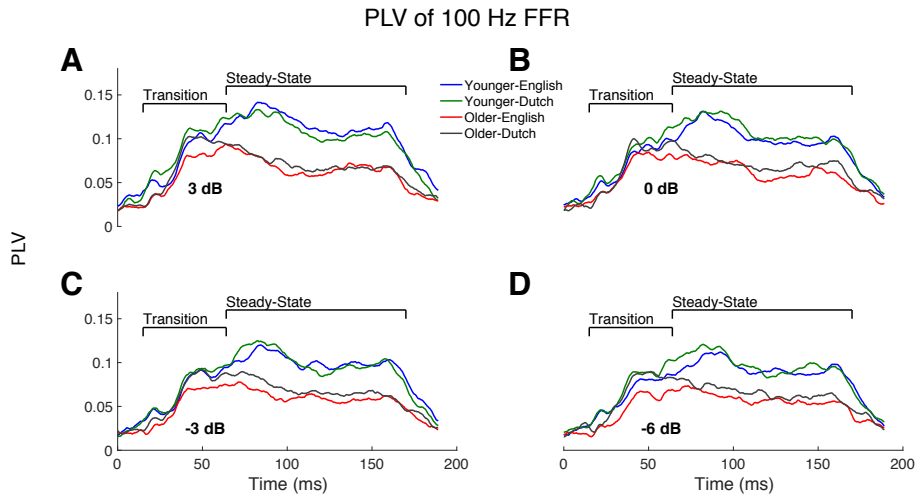

Figure 3.5. The PLV of the 100-Hz FFR is shown for all SNR levels, averaged across subjects, with colors indicating age and masker language, with younger listeners in blue (English) and green (Dutch) and older listeners in red (English) and gray (Dutch). A-D correspond to the four SNR levels: 3, 0, -3, -6 dB SNR. Younger listeners have visibly higher phase locking than older listeners. Older listeners have significantly better phase locking for the Dutch masker than for the English.

3.3.3 Information in phase of FFR

Phase information at 100 Hz

For the phase response at 100 Hz, the linear model, $I \sim age \times masker\ type + age \times SNR$ is significant ($F_{(5,250)} = 5.45, p < 0.001$ for the entire region; $F_{(5,248)} = 3.27, p = 0.007$ for the transition region; $F_{(5,248)} = 6.24, p < 0.001$ for the steady-state region). Outliers are excluded to satisfy homoscedasticity assumption (2 samples from transition and 2 samples from steady-state regions). Results show no significant interactions between age and masker type ($t_{(250)} =$

$0.56, p = 0.578$ for the entire region; $t_{(248)} = 0.22, p = 0.825$ for the transition region; $t_{(248)} = 0.06, p = 0.954$ for the steady-state region), and between age and SNR ($t_{(250)} = 0.86, p = 0.393$ for the entire region; $t_{(248)} = 1.05, p = 0.297$ for the transition region; $t_{(248)} = 0.66, p = 0.511$ for the steady-state region). A linear model with no interactions was then constructed and tested, i.e., $I \sim age + masker\ type + SNR$. The model itself is significant ($F_{(3,252)} = 8.77, p < 0.001$, $F_{(3,250)} = 5.08, p = 0.002$, $F_{(3,250)} = 10.32, p < 0.001$ for the entire region, the transition and steady-state regions, respectively). Comparisons show that younger listeners' responses contain significantly more information than older listeners' responses in the steady-state region ($t_{(252)} = 4.52, p < 0.001$ for the entire region; $t_{(250)} = 2.12, p = 0.035$ for the transition region; $t_{(250)} = 5.19, p < 0.001$ for the steady-state region), and that information increases as SNR increases ($t_{(252)} = 2.31, p = 0.022$ for the entire region; $t_{(250)} = 2.63,\ p = 0.009$ for the transition region).

Mutual information between stimulus and the response phase is analyzed analogously to that of the response amplitude. Phase information at 100 Hz is examined separately from the higher harmonics. To examine the effect of age and noise level, a linear regression line is fitted for information-by-SNR for each subject in both noise contents. The fitted y-intercept is compared for group differences. A one-tailed *t*-test (younger > older) effect size analysis on the y-intercept shows a significantly larger amount of information in younger than older listeners for the English masker ($t_{(30)} = 1.80, p = 0.041; d = 0.82, 95\%\ CI = [0.095, 1.540]$) ; the difference is not

significant for Dutch ($t_{(30)} = 1.36, p = 0.092;\ d = 0.58, 95\%\ CI = [-0.133, 1.284]$) (Figure 3.6A). Both age groups demonstrate decreasing information with worsening SNR: a one-tailed *t*-test on the negativity of the regression slope shows information loss, however, the negativity is not significant for older listeners in Dutch masker ($t_{(16)} = 3.31, p = 0.002$ and $t_{(16)} = 2.61, p = 0.013$ for younger listeners in English and Dutch maskers, respectively; $t_{(14)} = 2.17, p = 0.036;\ d = 0.56, 95\%\ CI = [3.19 \times 10^{-5}, +\infty]$, $t_{(14)} = 2.55, p = 0.061;\ d = 0.66, 95\%\ CI = [3.84 \times 10^{-5}, +\infty]$ for older listeners in English and Dutch maskers, respectively) (Figure 3.6B). No significant difference is seen between the slopes across age groups ($t_{(30)} = 1.36, p = 0.091$ and $t_{(30)} = 1.34, p = 0.095$ for English and Dutch masker, respectively). All tests have been corrected for multiple comparisons across the 6 frequency bands.

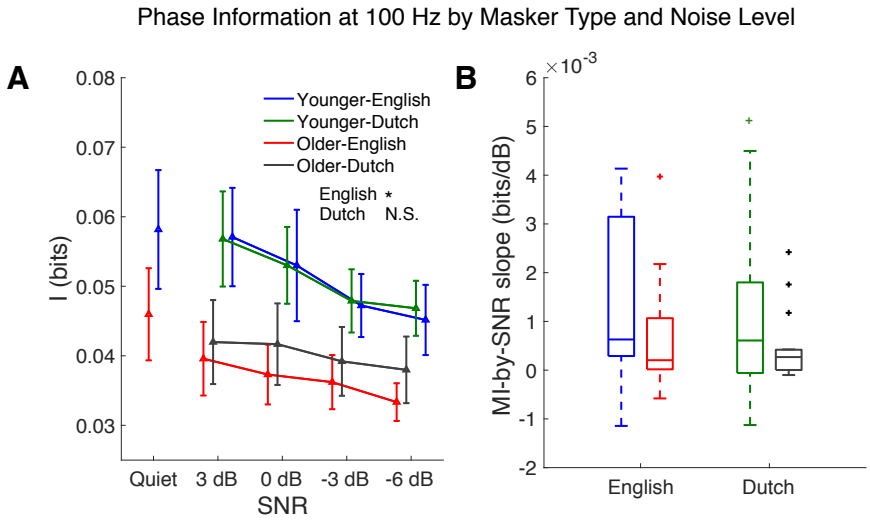

Figure 3.6. Mutual information between the stimulus and response phase as a function of noise level for each age group and masker condition (masker language). A:

Mutual information at the fundamental frequency as a function of noise level. The response in younger listeners conveys noticeably more information than the response in older listeners for the English masker condition, but the difference for Dutch is not significant at 100 Hz. Older listeners show consistently higher mutual information for the Dutch masker than for the English (the younger listeners show no consistent difference), but the difference is not significant at 100 Hz. B: The MI-by-SNR slopes of the previous plots show decreasing trends as SNR worsens, regardless of masker type, for both age groups. Younger listeners show a steeper decrease than older listeners but the difference is not significant at the 100-Hz response. Error bars indicate one SEM. ($* p < 0.05$)

Phase information in harmonics of 100 Hz

To examine information in the harmonics of 100 Hz, a linear regression line is fitted for mutual information as a function of SNR for each subject under each masker type. One-tailed (younger > older) *t*-tests on the y-intercept (with FDR correction) suggest that for all SNR levels, the aging midbrain contains significantly less information than the younger midbrain in all frequencies from 100 to 600 Hz (Figure 3.7A). For *p*-values near 0.05 (see Table 2), effect size analysis is further applied. For the English masker condition, the 100 and 200 Hz cases show consistent significance from both tests ($t_{(30)} = 1.80, p = 0.041;\ d = 0.82, 95\%\ CI = [0.095, 1.541]$ and $t_{(30)} = 1.83, p = 0.041;\ d = 1.06, 95\%\ CI = [0.317, 1.799]$), and similarly for the Dutch masker condition at 300, 400 and 500 Hz, respectively ($t_{(30)} = 2.12, p = 0.042;\ d = 1.39, 95\%\ CI = [0.613, 2.159]$ and $t_{(30)} = 1.97, p = 0.044;\ d =$

$0.84, 95\%\ CI = [0.116, 1.564]$ and $t_{(30)} = 2.28, p = 0.042;\ d = 1.64, 95\%\ CI = [0.838, 2.443]$) (see also Figure 3.7A). The results show significant decreasing slope in both groups and that the decrease with worsening SNR is faster for younger listeners than older listeners.

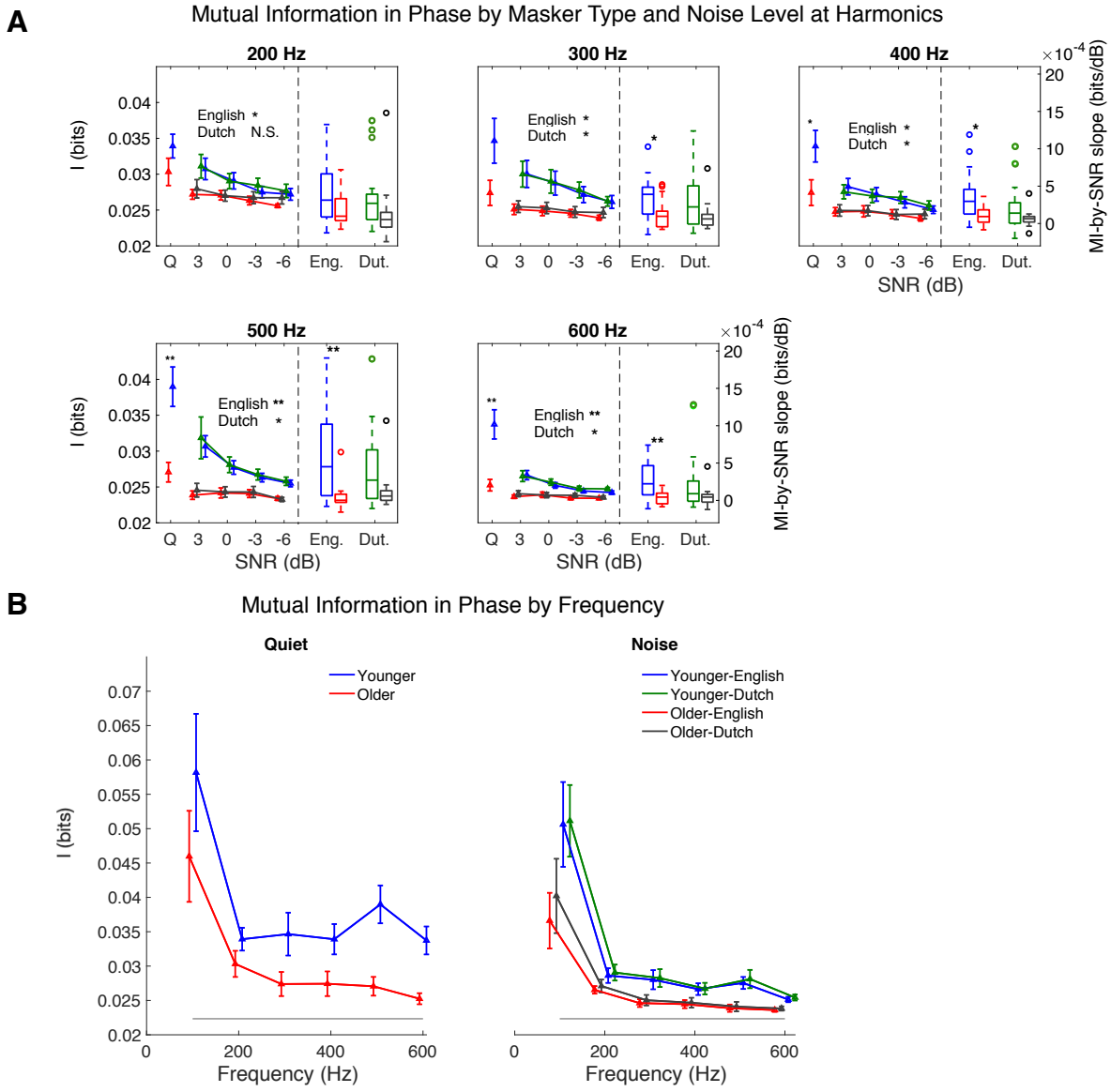

Figure 3.7. A: Mutual information for phase across frequency bands from 200 Hz to 600 Hz (separate subplot for each band). Within each subplot, as in Figure 3.3,

the left panel shows the mutual information as a function of SNR, separately for age group and masker type; in the right panels, the bar plots depict the linearly fitted decreasing slopes (of the plots shown in the left panel) for the different age groups and masker types. B: Overall, both in quiet (left) and averaged over SNR levels (right), mutual information decreases with increasing frequency (except for a single increase at 500 Hz for younger listeners). For older listeners, the decreasing trend in mutual information levels off at 500 Hz, which is lower than the frequency at which phase information levels off in younger listeners. The lower gray line represents the noise floor. Error bars indicate one SEM. ($* p < 0.05, ** p < 0.01$)

Table 3.2. Phase information: one-tailed *t*-test (younger > older) results applied to the fitted y-intercepts (3 dB values) and slopes from the linear regression analysis of mutual information (for response phase) as a function of SNR, for each harmonic. *p*-values are corrected for multiple comparisons by FDR correction. Boldfaced entries indicate the corresponding tests are statistically significant.

Harmonic (Hz)	Quiet (Y>O)		English masker (Y>O)				Dutch masker (Y>O)			
			y-intercept		slope		y-intercept		slope	
	t(30)	*p*	*t*(30)	*p*	*t*(30)	*p*	*t*(30)	*p*	*t*(30)	*p*
100	1.072	0.146	**1.798**	**0.041**	1.363	0.092	1.526	0.069	1.344	0.095
200	1.386	0.106	**1.833**	**0.041**	1.757	0.053	1.530	0.069	1.479	0.090
300	**1.898**	**0.050**	**2.219**	**0.026**	**2.089**	**0.034**	**2.122**	**0.042**	1.909	0.090
400	**2.170**	**0.038**	**2.407**	**0.022**	**2.694**	**0.011**	**1.967**	**0.044**	1.493	0.090
500	**3.609**	**0.002**	**3.740**	**0.001**	**3.352**	**0.003**	**2.280**	**0.042**	1.615	0.090
600	**3.579**	**0.002**	**3.738**	**0.001**	**3.446**	**0.003**	**2.690**	**0.035**	1.716	0.090

Phase information frequency limits

As seen in Figure 3.7B, the stimulus information contained in the response phase decreases with frequency for both age groups. Similar to amplitude analysis, the frequency-decreasing measure used here is phase information of y-intercept at 3dB of the fitted MI-by-SNR regression line. The measure at 600 Hz for older listeners is not statistically distinguishable from the noise floor ($t_{(14)} = 0.11, p = 0.917$ by one-sample *t*-test). For younger listeners, the measure is significantly higher than the noise floor at all frequencies ($t_{(30)} = 3.74, p < 0.001$ for English masker; $t_{(30)} = 2.69, p = 0.007$ for Dutch masker (younger > older), both at 600 Hz where lowest information is observed), i.e., the information for younger listeners has not yet reached floor by 600 Hz. In contrast, the cutoff frequency for older listeners is 500 Hz: the information measure at 500 Hz is not significantly greater than that at 600 Hz ($t_{(14)} = 0.74, p = 0.235$ under English masker; $t_{(14)} = 1.07, p = 0.152$ under Dutch masker). Therefore, results suggest a lower frequency limit of 500 Hz for older listeners than beyond 600 Hz for younger listeners.

Effect of masker type on phase information

As seen in Figure 3.6B, older listeners demonstrate a slower fall-off in phase information as a function of SNR when the noise masker is Dutch than for English. Analogous to amplitude analysis, the difference in mutual information between the Dutch and English maskers is calculated for each subject in all SNR levels (for both transition and steady-state regions) to examine phase information benefit from the

Dutch masker over the English masker, and a linear model of $I_{Dutch} - I_{English} \sim SNR$ shows a significantly positive intercept for older listeners in the transition region ($t_{(56)} = 4.64, p < 0.001$ with 2 samples omitted) but not in the steady-state region ($t_{(54)} = 1.77, p = 0.083$ with 4 samples omitted). Younger listeners, however, do not show significant positive intercept in either transition ($t_{(64)} = 1.75, p = 0.085$ with 2 samples omitted) or steady-state region ($t_{(66)} = -0.64, p = 0.522$). Samples were omitted from the tests to satisfy the homoscedasticity requirement. For justification, a regression line was fitted as a function of SNR to reduce within-subject variance. Using a one-tailed *t*-test on the y-intercept (effective mutual information benefit at 3 dB SNR) of the regression line against zero, the mutual information benefit from the Dutch masker over the English masker is significantly higher for older listeners in the transition region ($t_{(14)} = 2.31, p = 0.018$), but not the steady-state region ($t_{(14)} = 1.55, p = 0.072$). No significant benefit is found for younger listeners in either region ($t_{(16)} = 1.33, p = 0.102$ and $t_{(16)} = 0.44, p = 0.332$ for transition and steady-state region, respectively). The regression slope is not significantly positive or negative for either group ($p > 0.05$ by two-tailed *t*-tests), as seen in the bar plots in the right panels of figure 3.8C and 3.8D.

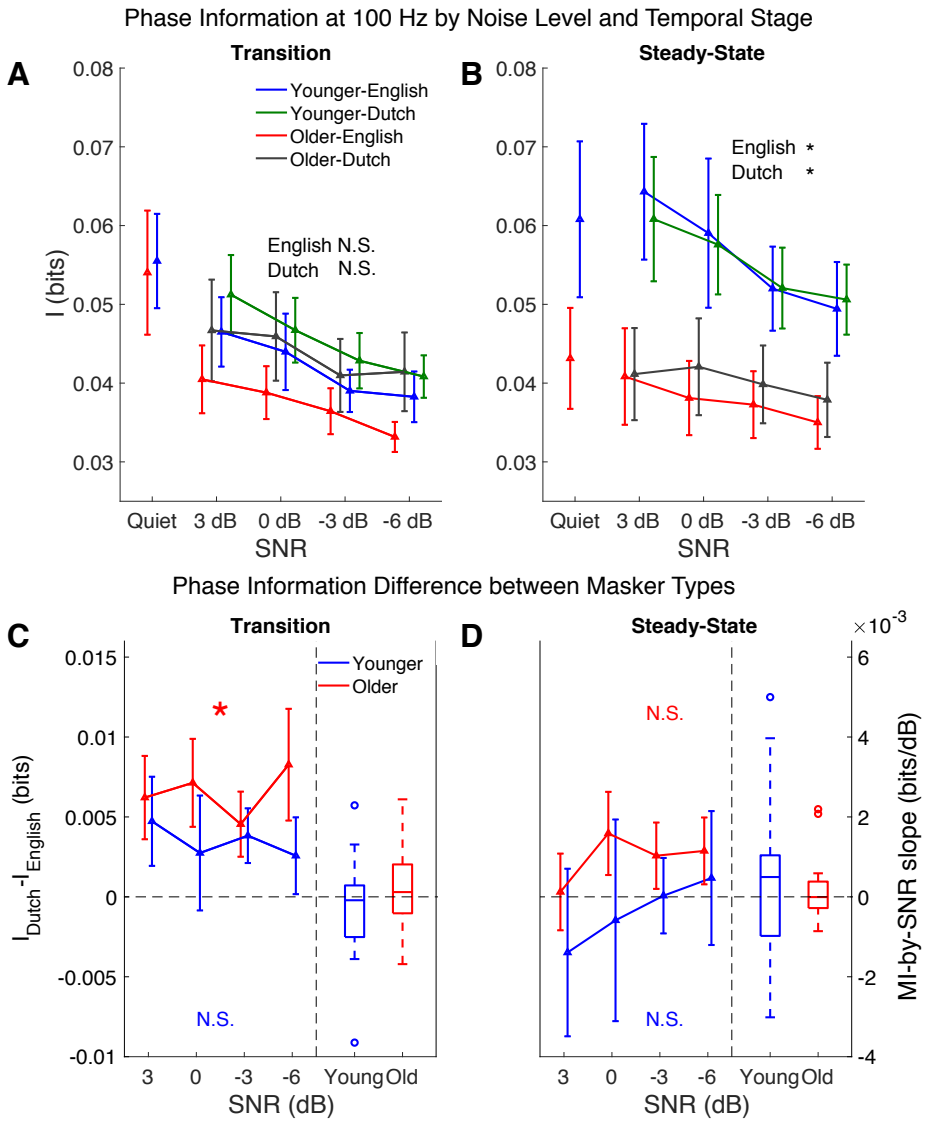

Figure 3.8. Mutual information of phase response by masker type and response region for younger listeners in blue (English) and green (Dutch) and older in red (English) and gray (Dutch). A and B demonstrate the mutual information as a function of SNR in the transition and steady-stage regions, respectively. In the steady-state region, group differences are significant for both masker types, indicated by asterisks. C and D illustrate the mutual information difference between masker types (denoted $I_{Dutch} - I_{English}$) in the transition region and steady-state region, respectively. In each

plot, the left panel displays information as a function of SNR, and the right panel displays a bar plot showing the slopes of the linear fits. The y-intercepts (corresponding to the fit at 3 dB SNR) are tested against 0 bits. Older listeners show significant benefit from the Dutch masker over English (denoted by asterisk), but only in the transition region. Error bars in all plots indicate SEM. ($* p < 0.05$)

3.4 Discussion

Based on these results from the mutual information analysis of FFR amplitude and phase, this study provides supporting evidence that the neural response of the midbrain of older listeners is not merely less well synchronized than for younger listeners (Anderson et al. 2012; Presacco et al. 2016a, 2016b) but also actually contains less information, in both amplitude and phase. At the fundamental frequency, the informational loss for older listeners was seen only in the presence of a competing talker. In contrast, for higher frequencies, the informational loss for older listeners was seen in both quiet and noisy conditions. Furthermore, the masker type (Dutch vs. English) significantly affects the amount of stimulus information carried in the response at the fundamental frequency in the transition region, for older listeners but not younger. This last finding arises for the first time from this mutual information analysis and demonstrates that mutual information analysis provides access to response properties otherwise hidden by response variability.

3.4.1 Aging

Aging has different effects on subcortical and cortical auditory stages along the ascending pathway. Here this study addresses its effect on midbrain representations of FFR from an information point of view. First results show a broad-band (100-600 Hz) informational loss associated with aging in both quiet and noisy conditions, which is reflected in both the amplitude and phase of the responses. The informational loss at the fundamental frequency can be attributed to the delayed and weakened responses in the aging midbrain (Anderson et al. 2012; Burkard and Sims 2002; Clinard and Tremblay 2013), which can be linked to age-related loss of neural inhibition. For example, dorsal cochlear nucleus (DCN) has been shown to represent signal and suppress background noise aided by glycinergic neurotransmitters, and aging rats display decreased glycinergic inhibition in DCN (Caspary et al. 2005, 2006). Another contribution may come from synaptopathy arising from a loss of inner hair cell (IHC) ribbons and degeneration of ganglion cells (Sergeyenko et al. 2013), or from a decline in low-spontaneous-rate nerve fibers as has been seen in aging gerbils (Schmiedt et al. 1996). Together, synaptopathy and loss of neural inhibition in midbrain may both contribute to less information in midbrain FFR in older listeners.

3.4.2 Noise level

In these results, the amount of information in FFR (both phase and amplitude) decreases as noise level increases (i.e., SNR decreases) for both younger and older listeners. This result is consistent with previous findings (Presacco et al. 2016a, 2016b) where the amplitude of FFR decreases with worsening noise level. Via linear regression,

it is also seen that younger listeners have a more steeply decreasing slope (as a function of noise level) than the older listeners, at both the fundamental frequency and its harmonics. This result may also be due to disrupted synchrony at auditory nerve fibers (Schmiedt et al. 1996) and the synapse (Sergeyenko et al. 2013). A loss of auditory nerve fibers in older listeners may lead to a reduced brainstem response, causing a decrease in information even in the quiet condition, leading to a slower rate of additional decrease with increasing noise level.

3.4.3 Masker type

In this experiment background masker types included English (meaningful to all listeners) and Dutch (meaningless to all listeners). The results suggest that the informational content of the noise affects information in the midbrain FFR, in both amplitude and phase (in the transition region): older listeners benefit *neurally* from the masker being meaningless over meaningful. It is unexpected that a high-level feature such as language would affect midbrain neural responses, though this has been seen before for younger listeners (Presacco et al. 2016b). One explanation for the language-dependent response difference in the aging midbrain could be top-down modulation from cortical areas. Descending pathways from primary auditory cortex to inferior colliculus (IC) in the midbrain have been reported to mediate learning-induced auditory plasticity (Bajo et al. 2010), and IC neurons' sensitivity to sound frequency and intensity can be modified by cortical projections (King and Bajo 2013). Since older listeners benefit behaviorally from competing speech being non-meaningful (Pichora-

Fuller 2008; Tun et al. 2002), the cortical processing underlying this difference may also project back upstream to the midbrain.

Another explanation for this difference in FFR due to masking language is that the difference might be purely cortical. i.e., purely cortical FFR. Recent studies (Coffey et al. 2016a, 2017) have shown that traditional EEG-measured FFR may not be purely subcortical at all. It would be substantially less surprising to see language-specific effects originating from cortex than midbrain, although, even so, these effects from the transition region (15-65 ms) are earlier than might be expected from a language-influenced cortical response.

3.4.4 High frequency limit

Results show that for both amplitude and phase information, responses from older listeners in speech-in-noise conditions contain less information in the higher frequencies, and have lower high frequency limits, than younger listeners. Such deficits might be also associated with lowered temporal precision arising from a loss of auditory nerve fibers and ganglion cells (Schmiedt et al. 1996; Sergeyenko et al. 2013), which affect all frequencies. The same analysis carried out on single sweeps (see Appendix) suggests that the decrease in information at high frequencies may not be due to the average of the two polarities.

3.4.5 Relation to cortical representation

Even though the stimulus representation at the level of auditory midbrain is weaker for older listeners, whether based on RMS, correlation, or mutual information

measures, it is paradoxically amplified at the level of auditory cortex (Brodbeck et al. 2018a; Presacco et al. 2016a, 2016b). A negative association between subcortical FFR and cortical responses, as measured with mutual information, has been shown in older listeners in a task of categorical syllable perception (Bidelman et al. 2014). The analogous correlation between cortical speech representation and midbrain response amplitude was not seen, however, for temporal speech processing (Presacco et al, 2016b). Both attention and behavioral inhibitory control are used to enhance understanding of speech in noise, but the extent to which these high-level cortical processes are altered by auditory periphery deficits is not well known (Presacco et al. 2019). Furthermore, it is unclear where and how the neural representation of speech in older listeners shifts from degraded in midbrain to exaggerated in cortex, but mutual information is a promising tool to address these issues (Bidelman et al. 2014).

3.4.6 Summary

The approach employed here, using mutual information to analyze the relationship between a speech-in-noise stimulus and the FFR response, can be seen in at least two different lights. At one level it can be viewed as a mathematical measure derived from information theory (Cover and Thomas 1991; Shannon 1948). This places the present analysis on firm mathematical grounds, using concepts and measures from a well-established field of mathematical signal processing. At another level, the analysis can be viewed as an acknowledgement that the relationship between stimulus and response may have strongly non-linear aspects, with mutual information being just one of several available non-linear measures that allow us to move beyond

conventional linear analysis methods (e.g. evoked response analysis) and conventional phase coherence methods.

Chapter 4. Mutual information analysis in the auditory cortex and the effects of aging

4.1 Introduction

The human brain is capable of separating attended speech from background distractions. However, this capability degrades with aging. Behavioral studies have shown age-related temporal processing deficits in different auditory tasks, such as pitch discrimination (Fitzgibbons and Gordon-Salantt 1996), gap-in-noise detection (Fitzgibbons and Gordon-Salant 2001) and recognition of speech in noise (Frisina and Frisina 1997; Gordon-Salant et al. 2006; He et al. 2008). Neurophysiological studies show that although the auditory brain robustly segregates speech from either a competing speaker (Ding and Simon 2012b) or spectrally matched noise (Ding and Simon 2013), the temporal processing degrades by demonstrating age-related changes in response latency and strength in midbrain (Anderson et al. 2012; Burkard and Sims 2002; Clinard and Tremblay 2013) and cortical evoked responses (Lister et al. 2011; Presacco et al. 2016a, 2016b). In animal studies, overrepresentation in central auditory nervous system has been seen in aging animals (Hughes et al. 2010). In aging rats, the altered neural inhibition, functional impairments in the cortex are mostly due to a regulated plasticity change and most of them are reversible (de Villers-Sidani et al. 2010). However, it remains an unsolved question how much plasticity change is in the aging human brain, and its effects on speech processing.

A recent magnetoencephalography (MEG) study suggests an exaggerated response to speech in noise for older listeners (Presacco et al. 2016a, 2016b) by demonstrating a higher speech envelope reconstruction accuracy than younger. Since the reconstruction is based on a linear decoder with window length of ~500 ms, it remains unclear which response components contribute to the exaggerated response, especially with respect to latency, i.e., 50 ms (M50) or 100 ms (M100). Since previous studies observe attention modulation of the M100 response (Ding and Simon 2012b, 2013), and older listeners may pay more attention in the listening tasks (Presacco et al. 2016a), the study hypothesizes that older listeners will have a higher mutual information level at responses of 100 ms and later. Furthermore, Presacco et al. (2016b) also shows a negative correlation between the speech envelope reconstruction accuracy and behavioral inhibitory control score, measured by a visual inhibitory task, for older listeners. It suggests that there might be responses, at M50, M100 or other later responses if any that contribute to the correlation. Since the multi-modal association of auditory and visual responses occurs later than auditory attention (Griffiths and Warren 2004), a correlation is expected between M100 or later response for older listeners. Here in this study, all the response properties, such as latency and amplitude, will be evaluated with respect to mutual information.

Earlier investigations the cortical coding of continuous speech have often relied on linear methods (Ding and Simon 2012b; Presacco et al. 2016a, 2016b). Auditory cortex, however, is well known to employ non-linear processing (Sahani and Linden 2003), and therefore a non-linear analysis framework might provide more insight.

Nonlinear approaches based on Shannon's information theory (Shannon 1948) have been successfully applied to spiking neurons in the auditory system (Nelken and Chechik 2007), to EEG subcortical recordings (Zan et al. 2019; Chapter 3), and even to MEG recordings from auditory cortex (Cogan and Poeppel 2011), where it was used to decode phase information in low-frequency responses to speech. By estimating mutual information between midbrain and cortical responses, recent study shows a redundant information transition for older listeners in the task of categorical perception of speech syllables (Bidelman et al. 2014).

Here, to investigate the information encoded in the cortical response phase-locked to continuous speech, the measure of temporal mutual information function (TMIF) is developed. It provides a non-linear measure of a general phase-locked response to speech, and is a non-linear analog to the TRF, or the evoked response to a brief sound. It also has response peaks at specific latencies, analogous to the TRF's $M50_{TRF}$ and $M100_{TRF}$ peaks, or the M50 and M100 response peaks of an evoked response. The main mutual information peaks of the TMIF are, by analogy, named the MI50, MI100 and MI200, and occur for early cortical latency (~50 ms), middle cortical latency (~100 ms), and late cortical latency (~200 ms). The actual TMIF peak levels and latencies depend on the specific stimulus, the properties of any maskers, the age of the subject, and other related factors. The main hypotheses above can be investigated by testing for associations between the properties of an individual's MI200 and their behavioral scores.

4.2 Materials and methods

4.2.1 Subjects

The dataset analyzed here was already obtained and analyzed for previous studies (Presacco et al. 2016a, 2016b). 22 subjects participated in the experiment: 17 younger adults ages 18 to 27 (3 male) and 15 older adults ages 61 to 73 (5 male). All participants were recruited from the greater Washington D.C. area (Maryland, Virginia and Washington D.C.), with clinically normal hearing. Specifically, participants had normal hearing thresholds ($\leq 25\ dB$ hearing level) from 125 Hz to 4000 Hz, no history of neurological or middle ear disorders or surgery, and normal intelligent quotient scores (≥ 85 on the Wechsler Abbreviated Scale of Intelligence (Zhu and Garcia 1999). Written informed consent was obtained from each subject, and they were compensated for their time. The experimental protocol and all procedures were reviewed and approved by the Institutional Review Board of the University of Maryland.

4.2.2 Behavioral tests

Flanker test

The ability to attend to a selected or goal-appropriate stimulus and to ignore other distracting stimuli is associated with behavioral inhibitory control (Neill et al. 1995), and this ability declines with aging (Diamond 2013). This ability may affect auditory suppression of a competing speaker while attending to another. To investigate broad aging effects on behavioral inhibitory control, including its relationship with complex auditory processing, a visual Flanker test (Ward et al. 2016) was given to all subjects. The Flanker test measured behavioral inhibitory control and attention control

by displaying five arrows in a row and asking only for the direction of the middle arrow, i.e., the flanking arrows serve only as distractors. Both reaction time and accuracy are taken into account for scoring (Weintraub et al. 2013), and a *higher* Flanker score indicates better performance, i.e., more behavioral inhibitory control.

QuickSIN test

The Quick Speech-in-Noise test (QuickSIN) measures listeners' ability to understand speech in noise (four-speaker babble), with subjects asked to recall words presented at different SNR levels, with performance rated by the number of words they failed to recall (Killion et al. 2004). A *lower* QuickSIN test score indicates better performance, i.e., superior ability to understand speech in noise.

Flanker and QuickSIN scores may be correlated across subjects and this is measured with Pearson's correlation test for both age groups.

4.2.3 Stimuli and MEG recording

The task and stimuli were the same as the ones described in previous studies (Presacco et al. 2016a, 2016b). For each subject, the MEG response was recorded with an 157 axial gradiometer whole head MEG system (KIT, Kanazawa, Japan) inside a magnetically shielded room (Vacuumschmelze GmbH & Co. KG, Hanau, Germany) at the University of Maryland, College Park, sampled at 1000 Hz with online low-pass filter of cut-off frequency at 200 Hz. The stimulus was continuous speech (a narrated audio book), either from a solo speaker or a mixture of two concurrent speakers. The solo-speaker speech stimuli were one-minute segments from an audiobook, *The Legend*

of Sleepy Hallow by Washington Irving, narrated by a male speaker (http://www.audiobooktreasury.com/legend-of-sleepy-hollow/). The mixture was composed of foreground speech to which the subject was instructed to attend and a background, which served as a distractor. The foreground speech was from the same source as the clean speech condition. The background stimuli were one-minute segments from an audiobook, *A Christmas Carol* by Charles Dickens, narrated by a female speaker (http://www.audiobooktreasury.com/a-christmas-carol-by-charles-dickens-free-audio-book/). The foreground and background speech segments were mixed together at four different power ratios, of 3 dB, 0 dB, -3 dB and -6 dB. The foreground speech used in -6 dB condition and the clean speech were the identical, and the clean speech was only presented after all the mixed speech stimuli had been presented. The subjects also listened to mixed speech stimuli where the background speaker spoke Dutch instead of English. The Dutch speech stimuli were not comprehensible to all subjects, and responses to these stimuli were not analyzed here. The stimuli were all presented to the subjects with E-A-RLINK earphones attached with sound tubing at about 70-dB sound pressure level.

For each subject, under each condition, the raw MEG recording was first denoised by time-shifted principle component analysis (de Cheveigné et al. 2007), in which three separate reference channels recording the environmental noise serve as a reference with which to eliminate environmental noise from the 157 neural data channels. A blind source separation approach called denoising source separation (DSS; de Cheveigné and Simon 2008; Särelä and Valpola 2005) is then used to extract the

dominant auditory components, based on the 2-8 Hz band-passed response (Ding and Simon 2013) to extract a spatial filter which is applied to the 1-8 Hz (FIR filter) bandpassed result of TSPCA (Ding and Simon 2012b). Finally, the first DSS component, which contain the responses contributing most reliably over repeated stimulus presentations, is analyzed further as described below.

4.2.4 Data analysis

Temporal mutual information function (TMIF)

For decoding the cortical phase-locked response to speech, the use of the temporal mutual information function (TMIF) shares properties with the analogous temporal response function (TRF) (Ding and Simon, 2012). While the TRF, typically 500 ms in length for cortical responses, linearly maps the stimulus envelope to the low-frequency response counterpart, the TMIF captures non-linear cortical modulations following the speech envelope. A typical TRF has prominent peaks at latencies of approximately 50 ms and 100 ms, meaning that any speech envelope feature will evoke a pair of cortical responses 50 ms and 100 ms later. Since this implies enhanced cortical processing of speech information at those latencies, therefore it may be expected that the mutual information between the speech envelope and the response at those latencies should also result in peaks (though both peaks would be positive since mutual information can never be negative). Only the TMIF of the first DSS component is computed.

Specifically, since the mutual information is determined by the joint distribution of the stimulus and response, quantization is necessary to specify the probability distribution. To estimate the TMIF, the method first quantizes both speech envelope and amplitude response into integer-labeled bins (1 to 8) based on the equipartition principle: that the number of samples assigned to each integer are approximately the same (limited necessarily by the divisibility of the number of samples into the number of bins). Here, $x(t)$ denotes the quantized speech envelope, and $y(t)$ denotes the quantized response at time point t, so then the mutual information at time t is

$$I_t(X;Y) = \sum_{x,y} p\big(x(\tau), y(\tau+t)\big) \log \frac{p\big(x(\tau), y(\tau+t)\big)}{p\big(x(\tau)\big)p\big(y(\tau+t)\big)}. \tag{4.1}$$

Let $S = \{1, 2, \dots, 8\}$ be the set from which the sample values are drawn. The joint probability distribution of $x(\tau) \in S$ and $y(\tau+t) \in S$, i.e., $p\big(x(\tau), y(\tau+t)\big)$, is drawn from different values of τ, which ranges from 0 to $L-1$, where L is the length of stimulus/response in ms. Since the analysis is done at a sampling rate of 1 kHz, i.e., a sampling period of 1 ms, L is also the sample size. Practically, the mutual information at each time point is estimated by its relation to entropy and conditional entropy, $I(X;\, Y) = H(Y) - H(Y|X)$. Specifically, the equation above could be written as,

$$I_t(X;Y) = \sum_{i\in S, j\in S} p(x(\tau)=i, y(\tau+t)=j) \log \frac{p(x(\tau)=i, y(\tau+t)=j)}{p(x(\tau)=i)p(y(\tau+t)=j)}. \tag{4.2}$$

Here, *i* and *j* are values drawn from set *S*; t and τ are integer numbers of ms. A time window of 500 ms is used for *t*, and mutual information is estimated with 2-ms steps, i.e., $t \in \{0, 2, \dots, 498\ (ms)\}$. Then the TMIF function is given by $TMIF(t) = I_t(X;Y)$. In summary, $TMIF(t)$ is estimated by mutual information between stimulus and response shifted forward by time t. Let Y_t be the response shifted forward by t, $TMIF(t) = I(X;Y_t)$. To prove that $TMIF(t)$ does not contain redundant information introduced by repeatedly shifting Y, we need to show whether $I(X;Y_t,Y_{t+1},\ \dots,Y_n) - I(X;Y_{t+1},\ \dots,Y_n) = I(X;Y_t)$. Based on the chain rule for mutual information (Cover and Thomas 1991),

$$I(X;Y_1,Y_2,\ \dots,Y_n) = \sum_{i=1}^{n} I(X;Y_i|Y_{i-1},Y_{i-2},\ \dots,Y_1). \tag{4.3}$$

Therefore,

$$\begin{aligned} & I(X;Y_t,Y_{t+1},\ \dots,Y_n) - I(X;Y_{t+1},\ \dots,Y_n) \\ &= \sum_{i=t}^{n} I(X;Y_i|Y_{i-1},Y_{i-2},\ \dots,Y_1) \\ &- \sum_{i=t+1}^{n} I(X;Y_i|Y_{i-1},Y_{i-2},\ \dots,Y_1) \\ &= I(X;Y_t), \end{aligned} \tag{4.4}$$

which proves that $TMIF(t)$ was not affected by repeatedly shifting the response Y.

After estimating the TMIF function for each subject, the distinctive peaks with approximate latency of 50, 100, and 200 ms are identified as the MI50, MI100 and

MI200 peaks (later their amplitude difference between age groups will be tested). Peaks are found by searching for the maximum value over a specific time range. Since the response latencies differ when in quiet condition and noise conditions, different ranges are applied for different conditions, with range boundaries determined by the trough latencies in the relevant TMIF when averaged over subjects. Specifically, for the quiet condition, the MI50 corresponds to the time point with the largest amplitude in the range of 2-86 ms in the time course, while MI100 and MI200 each corresponds to the maximum of ranges of 80-160 ms and 150-300 ms respectively. The group difference is tested for each peak by performing 2-sample one-tailed *t*-tests over amplitudes. For the four noise conditions the TMIF is analyzed analogously, but since the noise is just background speech, for each SNR condition two TMIFs are computed for each subject, based on the foreground and background speech respectively. The temporal ranges of foreground are 2-70 ms for the MI50, 50-200 ms for the MI100 and 200-300 ms for the MI200. The temporal ranges of background are 2-120 ms for the MI50, 120-230 ms for the MI100 and 200-350 ms for the MI200. The group difference is tested for each peak by performing the same *t*-tests over the averaged amplitude across SNRs.

Statistics

To systematically examine relationships among neural responses properties of the TMIF (MI50, MI100 and MI200) and behavioral scores, linear mixed effect models (LME) are used. For each neural response peak, a base model is constructed as a function of fixed effects from $age \times attention \times behavior + SNR$ and a random effect of subject-specific bias. Here, $attention$ has the value of either foreground or

background, and $behavior$ is either the Flanker or QuickSIN score. The 4-way interaction was not included due to the limited degrees of freedom. To scrutinize the significance of a factor (or an interaction) in the prediction of a neural response, a second model is constructed without the factor (or interaction) and is compared with the base model by ANOVA. Then non-significant factors or interactions are excluded from model, and the significant interaction is examined by dissecting it into all possible combinations of its categorical values and further analyzed by linear models. All linear model analysis is done in R (R. Core Team 2017), and LME analysis is performed with the toolbox *lme4* (Bates et al. 2015).

4.3 Results

By implementing the approaches established above, for each subject under each condition, TMIFs were computed for the first DSS component. Here, this section reports results under the conditions of clean speech and mixed speech with SNRs of +3, 0, -3 and -6 dB and source-space analysis in the worst SNR case.

4.3.1 Behavioral correlation

A slight negative correlation was found between Flanker score and QuickSIN score, for older listeners (Pearson's correlation $r = -0.52, p = 0.046$), as shown in Figure 4.1.

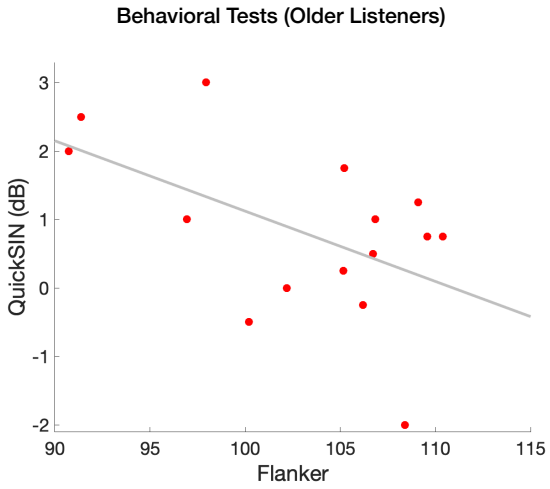

Figure 4.1. Behavioral test correlation. Flanker score (higher is better) is negatively correlated with Quick-SIN score (lower is better) in older listeners ($r = -0.52, p = 0.046$).

4.3.2 Neural Responses to Clean speech

To investigate the age-related exaggerated information representation in the quiet condition, peaks analogous to TRF peaks are identified, namely the MI50, MI100 and MI200, analogous to the M50, M100 and M200 MEG TRF (and evoked response) peaks. As for their evoked counterparts, peaks of different latencies may be associated with different stages of the processing chain. A one-tailed *t*-test is performed for each peak amplitude for younger against older. Results show that all the peaks from the older listeners are significantly larger than those of the younger, with $t_{30} = -1.85, p = 0.037$ for MI50, $t_{30} = -2.52, p = 0.009$ for MI100 and $t_{30} = -2.24, p = 0.031$ for MI200. The results suggest that *all* the processing stages in the aging cortex overly represents the clean speech envelope.

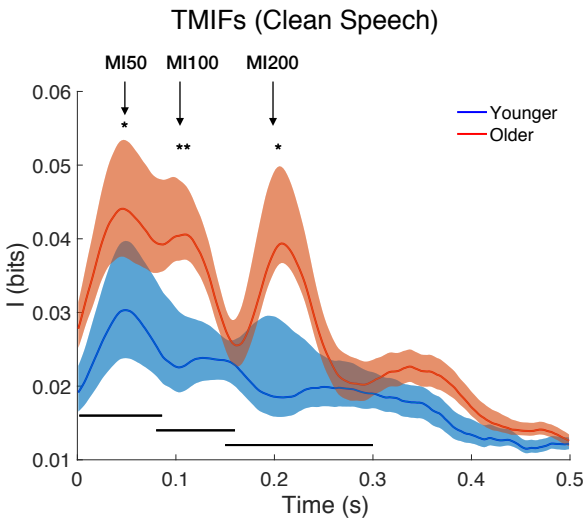

Figure 4.2. TMIF to clean speech. Shaded areas above and below the solid lines indicate the standard error of mean. The temporal ranges over which MI50, MI100 and MI200 for each subject are constrained are marked by the three black lines above x-axis. Asterisks show the significance of amplitude differences between the two groups from a one-tailed *t*-test (*p<0.05, **p<0.01).

4.3.3 Neural Response to Mixed Speech

In mixed speech conditions, separate TMIFs for both foreground and background speech are computed, shown in Figure 4.3 and Figure 4.4, respectively. Response peaks are extracted and effects from factors of age, attention and behavior are examined systematically by linear mixed effect models, $MI\ level \sim age \times attention \times behavior + SNR + (1|subject)$, where the random effect term, $(1|subject)$, accounts for subject-specific intercepts or bias. The two behavioral scores (Flanker and QuickSIN) were considered separately. When considering the Flanker score, the 3-way interaction is significant for models predicting the amplitude of the

MI50 ($\chi^2_{(4)} = 16.45, p = 0.002$), MI100 ($\chi^2_{(4)} = 98.08, p < 0.001$) and MI200 ($\chi^2_{(4)} = 91.38, p < 0.001$) compared with a null model with no interactions, i.e., $MI \sim age + attention + behavior + SNR + (1|subject)$. To examine the significance of interactions, variables *age, attention* and *behavior* are then separately released from the 3-way interaction. Those results show that the $age \times attention$ interaction is significant in predicting the amplitude of the MI50 ($\chi^2_{(3)} = 7.61, p = 0.055$ by releasing *behavior (Flanker)*; $\chi^2_{(3)} = 14.17, p = 0.003$ by releasing *age*; $\chi^2_{(3)} = 14.52, p = 0.002$ by releasing *attention*), and the 3-way interaction is significant in predicting the amplitude of the MI100 ($\chi^2_{(3)} = 66.89, p < 0.001$ by releasing *behavior (Flanker)*; $\chi^2_{(3)} = 70.89, p < 0.001$ by releasing *age*; $\chi^2_{(3)} = 83.92, p < 0.001$ by releasing *attention*) and MI200 ($\chi^2_{(3)} = 88.98, p < 0.001$ by releasing *behavior (Flanker)*; $\chi^2_{(3)} = 78.67, p < 0.001$ by releasing *age*; $\chi^2_{(3)} = 72.39, p < 0.001$ by releasing *attention*). Therefore, variables of *age* and *attention* interact with *behavior* in predicting the level of mutual information, and the prediction power changes for different combinations of $age \times attention$, such as younger and foreground vs. older and foreground. To examine the prediction differences, the model of $MI \sim behavior + SNR$ is constructed separately for different combinations of *age* and *attention*. The overall model significances are shown in Table 1, and the effects of *behaviors* are shown in Table 2.

Behavior	Attention	Age	MI50		MI100		MI200	
			$F_{(2,57(65))}$	*p*	*F*	*p*	*F*	*p*
	FG	Y	**5.52**	**0.006**	**3.84**	**0.026**	1.33	0.271

Flanker		O	**6.37**	**0.003**	**16.76**	**<0.001**	**32.44**	**<0.001**
	BG	Y	1.74	0.183	**6.44**	**0.003**	**4.35**	**0.017**
		O	0.34	0.715	2.41	0.099	0.41	0.668
Q-SIN	FG	Y	**4.85**	**0.011**	0.56	0.579	0.14	0.869
		O	2.64	0.080	2.28	0.112	**4.52**	**0.015**
	BG	Y	1.29	0.288	**8.05**	**<0.001**	**5.86**	**0.005**
		O	-0.42	0.677	1.98	0.147	0.42	0.656

Table 4.1. Model $MI \sim behavior + SNR$ significance. FG: foreground; BG: background. Boldfaced entries indicate the corresponding tests are statistically significant.

Behavior	Attention	Age	MI50		MI100		MI200	
			t	*P*	*t*	*p*	*t*	*P*
Flanker	FG	Y	**2.90**	**0.005**	**2.56**	**0.013**	1.56	0.124
		O	**-3.56**	**<0.001**	**-5.79**	**<0.001**	**-7.96**	**<0.001**
	BG	Y	1.30	0.199	-0.05	0.961	1.50	0.139
		O	0.14	0.893	-0.91	0.366	0.35	0.731
Q-SIN	FG	Y	**-2.67**	**0.010**	-0.27	0.792	-0.23	0.819
		O	**2.29**	**0.026**	**2.13**	**0.038**	**2.87**	**0.006**
	BG	Y	-0.87	0.385	-1.64	0.106	**-2.24**	**0.029**
		O	-0.42	0.677	-0.19	0.853	-0.39	0.696

Table 4.2. Effects of behaviors (Flanker and Quick-SIN) in prediction of mutual information. Boldfaced entries indicate the corresponding tests are statistically significant.

To investigate whether the exaggerated response associated with aging occurs for both the foreground and the background, and which peak contributes to this exaggerated information in older participants, mutual information levels of all three peaks, for each stimulus, under each SNR condition are found for each subject and

compared between groups. Older listeners show significantly larger mutual information levels in all three peaks to both foreground ($t_{30} = -2.07, p = 0.024$ for MI50, $t_{30} = -3.80, p < 0.001$ for MI100 and $t_{30} = -2.37, p = 0.012$ for MI200) and background ($t_{30} = -2.44, p = 0.010$ for MI50, $t_{30} = -2.57, p = 0.0076$ for MI100 and $t_{30} = -2.90, p = 0.0035$ for MI200). Therefore, both foreground and background are exaggerated for older listeners, and the MI100 exaggerated information is the most prominent for foreground.

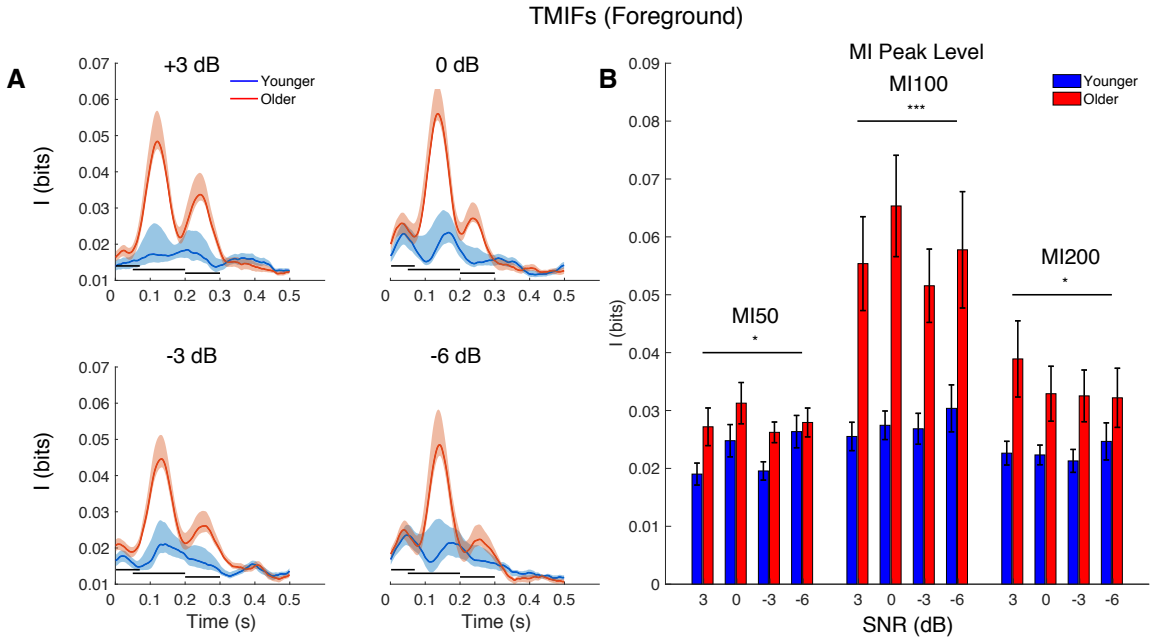

Figure 4.3. TMIFs of the foreground speech are amplified in older listeners. **A**. The four plots illustrate different SNR conditions of 3, 0, -3 and -6 dB SNR, with younger listeners in blue and older listeners in red. The three black horizontal lines in each figure indicates the ranges from which three peaks are extracted. Shaded areas: ± 1 SEM. **B**. MI peak level in older (red bars) and younger listeners (blue bars). 2-sample one-tailed *t*-tests on the averaged peak amplitudes over SNR conditions show that the

older listeners have significantly larger amplitudes (*$p<0.05$, **$p<0.01$, ***$p<0.001$). Error bars: ± 1 SEM.

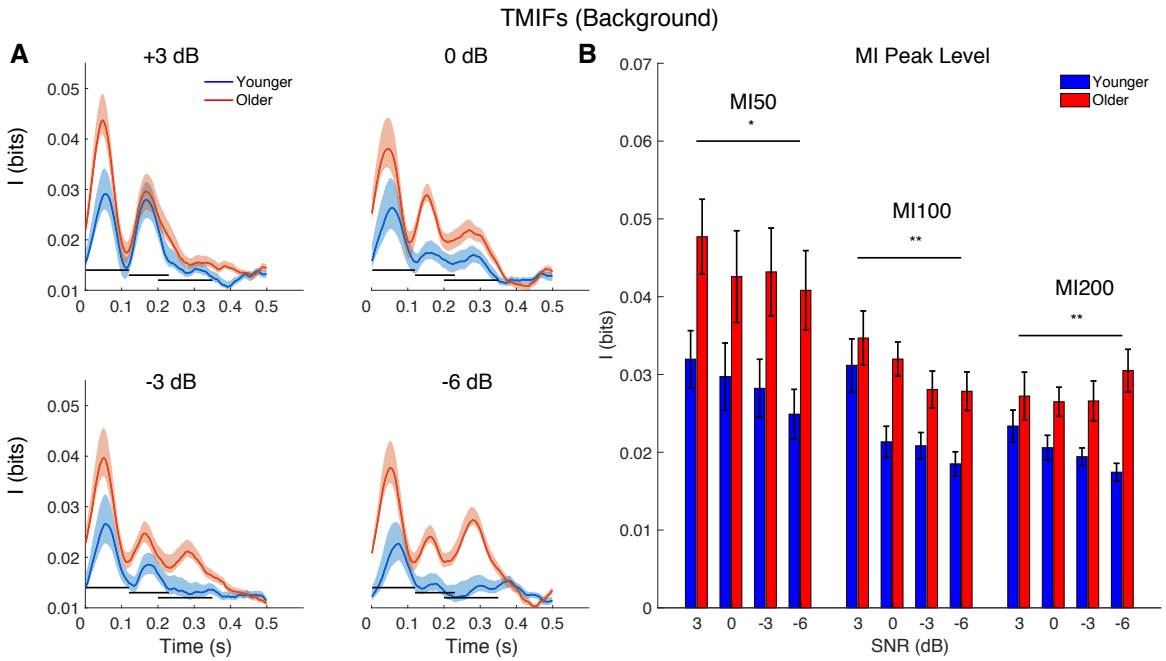

Figure 4.4. TMIFs of background speech are amplified in older listeners. Plots in **A** illustrate different SNR conditions of 3, 0, -3 and -6 dB, with younger listeners in light blue and older listeners in light red. The three black horizontal lines in each figure indicates the ranges from which three peaks are extracted. Shaded areas: ± 1 SEM. Figure **B** compares peak amplitudes in older listeners (red bars) with younger listeners (blue bars). Similar to the responses to foreground speech, the older listeners' responses have significantly larger peaks than younger listeners with 2-sample one-tailed *t*-tests on the averaged peak level over SNR conditions (*$p<0.05$, **$p<0.01$, ***$p<0.001$). Additionally, the MI50 level is notably larger than the other two peaks, for both groups. This pattern demonstrates representation-to-suppression mechanism for background processing. Error bars: ±1 SEM.

4.3.4 MI200 relationships with behavioral performance

As can be seen in Figures 4.3 and 4.4, the dependence of the MI200 peak level on SNR condition follows different trends for older and younger listeners. Notably, for younger listeners, the MI200 response remains steady as SNR decreases for foreground speech while it decreases for background speech. However, for older listeners, the response to foreground decreases as SNR decreases, while the response to background increases as SNR decreases. MI200 saliency can then be defined as the difference between foreground and background information (Figure 4.5A), and any trends as a function of SNR can be analyzed via the slope of difference-by-SNR linear regression line (Figure 4.5B). A right-tailed 2-sample *t*-test is performed on the slopes of younger listeners against the older, resulting in a significantly larger slope for younger than older listeners ($t_{30} = 2.31, p = 0.014$). To test the positivity of the ratio as SNR decreases in the younger participants, a right-tailed 1-sample *t*-test is conducted on the slopes of younger listeners, and the results show a significant positive trend as SNR decreases ($t_{16} = 1.83, p = 0.043$). Similarly, a left-tailed 1-sample *t*-test against zero on slopes of older listeners show a negative trend but not significant ($t_{14} = -1.47, p = 0.083$) (Figure 4.5B). In short, age affects the response pattern (with increasingly challenging mixed speech conditions) of this late cortical representation.

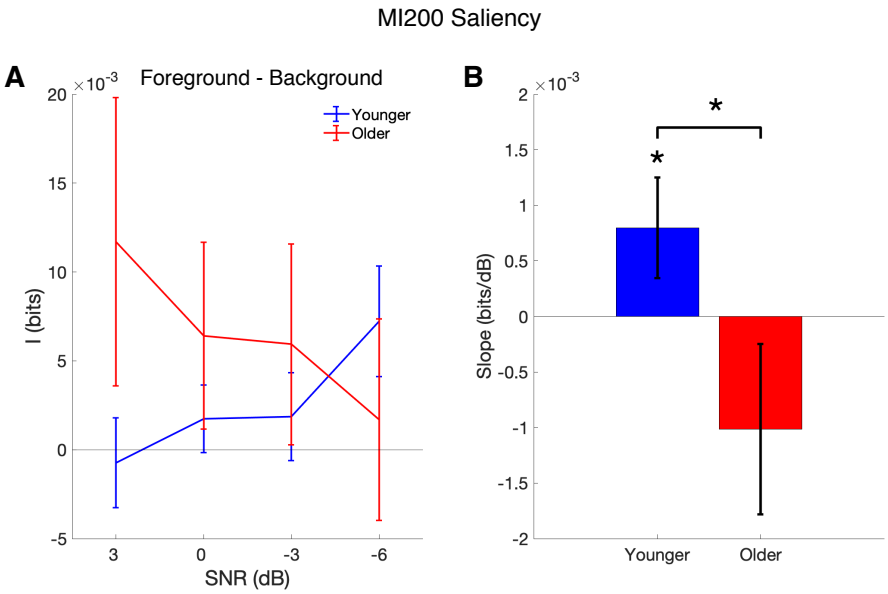

Figure 4.5. MI200 level difference between foreground and background as a function of SNR in younger and older listeners. **A**. Younger listeners (blue) demonstrate a significant increasing trend with decreasing SNR, while the older (red) demonstrate a decreasing trend. **B**. MI200 ratio slope as a function of SNR for the two age groups. Younger listeners have a significantly positive slope (linearly fitted regression to the data shown in panel A), while older listeners show a weakly negative slope (not statistically significant). The slope difference between groups is significant. (*$p<0.05$, **$p<0.01$, ***$p<0.001$)

The different MI200 saliency trend by age suggest functional differences in neural suppression of the background and/or enhancement of foreground representation for older listeners as SNR level decreases. These abilities may be related to behavioral inhibitory and attentional control. A Pearson's test was performed on correlations between foreground MI200 and Flanker score for each age group under every SNR

condition, with p-values corrected by Bonferroni-Holm for multiple tests across SNR conditions (Holm 1979), to examine if the foreground MI200 relates to behavioral inhibitory control. The result shows significantly negative correlations across all SNR condition ($r = -0.53, p = 0.04$ at +3 dB, $r = -0.77, p = 0.0014$ at 0 dB, $r = -0.82, p < 0.001$ at -3 dB, and $r = -0.89, p < 0.0001$ at -6 dB) for older listeners, and the correlation coefficient increases as SNR decreases. In contrast, younger listeners show no correlation between MI200 and Flanker score ($r = 0.40, p = 0.44$ at -6 dB). The overall results suggest that foreground MI200 relates to behavioral inhibitory control in older listeners. Correlations are shown in Figure 4.6.

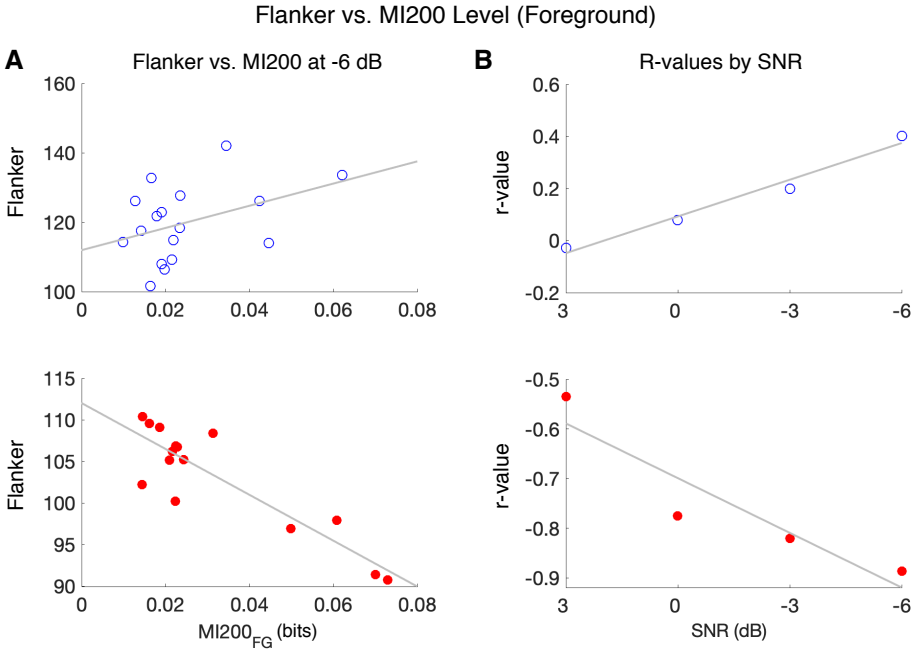

Figure 4.6. Correlation between foreground MI200 and Flanker test score by age and SNR. **A**. The relationships of foreground MI200 and Flanker test scores under the most challenging condition, -6 dB SNR, for younger (blue) and older listeners (red) The r-value of this regression becomes the rightmost data point in the corresponding

plot in panel B. **B**. Correlation tests were performed for every SNR level, with p-values corrected by Bonferroni-Holm corrections. Younger listeners show a weakly (non-significant) positive correlation, but the older listeners show a significantly negative correlation, with the correlation coefficient growing larger in its absolute value as SNR decreases.

Since the speech-in-noise behavioral score is negatively correlated with the Flanker behavioral inhibitory score in older listeners (Figure 4.1), the foreground MI200 may also correlate with the QuickSIN score. A Pearson's correlation test shows a significant positive correlation between the QuickSIN and foreground MI200 level ($r = 0.60, p = 0.018$) for older listeners. A stepwise regression, testing if Flanker and MI200 both contribute to QuickSIN performance, shows that only MI200 but not Flanker contributes to QuickSIN performance ($F_{(1,13)} = 7.27, p = 0.018$). These results demonstrate that higher higher MI200 level corresponds to worse hearing performance for older listeners.

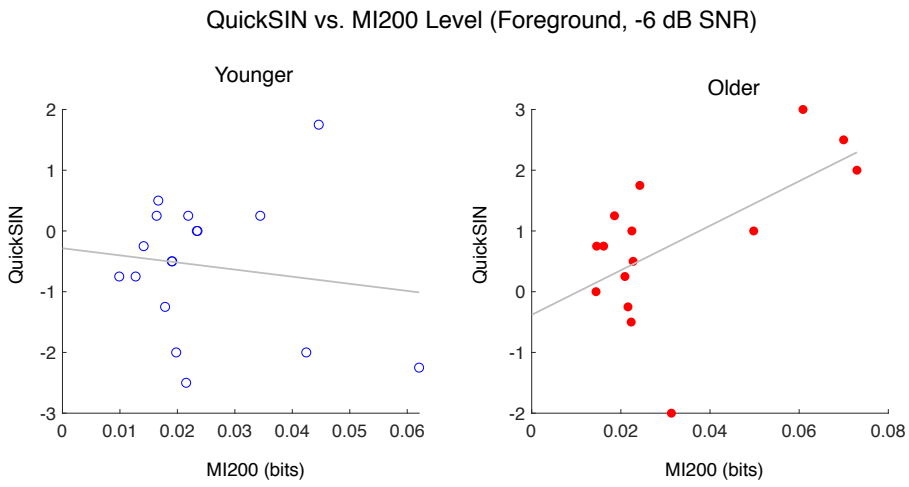

Figure 4.7. MI200 of the foreground correlations with speech-in-noise behavioral score. **A**. A significant correlation is not seen in younger listeners (blue). **B**. The correlation is significant in older listeners (red). A stepwise regression shows only the MI200, and not the Flanker, contributes to QuickSIN performance.

4.4 Discussion

By developing a novel approach based on information theory, phase-locked cortical responses to the low-frequency speech envelope can be measured without resorting to linear-only statistics. The TMIF unveils different processing stages in cortical response to speech, via the mutual information peaks MI50, MI100 and MI200. All three of these peaks in mutual information are larger for older adults than younger adults in all conditions. The MI200 stands out differently, however, since difference between foreground and background levels has a different pattern of dependences on SNR for the two age groups: while the ratio in younger listeners increases with worsening SNR, it decreases in older listeners. Plausibly related, the MI200 response to foreground negatively correlates with Flanker behavioral inhibitory scores regardless of SNR level, and it also correlates to the QuickSIN score in the most challenging noise condition. When compared to analysis results based on purely linear methods, such as speech envelope reconstruction accuracy (Presacco et al. 2016a), these mutual information based measures, e.g., the MI100 and MI200, reveal additional significant findings, even for the same data set.

4.4.1 Exaggerated information in older cortex: potential mechanisms

Exaggerated speech information for older listeners is seen at both short (MI50), intermediate (MI100), and long (MI200) latencies and in both clean speech and adverse conditions. One possible contribution to this may be loss of neural inhibition between synapses (Caspary et al. 2008; Takesian et al. 2012). Animal studies show decreased release of inhibitory neurotransmitters, such as gamma-aminobutyric acid (GABA), in auditory cortex (Juarez-Salinas et al. 2010; de Villers-Sidani et al. 2010). Such a reduction in neural inhibition might occur as part of a compensatory gain mechanism (Caspary et al. 2008; Takesian et al. 2012). The aging midbrain shows deficits in temporal processing acuity in normal-hearing CBA mice (Walton et al. 1998), and the cortex is able to restore auditory processing even with a cochlear denervation and virtually eliminated brainstem response (Chambers et al. 2016). Similar exaggerated responses are also seen in cases of tinnitus and hyperacusis, at multiple levels along the auditory pathway (Auerbach et al. 2014). The early response of MI50 is pre-attentive and based on acoustics not the perception. This response might be related to the detection of auditory scene. The enlarged response at subcortical level may also induce an enlarged early response at the cortical level.

Another potential contributor to exaggerated information in the aging cortex might be due to the utilization of more neural resources in cognitive processing, such as redundant local processing (Peelle et al. 2010) and exaggerated effort and attention (Presacco et al. 2016a). Older listeners allocate more neural resources outsize the core sentence-processing network and demonstrate reduced coherence between activated

regions (Peelle et al. 2010). Cortical responses with ~100 ms latency have been shown to be enhanced by attention (Ding and Simon 2012b, 2013). Therefore, MI100 response level may be more affected by attention than acoustics, and older listeners may allocate more neural sources or higher power for auditory attention.

Multiple cortical representations of contextual information in older listeners might also contribute to this age-related exaggerated information. Older listeners' speech understanding benefits from different levels of supportive context, such as sentential, lexical, phonological and sub-phonemic levels (Pichora-Fuller 2008). Embedded within the frequency range of 1-8 Hz (Cogan and Poeppel 2011), such contextual information enhancement for older listeners not only corresponds to a larger amplitude (Presacco et al. 2016a) but also induces more information in cortical response to low-frequency speech envelope as shown by the present study. MI200 is a late response that may relate to the contextual information processing, and shows an age-related enlargement.

4.4.2 Long latency processing, distractor suppression and speech-in-noise intelligibility

The MI200 is the latest of the three peaks, and is stronger in older listeners than younger, so it becomes a viable candidate for the processing of redundant speech information. The negative correlation between the MI200 and the Flanker score suggest that this later neural activity might be related to the same neural source with the behavioral inhibitory control for older listeners. It may also lend support to a compensatory hypothesis (Caspary et al. 2008; Presacco et al. 2016a, 2016b; Takesian

et al. 2012) that an insufficient speech representation at earlier latencies induces larger responses at later latencies, where the compensated response would have a negative correlation with speech intelligibility. Here, for older listeners, the MI200 salience decreases with worsening SNR, as the response to the background is strengthened, suggesting that compensation can no longer suppress background speech when it reaches higher sound levels. Older listeners show a trend, as SNR decreases, for MI200 saliency (foreground over background) that is consistent with this view. The MI200 saliency for younger listeners, however, for whom these SNRs cause only modest difficulty, show a trend in the opposite direction.

The EEG correspondence of MI200 could be P2 (P200) or N2 (N200) based on its response latency (~200 ms). However, MEG shows little consistent activity at the time of occurrence of N2 and P3 (Siedenberg et al. 1996). The current study demonstrates consistent MI200 across SNR levels, and thus it more likely corresponds to P2 response in EEG. A previous study with syllable stimulus supports evidence for P2(m) response in both EEG and MEG, and both radial and tangential neural sources contribute to the response (Shahin et al. 2007). Various hypotheses have been proposed to explain the functions of P2. Shahin et al. (2007) suggest that P2 is modulated by spectral complexity of sounds. Speech-sound training studies have shown that N1-P2 complex changes after training, which reflect central auditory plasticity change (Tremblay et al. 2001). A further study suggests P2 amplitude changes after training are retainable for months, and are attributed to neural activity changes associated with the acquisition process (Tremblay et al. 2014). These studies support a possible

hypothesis of a retainable neural plasticity change associated with MI200 over the lifetime listening for older listeners. Other studies support hypothesis that this late response of ~200 ms in latency also associates with auditory working memory and may have potentials to serve as neurophysiological markers for the assessment of working memory capacity (Lefebvre et al. 2005). Visual studies suggest visual P2 may be associated with processing contextual information to prepare for the visual analysis of upcoming stimuli (Federmeier and Kutas 2002). These studies suggest that P2 might be a cross-modality response that relate to processing of contextual information and working memory. The current study extends the hypothesis by demonstrating a strong correlation between the late auditory response and visual behavioral inhibitory ability, suggesting, possibly that a similar neural source or mechanism may contribute to both P2 or MI200 and these behavioral abilities such as working memory and behavioral inhibitory control.

In conclusion, mutual information analysis provides a non-linear approach towards decoding temporal response function to continuous speech. The mutual information representation has higher predictive power of behavioral measures compared to linear representations. By this novel approach, the current study shows that with aging, the cortical response to speech is not only larger in amplitude but also redundant in information. And the late response at latency about 200 ms is an important response component for older listeners, predicting both behavioral inhibitory control and speech intelligibility.

Chapter 5. Cortical High-gamma Response to Speech in Noise and the Effects of Aging, using Mutual Information

5.1 Introduction

The auditory brainstem response (ABR) is phase-locked neural response that tracks both the onset transient and any periodic component of a sound (Skoe and Kraus 2010). The sustained response to the latter, which has been called envelope-following responses (Aiken and Picton 2006) and auditory steady-state responses (Dimitrijevic et al. 2004), is now known as frequency following response (FFR) (Aiken and Picton 2008b; Galbraith et al. 1995; Greenberg 1980; Krishnan et al. 2004; Russo et al. 2004) where the oscillation rate is at the acoustic fundamental frequency (F0), the lowest frequency of a periodic waveform. Neural sources for the onset responses and FFR have been hypothesized originate mainly from subcortical stages of the auditory system, including the cochlear nucleus (CN), geniculate nucleus (MGB) and inferior colliculus (IC) (Batra et al. 1986; Chandrasekaran and Kraus 2010; Schnupp et al. 2011; Worden and Marsh 1968). Recent studies using magnetoencephalography (MEG) have modified this hypothesis by providing evidence of cortical contributions to FFR, which falls within the frequency range of gamma band (Coffey et al. 2016b; Hertrich et al. 2012). In this study, high gamma waves is defined as rhythms from 60 to 100 Hz. Recent research generalizes FFR to continuous speech by estimation of temporal response function (TRF) (Forte et al. 2017; Maddox and Lee 2018). The use of continuous speech stimuli also allows a natural extension to the case of competing speakers, where one speaker is attended (foreground) and the other unattended

(background). FFR may, or may not, show modulation due to selective attention, depending on the specific methodology employed (Forte et al. 2017; Hoormann et al. 2004; Lehmann and Schönwiesner 2014; Varghese et al. 2015).

Behavioral studies have demonstrated age-related temporal processing deficits in different auditory tasks (Fitzgibbons and Gordon-Salant 2001; Fitzgibbons and Gordon-Salantt 1996; Frisina and Frisina 1997; Gordon-Salant et al. 2006). Results based on neurophysiological studies are consistent with observed age-related changes in response latency and strength in midbrain, i.e., delayed and decreased midbrain FFR (Anderson et al. 2012; Burkard and Sims 2002; Clinard and Tremblay 2013) and cortical evoked response, i.e., exaggerated cortical response to low-frequency (1-8 Hz) speech envelope (Lister et al. 2011; Presacco et al. 2016a, 2016b). However, whether cortically generated high-gamma response would be enhanced or reduced in older listeners remains unknown.

Mutual information, interpreted as a reduction in auditory response variability due to the presentation of a stimulus (Nelken and Chechik 2007), has been proved to be useful in auditory research (Rieke et al. 1995). It is applied to measure the amount of information contained in magnetoencephalography (MEG) auditory responses about the continuous speech (Cogan and Poeppel 2011). Bidelman et al. (2014) applies it to measure the information transmitted from midbrain to auditory cortex, and shows redundancy in older listeners. My recent work proves it to be fruitful and powerful in

studying aging effects of midbrain responses (Zan et al. 2019; Chapter 3) and cortical low-frequency responses (Chapter 4).

In the study, neural responses of both attended (foreground) and unattended (background) speech, represented by mutual information, are presented, but due to large difference in high-gamma component between the foreground and background speakers, and that each speaker is only ever attended or unattended (without switching across trials), one representation being larger than the other does not lead to any evidence for attentional modulation. The focus of this investigation is cortically generated high-gamma response to both foreground and background speech, including the effect of changing SNR, and to what extend aging has any effect. Specifically, MEG high-gamma responses to continuous speech were analyzed in frequency range of 60-100 Hz. Here, by integrating informational theoretical measures suitable for auditory responses (Nelken and Chechik 2007) and MEG source localization (Gramfort et al. 2013, 2014), it is expected to estimate the time-locked response to continuous speech in a cocktail party scenario.

Based on midbrain-generated FFR results of Zan et al. (2019), it is hypothesized that the cortical high-gamma response to foreground speech decreases as SNR decreases. It is also expected that the background response increases as SNR decreases. Furthermore, since earlier studies have shown an enlargement of low frequency cortical auditory responses in older listeners, e.g., (Presacco et al. 2016a), the same effect may hold true for the high frequency cortical responses analyzed here.

5.2 Materials and Methods

5.2.1 Dataset

The dataset was collected previously and has already been analyzed using other methods (Presacco et al. 2016a, 2016b).

5.2.2 Subjects

A total number of 32 subjects participated in the experiment, with 17 younger adults aged between 18 and 27 (3 male), and with 15 older adults aged from 61 to 73 (5 male). All participants were recruited from the greater Washington area (Maryland, Virginia and Washington D.C.) and had clinically normal hearing (see Presacco et al. 2016a, 2016b for details). Written informed consent was obtained from each subject before the experiment. The experiment protocol and all procedures were reviewed and approved by Institutional Review Board of the University of Maryland.

5.2.3 Stimuli and recording

For each subject, the magnetic fields responses were recorded in an electromagnetically shielded room (insert manufacturer and location) by a KIT 157-channel MEG scanner (Kanazawa, Japan), sampled at 1000 Hz with online low-pass filter of cut-off frequency 200 Hz. The stimuli were continuous speech (a narrated audio book), either from a solo speaker or a mixture of two concurrent speakers. The solo-speaker speech stimuli were one-minute segments from the audiobook, *The Legend of Sleepy Hallow* by Washington Irving, narrated by a male speaker (http://www.audiobooktreasury.com/legend-of-sleepy-hollow/). The mixture was

composed of foreground speech to which the subject was instructed to attend and a background, which served as a distractor. The foreground speech was from the same source as the clean speech condition. The background stimuli were one-minute segments from the audiobook, *A Christmas Carol* by Charles Dickens, narrated by a female speaker (http://www.audiobooktreasury.com/a-christmas-carol-by-charles-dickens-free-audio-book/). The foreground and background speech segments were mixed together at four different power ratios, of 3 dB, 0 dB, -3 dB and -6 dB. The foreground speech used in -6 dB condition and the clean speech were the identical, and the clean speech was only presented after all the mixed speech stimuli had been presented. The stimuli were all presented to the subjects with E-A-RLINK earphones attached with sound tubing at about 70-dB sound pressure. More details of the experiment can be found in (Presacco et al. 2016a, 2016b).

5.2.4 Cortical responses across frequency bands and high-gamma response

The MEG recordings were first cleaned by time-shifted principle component analysis (TSPCA) (de Cheveigné et al. 2007), which uses the three reference (noise) channels to subtract off environmental noise in the data. The denoised data were then feed into two separate analysis pipelines for auditory component analysis and source space analysis.

Frequency bands

Different frequency bands were analyzed to estimate the amount of information contained in these bands. The cutoff frequencies for these bands are 1-12 Hz, 7-13 Hz,

12-30 Hz, 35-45 Hz, 45-55 Hz, 65-75 Hz, 75-85 Hz, 85-95 Hz, 95-105 Hz, and 100-115 Hz. For each band, an FIR band-pass filter with the corresponding cutoffs was applied to the signal. For the full high-gamma response, the 60-100 Hz frequency band was analyzed.

Auditory components

Denoising source separation (DSS) (de Cheveigné and Simon 2008b) was applied to the TSPCA-denoised signals to extract auditory components. The bias function was based on the band-passed signal. For example, for the high-gamma response, a bias function of band-passed 60-100 Hz response averaged across trials was applied. Then the signal was projected into DSS space and filtered to 60-100 Hz by an FIR filter. For the low-frequency speech envelope response, the signal was band-pass filtered to 1-12 Hz band by an FIR filter after projection to a DSS space computed by bias function of averaged 1-12 Hz response (Ding and Simon 2013).

Source space high-gamma response

Before and after each experimental session, head positions for each subject were recorded and used to locate the subject's head shape in MEG coordinate (Brodbeck et al. 2018c). An averaged brain from FreeSurfer was then used to co-register a single brain map to each subject's individual digitized head shape. A source space with 5124 dipoles was then constructed. All these steps were done using mne-python (Gramfort et al. 2013, 2014), described in greater detail in (Brodbeck et al. 2018c). Finally, the TSPCA-denoised MEG sensor-space data was filtered to 60-100 Hz for the FFR response analysis. The filtered data was then projected to source space

with fixed orientation by minimum norm estimate (MNE) in mne-python. A response matrix of size 5124 sources-by-1 min for each subject was obtained for each noise condition separately.

Mutual information analysis

Both the auditory component, represented by the first DSS component, and the source space responses were analyzed using a mutual information approach, which is called temporal mutual information function (TMIF) analysis (Chapter 4). First, both the speech representation and response levels were quantized into 8 bins based on amplitude. Then the TMIF was estimated iteratively by shifting the response forward by a step size of 2 ms (within a 500-ms window) and computing the mutual information at each time point based on the joint distribution of integer-encoded (one of eight bins) speech representation and the shifted response. More detailed explanations are contained in the previous chapter (Chapter 4). Two main speech representations are used, the speech waveform band-passed to 60-100 Hz for high-gamma analysis, and the 1-8 Hz Hilbert envelope for low-frequency response analysis. For the auditory component analysis, the TMIF was estimated for the first DSS component, separately for high-gamma response and low-frequency response, which were denoted as $TMIF_{HG}$ and $TMIF_{low}$, respectively. For the source space analysis, the TMIF was estimated separately for each source, for both the high-gamma response and low-frequency response, respectively. In competing-speaker conditions, the TMIF was estimated for attended speech (foreground), and unattended speech (background) separately.

Measurements and statistics

For auditory component measures, $MI50_{HG}$ peaks for both foreground and background were extracted by finding the maximum value in the time range of 45 ms and 55 ms for each individual, and their amplitude was analyzed. For source space analysis, a noise floor $TMIF_{HG}$ was calculated by mutual information estimation between the response and a speech sample not used in the experiment, and then was used to test significance of the response across a time window of 0-100 ms and in the region of interest (ROI) of Heschl's gyrus and superior-temporal gyrus (Brodbeck et al. 2018b). One-sample *t*-tests were used for virtual dipoles and time points, and the multiple comparisons were compensated by threshold-free cluster enhancement (TFCE) (Smith and Nichols 2009).

5.3 Results

5.3.1 Cortical responses across frequencies

TMIFs were estimated for the auditory component, separately for a large set of frequency bands ranging from 1-12 Hz to 100-115 Hz (example shown in figure 5.1B). For each frequency band, the peak value in the time range of 20-150 ms was compared across subjects. As seen in Figure 5.1C and 5.1D, younger and older listeners appear to show approximately equal levels for all bands but the lowest. Even for the largest apparent age difference, the frequency bands centered at 70, 80, 90 and 100 Hz for the -6 dB SNR condition, the group difference was not significant ($p > 0.05$). An ANOVA test over all frequency bands, MI ~ age × frequency, showed significant age × frequency interaction ($F_{(1,\ 326)} = 9.58$, $p = 0.002$), suggesting a different mutual

information trend as a function of frequency band for older listeners. To examine age effects on low frequency (< 21 Hz) and high frequency (>=21 Hz), respectively, separate linear models of MI ~ age × frequency were tested on the subset data with these two frequency restrictions. Results showed that for low frequency model (significant in itself with $F_{(3, 62)} = 16.92$, $p < 0.001$), older listeners had a significant 0.05 bits larger mutual information than younger listeners on average ($t_{(62)} = 2.64$, $p = 0.010$), and mutual information significantly decreased at a rate of 0.01 bits/Hz averaged over the two groups ($t_{(62)} = -5.90$, $p < 0.001$). It also demonstrated a significant age × frequency interaction with coefficient of -0.005 bits/Hz with younger listeners as reference for age factor ($t_{(95)} = -2.13$, $p = 0.037$), suggesting the decreasing slope for MI-by-frequency is significantly steeper (0.005 bits/Hz steeper) for older listeners. For frequencies above or equal to 21 Hz, a same linear model was constructed (significant in itself with $F_{(3, 260)} = 3.70$, $p = 0.012$). The testing results showed that no significant age ($t_{(260)} = 0.51$, $p = 0.609$) or frequency ($t_{(260)} = 0.37$, $p = 0.709$) effects, and that no significant age × frequency interaction was observed ($t_{(260)} = -1.45$, $p = 0.147$), suggesting that older listeners had comparable responses in high frequency bands as in younger listeners.

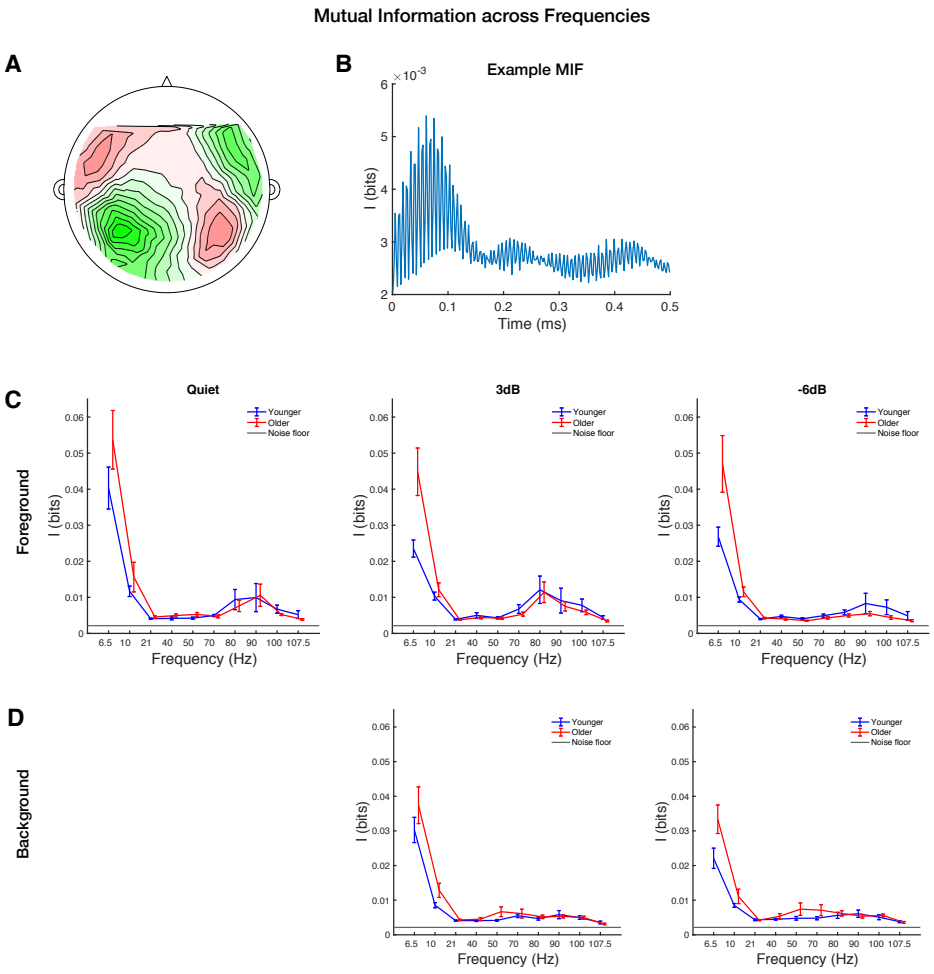

Figure 5.1. Mutual information across frequency bands. A. The magnetic field distribution associated with an auditory component for a representative subject. B. An example mutual information function. C. Mutual information trend as a function of frequency in quiet, 3 dB and -6 dB condition for foreground speech. Older listeners have stronger response and a steeper decreasing rate, than younger listeners in low frequency bands, 1-12 Hz and 7-13 Hz, but comparable responses in frequency bands above the former two. D. Mutual information trend for background speech.

5.3.2 Cortical high-gamma response

Auditory component $TMIF_{HG}$

The $TMIF_{HG}$ was estimated based on the response of the auditory component for the frequency band 60-100 Hz (Figure 5.2A and 5.2B)

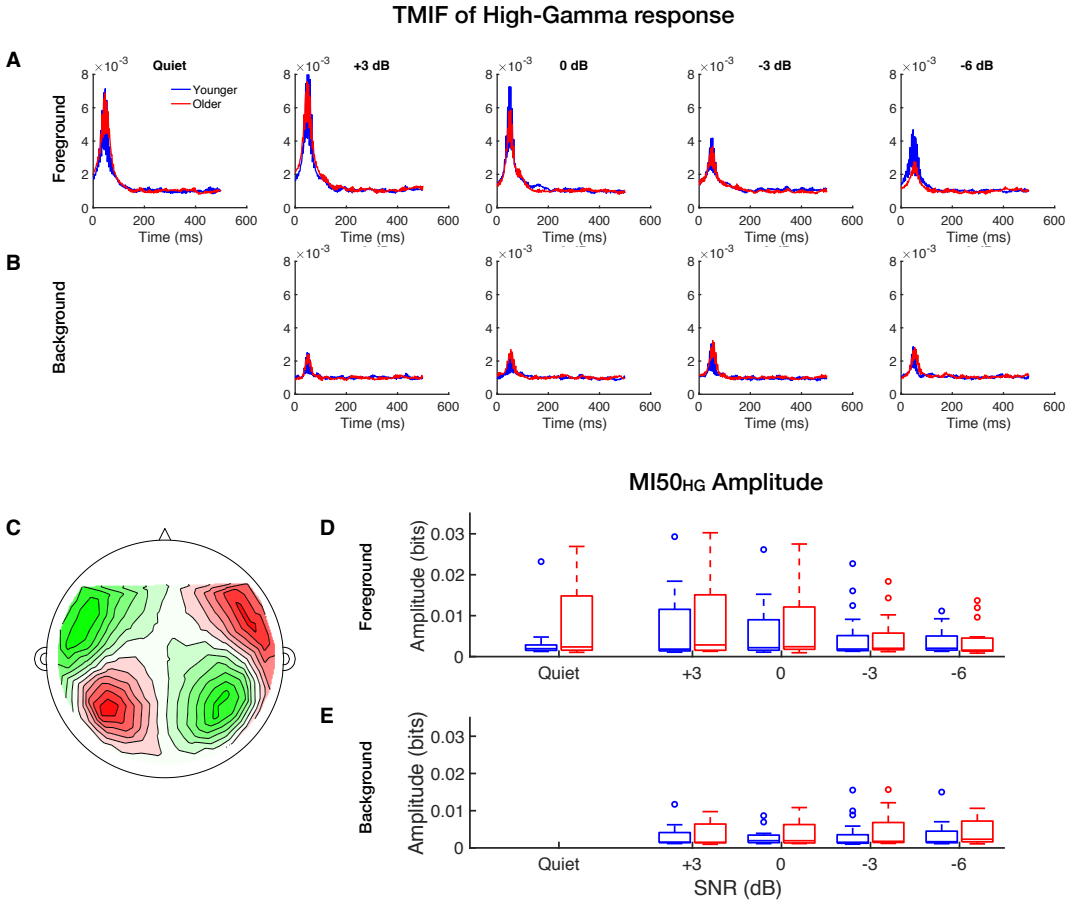

Figure 5.2. TMIFs from responses in the high-gamma band (60-100 Hz). A. $TMIF_{HG}$ for speech in quiet and for foreground speech. Younger listeners (blue) and older (red) show both showed peak responses at around 50 ms. B. TMIF for background speech. The peak latency remains near 50 ms. C. The magnetic field distribution

associated with an auditory component for a representative subject. D. $MI50_{HG}$ amplitude for foreground speech for all conditions, with younger in blue and older in red boxplots. E. $MI50_{HG}$ amplitude for background speech.

Compared with conventional slow cortical TMIF, the $TMIF_{HG}$ only has one peak, with latency around 50 ms, named here the $MI50_{HG}$. When compared between age groups for each condition and attentional focus (Figure 5.2D and 5.2E) no significant difference was found ($p > 0.05$).

To test whether foreground and background $MI50_{HG}$ was correlated and the aging effects, an ANOVA test was performed on $MI50_{HG}$ (foreground) ~ $MI50_{HG}$ (background) × age. The results showed a significant $MI50_{HG}$ (background) × age interaction ($F_{(1, 124)} = 16.65$, $p < 0.001$). Linear models were constructed for testing the prediction slope difference between groups. A linear model of $MI50_{HG}$(FG) ~ $MI50_{HG}$(BG) × age was tested. The linear model was significant ($F_{(3, 124)} = 37.06$, $p < 0.001$). Consistent with ANOVA test, results showed that $MI50_{HG}$(BG) had a statistically significant prediction power at a rate of 1.15 ($t_{(124)} = 4.25$, $p < 0.001$), and that older listeners showed a significantly shallower (1.60 shallower than younger) prediction slope ($t_{(124)} = 4.08$, $p < 0.001$). Then a linear model of $MI50_{HG}$ (foreground) ~ $MI50_{HG}$ (background) was tested separately for each group to examine the prediction slope in each group. The linear model shows a slope of 2.75 ($t = 10.04$, $p < 0.001$) for younger listeners and a slope of 1.15 ($t = 4.07$, $p < 0.001$) for older.

Given testing results from ANOVA and linear regression, Pearson's correlation test was performed to test $MI50_{HG}$ foreground-background correlation for each group for different SNR conditions with false discovery rate correction for multiple comparisons (Benjamini and Hochberg 1995). The results are shown in figure 5.3. Younger listeners showed significant correlations for all SNR conditions, while older listeners showed significant correlations in all but -3 dB condition. Figure 5.3A showed the scatter plot of $MI50_{HG}$ averaged across SNR levels ($r = 0.95$, $p < 0.001$ for younger listeners, $r = 0.65$, $p = 0.006$ for older listeners).

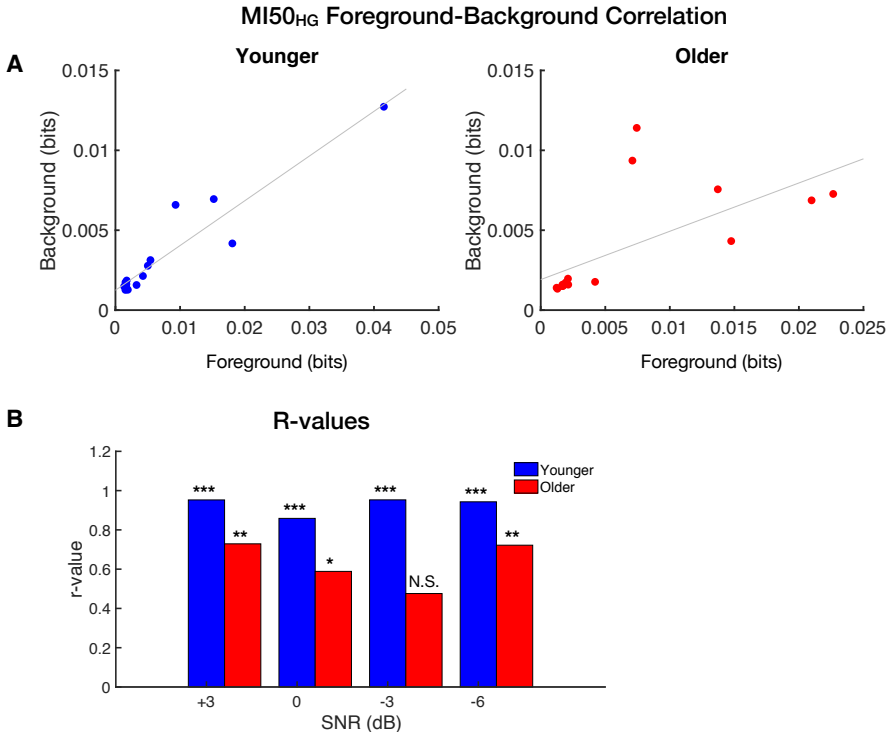

Figure 5.3. Relationship between $MI50_{HG}$ of foreground and background. Both younger (blue) and older (red) listeners showed significant correlations, but the younger had a higher prediction slope. A. Foreground-background $MI50_{HG}$ relationships demonstrated for the single condition of +3 dB SNR for younger (left)

and older (right). B. Correlation coefficients at all SNR levels. Stars indicates the statistical significance (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

TMIF for source space analysis

To investigate whether the neural sources localize to auditory cortex, the $TMIF_{HG}$ for 5124 sources across the whole brain was estimated for the cleans speech condition (Figure 5.4). By statistical testing mentioned in the methods section, the results showed significant responses for both younger (p<0.001) and older listeners (p<0.001). Independent samples t-tests showed no significant difference between the two groups (p>0.05).

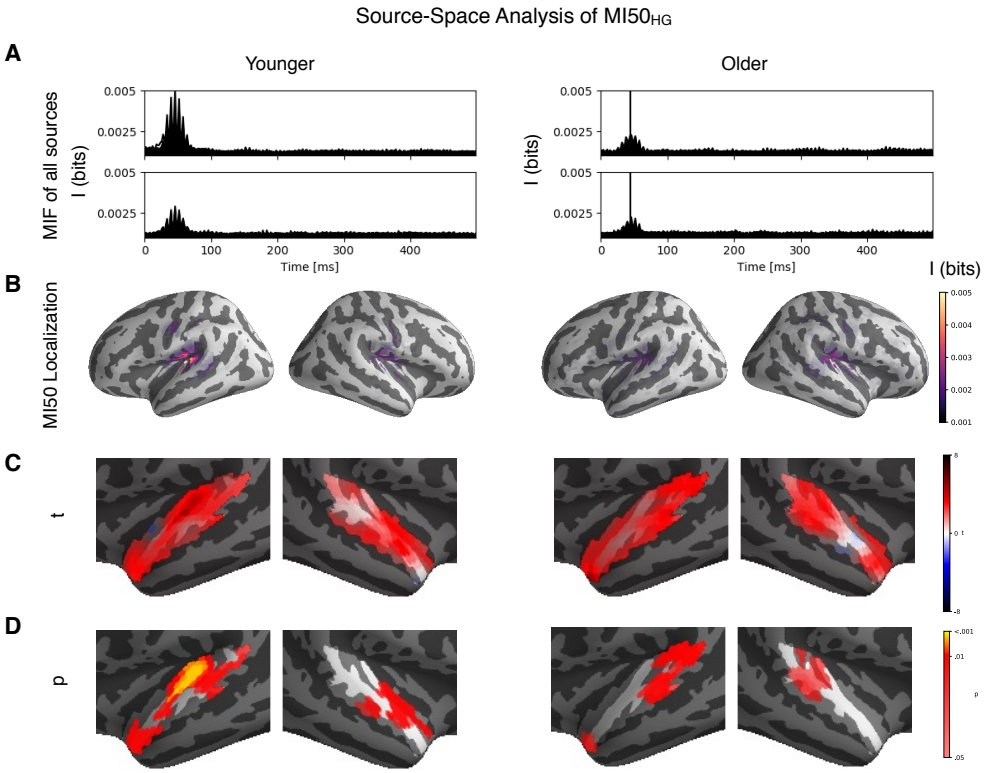

Figure 5.4. Source space TMIF in the clean speech condition. Neural sources for the $MI50_{FFR}$ localized to auditory cortex for both younger and older listeners. Statistical test on response in the ROI region showed significant response against noise floor for both groups. A. TMIF for all 5124 sources. B. $MI50_{HG}$ distribution around the whole brain. C. *t*-values of significance test across ROI region. D. *p*-values across the ROI.

5.3.3 Cortical low-frequency response

The mutual information for the cortical low-frequency response was also examined, to compare with the cortical high-gamma response results above. Both the $MI50_{TRF}$ and $MI100_{TRF}$ low-frequency response peaks were tested for associations between foreground and background measures with false discovery rate correction for multiple comparisons across SNR conditions. The results showed that the younger listeners had significant correlations for the $MI50_{TRF}$ in the two worst noise conditions, -3 dB ($r = 0.66$, $p = 0.015$) and -6 dB ($r = 0.59$, $p = 0.026$). Older listeners did not show such significance ($p>0.05$) for any condition, nor did the $MI100_{TRF}$, for either group.

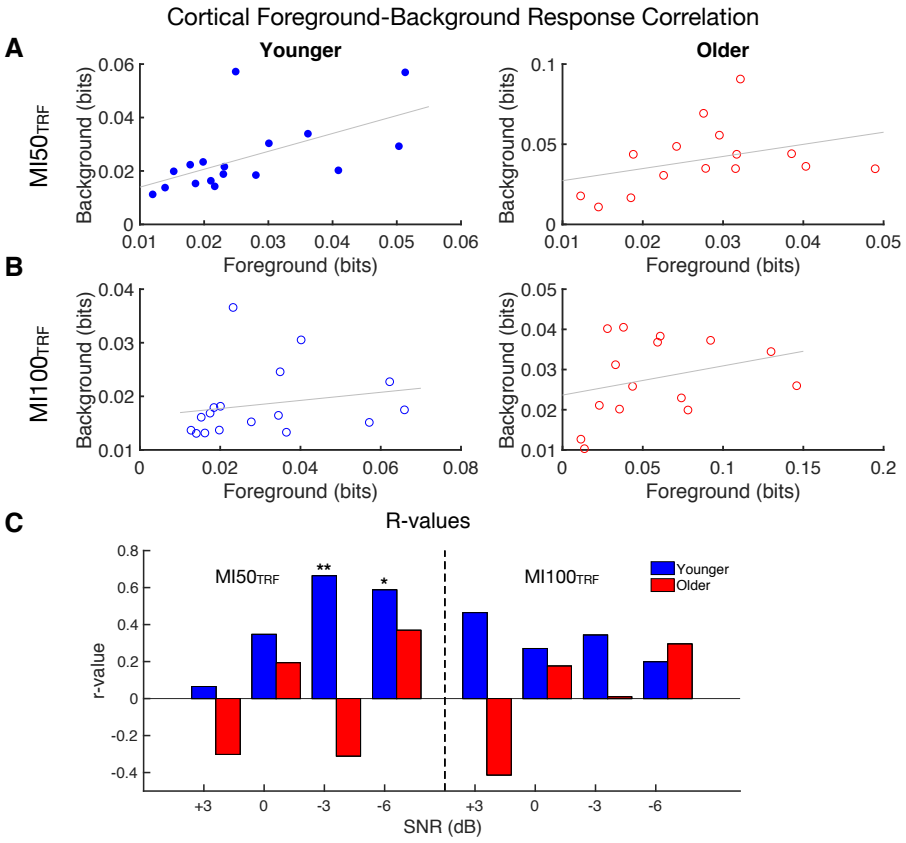

Figure 5.5. Relationship between foreground and background low-frequency speech envelope (1-8 Hz) mutual information representations in cortex in an example condition of -6 dB SNR. A. Younger listeners show a significant correlation between responses to foreground and background (left plot, filled blue circles), while older listeners do not show such a relationship (right plot, empty red circles). B. No significant correlation is found for the $MI100_{TRF}$ in the same condition for either group. C. r-values for each condition.

5.3.4 Noise level effects

Auditory DSS component

To examine the effect of noise level on the $MI50_{HG}$, a linear regression was performed for $MI50_{HG}$-by-SNR for each subject and for foreground and background speech, respectively (Figure 5.6). The slope was tested against 0 by a one-tailed *t*-test separately for younger and older listeners. The results for foreground showed both groups have significantly positive slope in the direction of increasing SNR ($t = 1.79$, $p = 0.046$ for younger; $t = 2.54$, $p = 0.012$ for older). However, the results for background speech showed significantly negative slope in the direction of increasing SNR only for older listeners ($t = -0.10$, $p = 0.167$ for younger and $t = -1.84$, $p = 0.044$ for older).

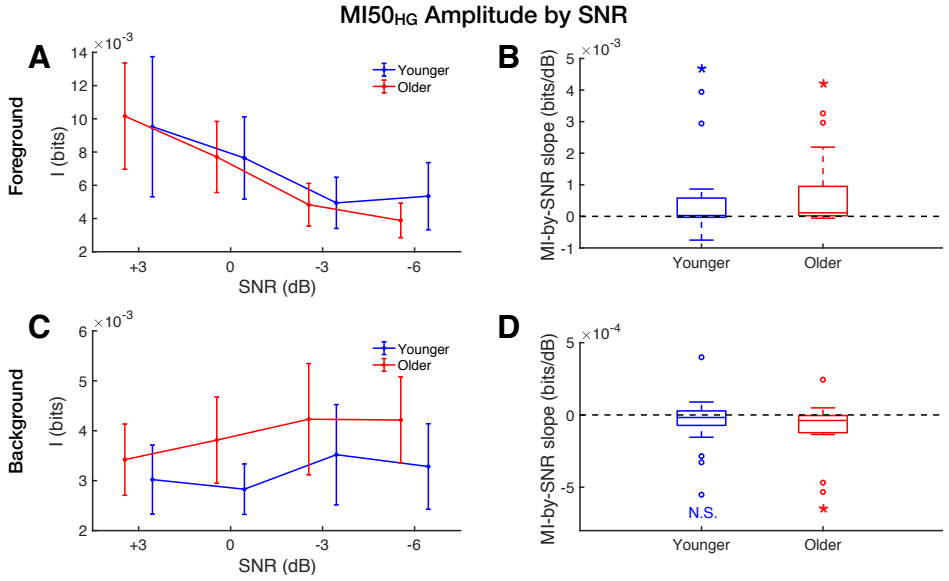

Figure 5.6. $MI50_{HG}$ amplitude as a function of SNR level for each age group. A. Foreground $MI50_{HG}$ amplitude as a function of noise level for younger (blue) older (red) listeners. B. The MI-by-SNR slopes (e.g. of the regression indicated in A) show decreasing trends as SNR worsens for both age groups. Older listeners appear to show

a steeper decrease than younger listeners, but the difference is not significant. C. Background $MI50_{HG}$ amplitude at the fundamental frequency as a function of noise level. The background showed an increasing trend as SNR decreases, or as background sound level increases. D. The MI-by-SNR slope was significantly smaller than zero for older listeners in red, but not significant for younger listeners. Stars indicates the slope significance against zero (*$p < 0.05$).

5.4 Discussion

These results show a cortical high-gamma response, time-locked to the relevant features of continuous speech, with a peak latency of around 50 ms and localized to auditory cortical areas. The results also strongly suggest that any cortical exaggerated representation seen for older listeners in time-locked low-frequency responses are not present at higher frequencies. Nevertheless, the foreground-background $MI50_{HG}$ slope is significantly decreased by aging. This cortical response is affected by noise level: specifically, the foreground $MI50_{HG}$ decreases with worsening SNR for both age groups. In contrast, the background $MI50_{HG}$ grows with increasing noise level but only for older listeners.

5.4.1 Cortical representation across frequencies and high-gamma response

According to previous studies, older listeners have exaggerated cortical responses to continuous speech (Presacco et al. 2016a, 2016b). By examining frequency bands from theta to high gamma, the results show that this overrepresentation is only present for the low-frequency response to speech envelope,

e.g., 1-12 Hz. Older listeners have comparable responses to younger listeners in the beta and gamma bands. This suggests that aging does not affect different processing rates in cortex in the same way. On the other hand, a recent study showed lower mutual information for midbrain-based FFR for older listeners (Zan et al. 2019; Chapter 3). Here, since no group difference was seen for frequency bands between 20 and 100 Hz, it may also suggest that the reduction for older listeners in the midbrain representation does not directly affect the cortical representation at the same frequencies.

Conventional auditory high-gamma responses are defined as event-related changes in spectral power in the 60-150 Hz frequency range (Cervenka et al. 2011). However, after examination of frequencies ranges from 1 to 140 Hz, for gamma-band response, by the proposed approach, analysis results demonstrate significant responses only in the frequency range of 60-100 Hz (Figure 5.1). Therefore, in the study, high-gamma response is defined in the frequency range of 60-100 Hz. The $TMIF_{HG}$ results show a consistent response peak around 50 ms across all SNR levels. Demonstrated by electrocorticographic studies, conventional auditory high-gamma responses (60-150 Hz) occur approximately 100 ms (75-120 ms) after stimulus presentation onset in response to phonemes (Crone et al. 2001), tones (Edwards et al. 2005) and click trains (Brugge et al. 2009; Howard et al. 2000). By examining cortical sources, recent MEG studies show clear cortical contributions to FFR in response to tones with a latency of 48-60 ms (Coffey et al. 2016a). The time-locked response to continuous speech shown in this study by TMIF demonstrates similar response latency. The neural sources localize to Heschl's gyrus and superior-temporal gyrus, which is consistent with the

previous studies (Coffey et al. 2016a, 2017; Steinschneider et al. 2011; Trautner et al. 2006).

5.4.2 Aging effects on high-gamma response

The current study then further examined the aging effects on high-gamma response. The amount of information contained in foreground response decreases as worsening SNR for both age groups. However, by demonstrating a significant increase in background information as SNR decreased only for older listeners, the results suggested the high-gamma response for background speech for older listeners is more easily affected by noise than younger listeners. Notice that to create different SNR conditions, sound level of foreground is fixed, and the SNR is decreased by increasing the background sound level. Therefore, the background response is expected to grow as worsening SNR unless it is successfully suppressed. The results, which demonstrated background response growth only for older listeners but not for younger listeners, suggested that the noise conditions are more challenging for older listeners, and the younger listeners are better at representing the actual acoustics than older listeners. The results of foreground $MI50_{HG}$ also agree with the previous findings in midbrain FFR, where the amplitude decreases with worsening noise level (Anderson et al. 2012; Presacco et al. 2016a, 2016b; Zan et al. 2019; Chapter 3), and the older listeners may spend more efforts suppressing the background (Presacco et al. 2016a, 2016b). According to previous studies, high-gamma responses are associated with multiple functions of auditory processing, including sound discrimination (Crone et al. 2001; Edwards et al. 2005; Fishman et al. 2004), phonological processing (Chang et al.

2010; Steinschneider et al. 2011), auditory selective attention (Herrmann and Knight 2001; Ray et al. 2008), auditory verbal memory (Herrmann et al. 2004; Kaiser et al. 2003) and auditory comprehension (Towle et al. 2008). The results of current study support the evidence of age-related deficits in selective attention in high-gamma response to adverse listening conditions.

The current analysis also shows a decreased foreground-background correlation in cortical high-gamma response for older listeners, compared with younger listeners. One possible reason would be that the neural oscillations in the aging cortex may contain more noise than younger (Presacco et al. 2016a).

5.4.3 Noise level effects on high-gamma response

In the results, the amount of information in the cortical high-gamma for foreground decreases as worsening SNR for both younger and older listeners, which are consistent with previous findings in midbrain FFR (Presacco et al. 2016, 2016; Zan et al. 2019; Chapter 3). The similar pattern against SNR for midbrain and cortex suggests a limited cortical modulation for representation of the acoustics. However, the finding that the older listeners also show an increasing trend for background as SNR worsens suggests that the suppression of background may not only happen in the low-frequency response (Ding and Simon 2012b, 2013), but may also happen around the fundamental frequency. And the older listeners are worse at background suppression in high-gamma band than younger listeners. This also suggests a selective attention starting at 50 ms with reference to the stimulus onset, which may degrade as aging.

Chapter 6. Summary and future work

6.1 Summary and discussion

6.1.1 Information theory and brain information processing

This dissertation studies auditory processing and the effects of aging by informational measures adopted and developed from information theory (Shannon 1948). Chapter 3 applies mutual information to study the amount of information contained in subcortical FFR. Chapter 4 modifies mutual information estimates to reveal the amount of information processed across time phase-locked to continuous speech envelope. Chapter 5 applies the same method to study information contained in higher frequency bands, with reference to low frequency of 1-12 Hz as in Chapter 4 and with emphasis on 60-100 Hz response, and also extends the method into source space analysis. Behind these studies is the analog of information processing for communication system and the human brain. According to John von Neumann (1958), the transmission error rate for a communication channel can be reduced by increasing the transmission redundancy, and this might also be a basis for the reliability of the brain information processing. The results from the dissertation support that for younger listeners, higher redundancy, as measured by mutual information may contribute to decreased error rate in auditory cortical representation. However, for older listeners, the increased error rate for subcortical representation may be compensated by an increased redundancy in cortical representation.

In Chapter 4, based on the MI peak latency, responses at ~50, ~100 and ~200 ms phase-locked to the speech envelope are named MI50, MI100 and MI200, which correspond to M50, M100 and M200, respectively in TRF. According to previous electrophysiological studies, M50 (MI50) is pre-attentive and more responsive for acoustics, and M100 (MI100) is more attentional modulated than M50 (Ding and Simon 2012b). Furthermore, shown by MEG studies, selective listening to sound in a complex auditory scene modulates longer-latency (~100-250 ms) responses, i.e., MI100 and MI200 in auditory cortex but not the shorter latency response (50 ms) (Ding and Simon 2012a; Gutschalk et al. 2008; Okamoto et al. 2011). Previous studies also show that P2, with the same latency as MI200, is modulated by spectral complexity of sounds (Shahin et al. 2007), associates with training-related neural plasticity change (Tremblay et al. 2001, 2014), and relates to auditory working memory (Lefebvre et al. 2005).

6.1.2 Age-related changes in speech representation in both cortical and subcortical responses

The three studies from Chapter 3 to Chapter 5 demonstrate age-related deficits in speech processing in both subcortical and cortical representations. Chapter 3 shows that the aging midbrain processes less information in FFR than younger listeners and is affected by informational masking. Chapter 4 shows age-related enlargement in the response to low-frequency speech envelope. Chapter 5 demonstrates in high-gamma band, the age-related enlargement from lower frequencies is absent. By the non-linear measure of mutual information, age-related changes are revealed for both subcortical and cortical responses that are otherwise hidden by linear methods. In Chapter 3, the

aging midbrain benefits from changing background speech from English to Dutch while the younger does not. This is not seen by a RMS measure (Presacco et al. 2016b). Chapter 4 demonstrates by TMIF that the aging cortex processes more information for all response latencies, i.e., MI50, MI100 and MI200. However, TRF analysis on the same dataset only shows significant age-related enlargement in M100 (Brodbeck et al. 2018b). Chapter 5 demonstrates age-related changes in high-gamma responses that no linear methods have found in known published works.

Chapter 3 includes Dutch speech as informational masking because Dutch is relatively close to English in terms of phonological inventory and prosodic contours (Collier and Hart 1975). This study intends to test the subcortical representation fidelity of the speech syllable without effect of attention. During the experiment, subjects were watching a quiet movie while listening to the presented stimuli. Auditory segregation of speech from noise/speech is a relatively complicated problem. It requires both bottom-up acoustic cues, such as spatial cues, pitch cues and timbre cues (Brungart et al. 2001; Shamma 2001) and top-down attention (Ding and Simon 2012b, 2012a; Kerlin et al. 2010) and prior knowledge about sound streams (Wang et al. 2019). Here by comparison between English and Dutch, the segregation problem is investigated by examining the effect of language, a higher level of feature. On the other hand, attention is required for studies in Chapter 4 and 5.

6.2 Future work

6.2.1 Informational measures of information transduction between subcortical and cortical responses and the effects of aging

Throughout all three studies, mutual information is applied to estimate the amount of information contained in the response about the stimulus. However, brain information processing includes information perception, transduction, coordination, storage and information creation (Rabinovich et al. 2012). For auditory information processing, it remains an open question that how much information is transmitted from subcortical to cortical auditory structures across different frequency bands. A previous study with simultaneous MEG and EEG has shown complementary effects for the two modality in studying radial and tangential long-latency (low-frequency) neural activities, which has both cortical and subcortical contributions (Shahin et al. 2007). Coffey et al. (2016) and Chapter 5 in current research have shown the feasibility of MEG recordings of cortical high frequency responses. EEG, on the other hand, has long been utilized in subcortical FFR studies (Anderson et al. 2012; Coffey et al. 2017; Presacco et al. 2016a). However, few studies have applied mutual information to study the amount of information transmitted from subcortical to cortical neural sources (Bidelman et al. 2014). Furthermore, even though discussions of MEG and EEG may clarify sensitivity difference with respect to the measured neural sources (Lopes da Silva 2013; Shahin et al. 2007), it remains an open question how much information in the response recordings by M/EEG comes from subcortical or cortical neural sources. Based on the developed methods described in Chapter 4 and 5, together with source localization (Gramfort et al. 2013, 2014) and information processing

inequality (Cover and Thomas 1991), the question of information transduction between subcortical and cortical neural sources can be attempted by a two-step random process, $X \rightarrow Y_{sub} \rightarrow Y_{cor}$, where X is stimulus representation, Y_{sub} is subcortical response, and Y_{cor} is cortical response. An extra model of Y_{sub} and Y_{cor} as a linear combination of Y_{MEG} and Y_{EEG} might be needed to link them to MEG and EEG recordings. The combination parameters can be solved by examining the source localization results.

Appendices

The mutual information results without averaging polarities from Chapter 3.

Analogously to the case of averaged polarities presented above, even without such polarity averaging, older listeners still demonstrate a slower fall-off in information as a function of SNR when the noise masker is Dutch than for English.

Information in amplitude of FFR without averaging polarities

For amplitude information, a regression line was fitted as a function of SNR to reduce within-subject variance. Using a one-tailed *t*-test on the y-intercept (effective mutual information benefit at 3 dB SNR) of the regression line against zero, the mutual information in amplitude benefit from the Dutch masker over the English masker is significantly higher for older listeners in the transition region ($t_{(14)} = 1.80, p = 0.046$), but not the steady-state region ($t_{(14)} = 1.61, p = 0.065$). No significant benefit is found for younger listeners in either region ($t_{(16)} = 1.04, p = 0.156$ and $t_{(16)} = 0.16, p = 0.439$ for transition and steady-state region, respectively). The regression slope is not significantly positive or negative for either group ($p > 0.05$ by two-tailed *t*-tests), as seen in the bar plots in the right panels of figure A1C and A1D.

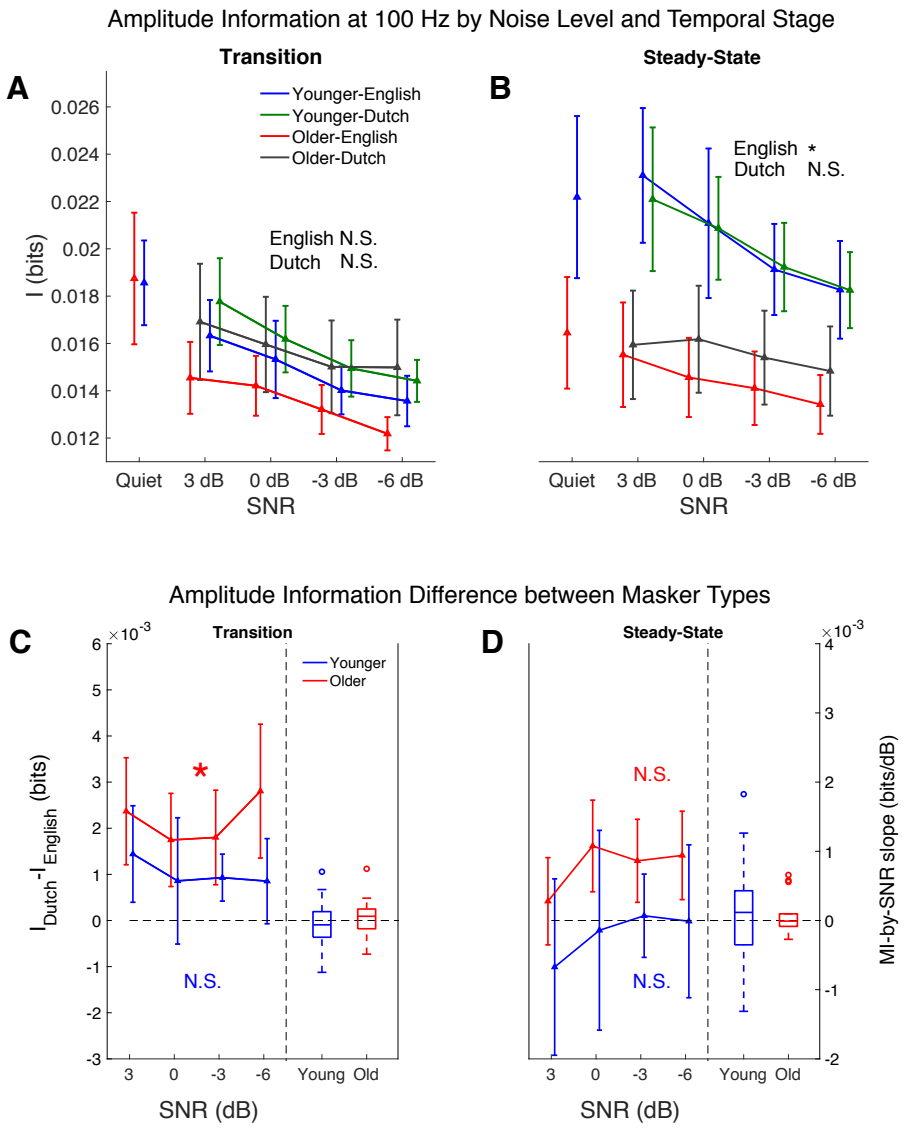

Figure A1. Mutual information of amplitude response by masker type and response region for younger listeners in blue (English) and green (Dutch) and older in red (English) and gray (Dutch). A and B demonstrate the mutual information as a function of SNR in the transition and steady-stage regions, respectively. In the steady-state region, group differences are significant for only the English masker, indicated by asterisks. C and D illustrate the mutual information difference between masker types (denoted $I_{Dutch} - I_{English}$) in the transition region and steady-state region,

respectively. In each plot, the left panel displays information as a function of SNR, and the right panel displays a bar plot showing the slopes of the linear fits. The y-intercepts (corresponding to the fit at 3 dB SNR) are tested against 0 bits. Older listeners show significant benefit from the Dutch masker over English (denoted by asterisk), but only in the transition region. Error bars in all plots indicate SEM. ($* \, p < 0.05$)

Information in phase of FFR without averaging polarities

Similarly, for phase information, a regression line was fitted as a function of SNR to reduce within-subject variance. Using a one-tailed *t*-test on the y-intercept (effective mutual information benefit at 3 dB SNR) of the regression line against zero, the mutual information in phase benefit from the Dutch masker over the English masker is significantly higher for older listeners in the transition region ($t_{(14)} = 1.90, p = 0.039$), but not the steady-state region ($t_{(14)} = 1.45, p = 0.085$). No significant benefit is found for younger listeners in either region ($t_{(16)} = 1.04, p = 0.156$ and $t_{(16)} = 0.25, p = 0.401$ for transition and steady-state region, respectively). The regression slope is not significantly positive or negative for either group ($p > 0.05$ by two-tailed *t*-tests), as seen in the bar plots in the right panels of figure A2C and A2D.

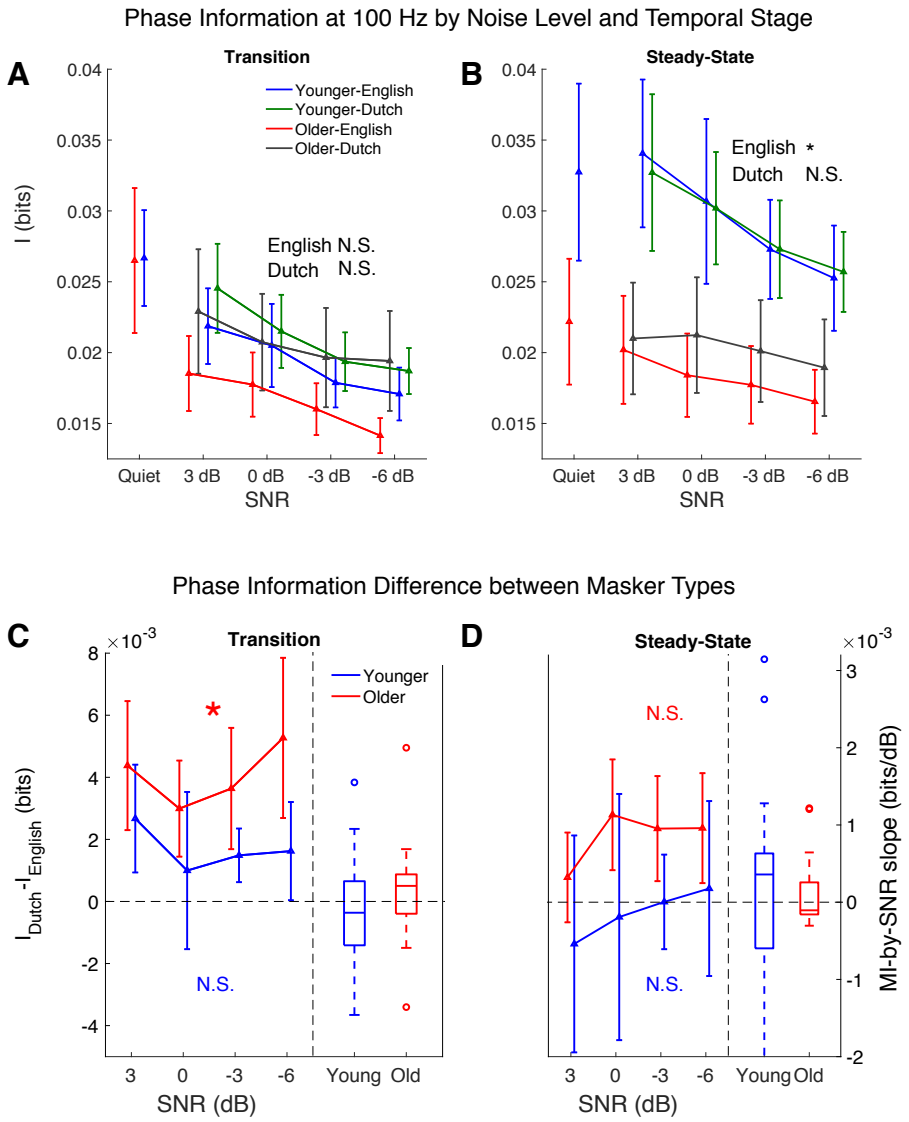

Figure A2. Mutual information of phase response by masker type and response region for younger listeners in blue (English) and green (Dutch) and older in red (English) and gray (Dutch). A and B demonstrate the mutual information as a function of SNR in the transition and steady-stage regions, respectively. In the steady-state region, group differences are significant for only English masker, indicated by asterisks. C and D illustrate the mutual information difference between masker types (denoted $I_{Dutch} - I_{English}$) in the transition region and steady-state region,

respectively. In each plot, the left panel displays information as a function of SNR, and the right panel displays a bar plot showing the slopes of the linear fits. The y-intercepts (corresponding to the fit at 3 dB SNR) are tested against 0 bits. Older listeners show significant benefit from the Dutch masker over English (denoted by asterisk), but only in the transition region. Error bars in all plots indicate SEM. ($* p < 0.05$)

Table A1. Amplitude information: one-tailed *t*-test (younger > older) results applied to the fitted y-intercepts (3 dB values) and slopes from the linear regression analysis of mutual information (for response amplitude) as a function of SNR, for each harmonic. *p*-values are corrected for multiple comparisons by FDR correction. Boldfaced entries indicate the corresponding tests are statistically significant.

Harmonic (Hz)	Quiet (Y>O)		English masker (Y>O)				Dutch masker (Y>O)			
			y-intercept		Slope		y-intercept		slope	
	t(30)	*p*	*t*(30)	*p*	*t*(30)	*p*	*t*(30)	*p*	*t*(30)	*p*
100	0.982	0.167	**1.700**	**0.050**	1.287	0.104	1.238	0.113	1.254	0.110
200	1.544	0.080	**1.918**	**0.039**	**2.583**	**0.011**	1.338	0.113	1.619	0.087
300	1.862	0.054	**2.161**	**0.029**	**2.060**	**0.029**	**2.138**	**0.041**	2.185	0.087
400	**2.441**	**0.021**	**2.380**	**0.024**	**2.699**	**0.011**	1.795	0.062	1.670	0.087
500	**3.466**	**0.002**	**3.640**	**0.003**	**3.612**	**0.002**	**2.247**	**0.041**	1.696	0.087
600	**3.536**	**0.002**	**3.370**	**0.003**	**3.546**	**0.002**	**2.168**	**0.041**	1.281	0.110

Table A2. Phase information: one-tailed *t*-test (younger > older) results applied to the fitted y-intercepts (3 dB values) and slopes from the linear regression analysis of mutual information (for response phase) as a function of SNR, for each harmonic. *p*-values are corrected for multiple comparisons by FDR correction. Boldfaced entries indicate the corresponding tests are statistically significant.

Harmonic (Hz)	Quiet (Y>O)		English masker (Y>O)				Dutch masker (Y>O)			
			y-intercept		slope		y-intercept		slope	
	t(30)	*p*	*t*(30)	*p*	*t*(30)	*p*	*t*(30)	*p*	*t*(30)	*p*
100	1.005	0.161	**1.758**	**0.044**	1.334	0.096	1.313	0.100	1.302	0.101
200	1.514	0.084	**2.047**	**0.030**	**1.962**	**0.035**	1.782	0.051	1.947	0.061
300	1.822	0.059	**2.300**	**0.021**	**2.008**	**0.035**	**2.199**	**0.034**	2.167	0.061
400	**2.400**	**0.023**	**2.537**	**0.017**	**2.512**	**0.018**	**2.088**	**0.034**	2.054	0.061
500	**3.653**	**0.001**	**3.865**	**0.001**	**3.641**	**0.002**	**2.204**	**0.034**	1.556	0.081
600	**3.701**	**0.001**	**3.677**	**0.001**	**3.904**	**0.001**	**2.619**	**0.034**	1.533	0.081

Bibliography

Aiken SJ, **Picton TW**. Envelope following responses to natural vowels. *Audiol Neurootol* 11: 213–232, 2006.

Aiken SJ, **Picton TW**. Envelope and spectral frequency-following responses to vowel sounds. *Hear Res* 245: 35–47, 2008a.

Aiken SJ, **Picton TW**. Envelope and spectral frequency-following responses to vowel sounds. *Hear Res* 245: 35–47, 2008b.

Anderson S, **Parbery-Clark A**, **White-Schwoch T**, **Drehobl S**, **Kraus N**. Effects of hearing loss on the subcortical representation of speech cues. *J Acoust Soc Am* 133: 3030–3038, 2013.

Anderson S, **Parbery-Clark A**, **White-Schwoch T**, **Kraus N**. Aging Affects Neural Precision of Speech Encoding. *J Neurosci* 32: 14156–14164, 2012.

AnonAge-Related Hearing Loss [Online]. *NIDCD* 2015.https://www.nidcd.nih.gov/health/age-related-hearing-loss [7 Nov. 2019].

Auerbach BD, **Rodrigues PV**, **Salvi RJ**. Central Gain Control in Tinnitus and Hyperacusis. *Front Neurol* 5, 2014.

Babadi B, **Obregon-Henao G**, **Lamus C**, **Hämäläinen MS**, **Brown EN**, **Purdon PL**. A Subspace Pursuit-based Iterative Greedy Hierarchical solution to the neuromagnetic inverse problem. *NeuroImage* 87: 427–443, 2014.

Bajo VM, **Nodal FR**, **Moore DR**, **King AJ**. The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat Neurosci* 13: 253–260, 2010.

Bates D, **Mächler M**, **Bolker B**, **Walker S**. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw* 67: 1–48, 2015.

Batra R, **Kuwada S**, **Maher VL**. The frequency-following response to continuous tones in humans. *Hear Res* 21: 167–177, 1986.

Bear MF, **Connors BW**, **Paradiso MA**. Neuroscience exploring the brain [Online]. 3rd ed. Baltimore, MD Lippincott Williams & Wilkins.https://trove.nla.gov.au/work/20483561 [28 Sep. 2018].

Bedrosian E. A product theorem for Hilbert transforms. *Proc IEEE* 51: 868–869, 1963.

Benjamini Y, **Hochberg Y**. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol* 57: 289–300, 1995.

Bidelman GM, **Villafuerte JW**, **Moreno S**, **Alain C**. Age-related changes in the subcortical–cortical encoding and categorical perception of speech. *Neurobiol Aging* 35: 2526–2540, 2014.

Brenner N, **Bialek W**, **de Ruyter van Steveninck R**. Adaptive Rescaling Maximizes Information Transmission. *Neuron* 26: 695–702, 2000.

Brodbeck C, **Presacco A**, **Anderson S**, **Simon JZ**. Increased speech representation in older adults originates from early response in higher order auditory cortex. 2018a. doi:10.1101/294017.

Brodbeck C, **Presacco A**, **Anderson S**, **Simon JZ**. Over-representation of speech in older adults originates from early response in higher order auditory cortex. *Acta Acust United Acust J Eur Acoust Assoc EEIG* 104: 774–777, 2018b.

Brodbeck C, **Presacco A**, **Simon JZ**. Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. *NeuroImage* 172: 162–174, 2018c.

Brugge JF, **Nourski KV**, **Oya H**, **Reale RA**, **Kawasaki H**, **Steinschneider M**, **Howard MA**. Coding of Repetitive Transients by Auditory Cortex on Heschl's Gyrus. *J Neurophysiol* 102: 2358–2374, 2009.

Brungart DS, **Simpson BD**, **Ericson MA**, **Scott KR**. Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J Acoust Soc Am* 110: 2527–2538, 2001.

Burkard RF, **Sims D**. A Comparison of the Effects of Broadband Masking Noise on the Auditory Brainstem Response in Young and Older Adults. *Am J Audiol* 11: 13, 2002.

Burke DM, **Shafto MA**. Language and Aging. 111, 2008.

Carabellese C, **Appollonio I**, **Rozzini R**, **Bianchetti A**, **Frisoni GB**, **Frattola L**, **Trabucchi M**. Sensory Impairment and Quality of Life in a Community Elderly Population. *J Am Geriatr Soc* 41: 401–407, 1993.

Caspary DM, **Hughes LF**, **Schatteman TA**, **Turner JG**. Age-related changes in the response properties of cartwheel cells in rat dorsal cochlear nucleus. *Hear Res* 216–217: 207–215, 2006.

Caspary DM, **Ling L**, **Turner JG**, **Hughes LF**. Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *J Exp Biol* 211: 1781–1791, 2008.

Caspary DM, **Milbrandt JC**, **Helfert RH**. Central auditory aging: GABA changes in the inferior colliculus. *Exp Gerontol* 30: 349–360, 1995.

Caspary DM, **Tracy A. Schatteman**, **Larry F. Hughes**. Age-Related Changes in the Inhibitory Response Properties of Dorsal Cochlear Nucleus Output Neurons: Role of Inhibitory Inputs. *J Neurosci* 25: 10952–10959, 2005.

Cervenka Mackenzie C., **Nagle Stephanie**, **Boatman-Reich Dana**. Cortical High-Gamma Responses in Auditory Processing. *Am J Audiol* 20: 171–180, 2011.

Cervenka MC, **Nagle S**, **Boatman-Reich D**. Cortical High-Gamma Responses in Auditory Processing. *Am J Audiol* 20, 2011.

Chambers AR, **Resnik J**, **Yuan Y**, **Whitton JP**, **Edge AS**, **Liberman MC**, **Polley DB**. Central Gain Restores Auditory Processing following Near-Complete Cochlear Denervation. *Neuron* 89: 867–879, 2016.

Chandrasekaran B, **Kraus N**. The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47: 236–246, 2010.

Chang EF, **Edwards E**, **Nagarajan SS**, **Fogelson N**, **Dalal SS**, **Canolty RT**, **Kirsch HE**, **Barbaro NM**, **Knight RT**. Cortical Spatio-temporal Dynamics Underlying Phonological Target Detection in Humans. *J Cogn Neurosci* 23: 1437–1446, 2010.

Chase SM, **Young ED**. Limited Segregation of Different Types of Sound Localization Information among Classes of Units in the Inferior Colliculus. *J Neurosci* 25: 7575–7585, 2005.

de Cheveigné A, **Roux JL**, **Simon JZ**. MEG Signal Denoising Based on Time-Shift PCA. In: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*. 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07. 2007, p. I-317-I–320.

de Cheveigné A, **Simon JZ**. Sensor noise suppression. *J Neurosci Methods* 168: 195–202, 2008a.

de Cheveigné A, **Simon JZ**. Denoising based on spatial filtering. *J Neurosci Methods* 171: 331–339, 2008b.

Chi T, **Ru P**, **Shamma SA**. Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am* 118: 887–906, 2005.

Clinard CG, **Tremblay KL**. Aging Degrades the Neural Encoding of Simple and Complex Sounds in the Human Brainstem. *J Am Acad Audiol* 24: 590–599, 2013.

Coffey EBJ, **Herholz SC**, **Chepesiuk AMP**, **Baillet S**, **Zatorre RJ**. Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat Commun* 7: 11070, 2016a.

Coffey EBJ, **Herholz SC**, **Chepesiuk AMP**, **Baillet S**, **Zatorre RJ**. Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat Commun* 7: 1–11, 2016b.

Coffey EBJ, **Musacchia G**, **Zatorre RJ**. Cortical Correlates of the Auditory Frequency-Following and Onset Responses: EEG and fMRI Evidence. *J Neurosci* 37: 830–838, 2017.

Cogan GB, **Poeppel D**. A mutual information analysis of neural coding of speech by low-frequency MEG phase information. *J Neurophysiol* 106: 554–563, 2011.

Cohen D. Magnetoencephalography: Detection of the Brain's Electrical Activity with a Superconducting Magnetometer. *Science* 175: 664–666, 1972.

Collier R, **Hart J 't**. The Role of Intonation in Speech Perception. In: *Structure and Process in Speech Perception*, edited by Cohen A, Nooteboom SG. Berlin, Heidelberg: Springer, 1975, p. 107–123.

Cover TM, **Thomas JA**. *Elements of information theory*. New York: Wiley, 1991.

Crone NE, **Boatman D**, **Gordon B**, **Hao L**. Induced electrocorticographic gamma activity during auditory perception. *Clin Neurophysiol* 112: 565–582, 2001.

David SV, **Mesgarani N**, **Shamma SA**. Estimating sparse spectro-temporal receptive fields with natural stimuli. *Netw Comput Neural Syst* 18: 191–212, 2007.

Delorme A, **Makeig S**. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134: 9–21, 2004.

Diamond A. Executive Functions. *Annu Rev Psychol* 64: 135–168, 2013.

Dimitrijevic A, **John M**, **Picton T**. Auditory Steady-State Responses and Word Recognition Scores in Normal-Hearing and Hearing-Impaired Adults. *Ear Hear* 25: 68–84, 2004.

Ding N, **Chatterjee M**, **Simon JZ**. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage* 88: 41–46, 2014.

Ding N, **Simon JZ**. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J Neurophysiol* 107: 78–89, 2012a.

Ding N, **Simon JZ**. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc Natl Acad Sci* 109: 11854–11859, 2012b.

Ding N, **Simon JZ**. Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical Representation of Speech. *J Neurosci* 33: 5728–5735, 2013.

Edwards E, **Soltani M**, **Deouell LY**, **Berger MS**, **Knight RT**. High Gamma Activity in Response to Deviant Auditory Stimuli Recorded Directly From Human Cortex. *J Neurophysiol* 94: 4269–4280, 2005.

Escabí MA, **Miller LM**, **Read HL**, **Schreiner CE**. Naturalistic Auditory Contrast Improves Spectrotemporal Coding in the Cat Inferior Colliculus. *J Neurosci* 23: 11489–11504, 2003.

Federmeier KD, **Kutas M**. Picture the difference: electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia* 40: 730–747, 2002.

Fischl B. FreeSurfer. *NeuroImage* 62: 774–781, 2012.

Fishman YI, **Arezzo JC**, **Steinschneider M**. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J Acoust Soc Am* 116: 1656–1670, 2004.

Fitzgibbons PJ, **Gordon-Salant S**. Aging and temporal discrimination in auditory sequences. *J Acoust Soc Am* 109: 2955–2963, 2001.

Fitzgibbons PJ, **Gordon-Salantt S**. Auditory Temporal Processing in Elderly Listeners. 7, 1996.

Forte AE, **Etard O**, **Reichenbach T**. The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. 12, 2017.

Friedman J, **Hastie T**, **Rosset S**, **Tibshirani R**, **Zhu J**. Discussion of Boosting Papers. 7, date unknown.

Friedman J, **Hastie T**, **Tibshirani R**. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *Ann Stat* 28: 337–407, 2000.

Frisina DR, **Frisina RD**. Speech recognition in noise and presbycusis: relations to possible neural mechanisms. *Hear Res* 106: 95–104, 1997.

Furukawa S, **Middlebrooks JC**. Cortical Representation of Auditory Space: Information-Bearing Features of Spike Patterns. *J Neurophysiol* 87: 1749–1762, 2002.

Galbraith G, **Bhuta S**, **Choate A**, **Kitahara J**, **Mullen T**. Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport* 9: 1889–1893, 1998.

Galbraith GC. Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. *Electroencephalogr Clin Neurophysiol Potentials Sect* 92: 321–330, 1994.

Galbraith GC, **Arbagey PW**, **Branski R**, **Comerci N**, **Rector PM**. Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 6: 2363–2367, 1995.

Galbraith GC, **Jhaveri SP**, **Kuo J**. Speech-evoked brainstem frequency-following responses during verbal transformations due to word repetition. *Electroencephalogr Clin Neurophysiol* 102: 46–53, 1997.

Galbraith GC, **Olfman DM**, **Huffman TM**. Selective attention affects human brain stem frequency-following response. *Neuroreport* 14: 735–738, 2003.

Galbraith GC, **Threadgill MR**, **Hemsley J**, **Salour K**, **Songdej N**, **Ton J**, **Cheung L**. Putative measure of peripheral and brainstem frequency-following in humans. *Neurosci Lett* 292: 123–127, 2000.

Gandour J, **Dzemidzic M**, **Wong D**, **Lowe M**, **Tong Y**, **Hsieh L**, **Satthamnuwong N**, **Lurito J**. Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain Lang* 84: 318–336, 2003.

Gardi J, **Merzenich M**, **McKean C**. Origins of the scalp recorded frequency-following response in the cat. *Audiol Off Organ Int Soc Audiol* 18: 358–381, 1979.

Gordon-Salant S, **Yeni-Komshian GH**, **Fitzgibbons PJ**, **Barrett J**. Age-related differences in identification and discrimination of temporal cues in speech segments. *J Acoust Soc Am* 119: 2455–2466, 2006.

Gramfort A, **Luessi M**, **Larson E**, **Engemann DA**, **Strohmeier D**, **Brodbeck C**, **Goj R**, **Jas M**, **Brooks T**, **Parkkonen L**, **Hämäläinen M**. MEG and EEG data analysis with MNE-Python. *Front Neurosci* 7, 2013.

Gramfort A, **Luessi M**, **Larson E**, **Engemann DA**, **Strohmeier D**, **Brodbeck C**, **Parkkonen L**, **Hämäläinen MS**. MNE software for processing MEG and EEG data. *NeuroImage* 86: 446–460, 2014.

Greenberg S. WPP, No. 52: Temporal Neural Coding of Pitch and Vowel Quality [Online]. 1980https://escholarship.org/uc/item/0vp3920c [25 Sep. 2019].

Greenberg S, **Hollenback J**, **Ellis DPW**. INSIGHTS INTO SPOKEN LANGUAGE GLEANED FROM PHONETIC TRANSCRIPTION OF THE SWITCHBOARD CORPUS. 1996.

Griffiths TD, **Warren JD**. What is an auditory object? *Nat Rev Neurosci* 5: 887–892, 2004.

Gutschalk A, **Micheyl C**, **Oxenham AJ**. Neural Correlates of Auditory Perceptual Awareness under Informational Masking. *PLOS Biol* 6: e138, 2008.

Hämäläinen M, **Hari R**, **Ilmoniemi RJ**, **Knuutila J**, **Lounasmaa OV**. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev Mod Phys* 65: 413–497, 1993.

Hämäläinen MS, **Ilmoniemi RJ**. Interpreting magnetic fields of the brain: minimum norm estimates. *Med Biol Eng Comput* 32: 35–42, 1994.

Hari R, **Salmelin R**. Magnetoencephalography: From SQUIDs to neuroscience: Neuroimage 20th Anniversary Special Edition. *NeuroImage* 61: 386–396, 2012.

He N, **Mills JH**, **Ahlstrom JB**, **Dubno JR**. Age-related differences in the temporal modulation transfer function with pure-tone carriers. *J Acoust Soc Am* 124: 3841–3849, 2008.

Helfer KS, **Freyman RL**. Aging and Speech-on-Speech Masking. *Ear Hear* 29: 87–98, 2008.

Herbst KG, **Humphrey C**. Hearing impairment and mental state in the elderly living at home. *Br Med J* 281: 903–905, 1980.

Herrmann CS, **Knight RT**. Mechanisms of human attention: event-related potentials and oscillations. *Neurosci Biobehav Rev* 25: 465–476, 2001.

Herrmann CS, **Munk MHJ**, **Engel AK**. Cognitive functions of gamma-band activity: memory match and utilization. *Trends Cogn Sci* 8: 347–355, 2004.

Hertrich I, **Dietrich S**, **Trouvain J**, **Moos A**, **Ackermann H**. Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. *Psychophysiology* 49: 322–334, 2012.

Hillebrand A, **Barnes GR**. A Quantitative Assessment of the Sensitivity of Whole-Head MEG to Activity in the Adult Human Cortex. *NeuroImage* 16: 638–650, 2002.

Holm S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand J Stat* 6: 65–70, 1979.

Hoormann J, **Falkenstein M**, **Hohnsbein J**. Effects of spatial attention on the brain stem frequency-following potential. *Neuroreport* 15: 1539–1542, 2004.

Howard MA, **Volkov IO**, **Mirsky R**, **Garell PC**, **Noh MD**, **Granner M**, **Damasio H**, **Steinschneider M**, **Reale RA**, **Hind JE**, **Brugge JF**. Auditory cortex on the human posterior superior temporal gyrus. *J Comp Neurol* 416: 79–92, 2000.

Hsu A, **Woolley SMN**, **Fremouw TE**, **Theunissen FE**. Modulation Power and Phase Spectrum of Natural Sounds Enhance Neural Encoding Performed by Single Auditory Neurons. *J Neurosci* 24: 9201–9211, 2004.

Hughes LF, **Turner JG**, **Parrish JL**, **Caspary DM**. Processing of broadband stimuli across A1 layers in young and aged rats. *Hear Res* 264: 79–85, 2010.

Juarez-Salinas DL, **Engle JR**, **Navarro XO**, **Recanzone GH**. Hierarchical and Serial Processing in the Spatial Auditory Cortical Pathway Is Degraded by Natural Aging. *J Neurosci* 30: 14795–14804, 2010.

Kaiser J, **Ripper B**, **Birbaumer N**, **Lutzenberger W**. Dynamics of gamma-band activity in human magnetoencephalogram during auditory pattern working memory. *NeuroImage* 20: 816–827, 2003.

Kerlin JR, **Shahin AJ**, **Miller LM**. Attentional Gain Control of Ongoing Cortical Speech Representations in a "Cocktail Party." *J Neurosci* 30: 620–628, 2010.

Killion MC, **Niquette PA**, **Gudmundsen GI**, **Revit LJ**, **Banerjee S**. 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, 2004.

King AJ, **Bajo VM**. Cortical modulation of auditory processing in the midbrain. *Front Neural Circuits* 6, 2013.

Klatt DH. Software for a cascade/parallel formant synthesizer. *J Acoust Soc Am* 67: 971–995, 1980.

Krishnan A, **Xu Y**, **Gandour JT**, **Cariani PA**. Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* 189: 1–12, 2004.

Lefebvre CD, **Marchand Y**, **Eskes GA**, **Connolly JF**. Assessment of working memory abilities using an event-related brain potential (ERP)-compatible digit span backward task. *Clin Neurophysiol* 116: 1665–1680, 2005.

Lehmann A, **Schönwiesner M**. Selective Attention Modulates Human Auditory Brainstem Responses: Relative Contributions of Frequency and Spatial Cues. *PLOS ONE* 9: e85442, 2014.

Levänen S. Neuromagnetic Studies of Human Auditory Cortex Function and Reorganization. *Scand Audiol* 27: 1–6, 1998.

Lister JJ, **Maxfield ND**, **Pitt GJ**, **Gonzalez VB**. Auditory evoked response to gaps in noise: Older adults. *Int J Audiol* 50: 211–225, 2011.

Lopes da Silva F. EEG and MEG: Relevance to Neuroscience. *Neuron* 80: 1112–1128, 2013.

Luo H, **Poeppel D**. Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron* 54: 1001–1010, 2007.

Luo H, **Tian X**, **Song K**, **Zhou K**, **Poeppel D**. Neural Response Phase Tracks How Listeners Learn New Acoustic Representations. *Curr Biol* 23: 968–974, 2013.

Maddox RK, **Lee AKC**. Auditory Brainstem Responses to Continuous Natural Speech in Human Listeners. *eneuro* 5: ENEURO.0441-17.2018, 2018.

Middlebrooks JC, **Clock AE**, **Xu L**, **Green DM**. A Panoramic Code for Sound Location by Cortical Neurons. *Science* 264: 842–844, 1994.

Miller GA, **Nicely PE**. An Analysis of Perceptual Confusions Among Some English Consonants. *J Acoust Soc Am* 27: 338–352, 1955.

Møller AR, **Jannetta PJ**, **Sekhar LN**. Contributions from the auditory nerve to the brain-stem auditory evoked potentials (BAEPs): results of intracranial recording in man. *Electroencephalogr Clin Neurophysiol Potentials Sect* 71: 198–211, 1988.

Mosher JC, **Leahy RM**, **Lewis PS**. EEG and MEG: forward solutions for inverse methods. *IEEE Trans Biomed Eng* 46: 245–259, 1999.

Neill WT, **Valdes LA**, **Terry KM**. 7 - Selective attention and the inhibitory control of cognition. In: *Interference and Inhibition in Cognition*, edited by Dempster FN, Brainerd CJ, Brainerd CJ. Academic Press, p. 207–261.

Nelken I, **Chechik G**. Information theory in auditory research. *Hear Res* 229: 94–105, 2007.

Neumann J von. *The Computer and the Brain*. New Haven, CT, USA: Yale University Press, 1958.

Okamoto H, **Stracke H**, **Bermudez P**, **Pantev C**. Sound processing hierarchy within human auditory cortex. *J Cogn Neurosci* 23: 1855–1863, 2011.

Parthasarathy A, **Bartlett EL**. Age-related auditory deficits in temporal processing in F-344 rats. *Neuroscience* 192: 619–630, 2011.

Peelle JE, **Troiani V**, **Wingfield A**, **Grossman M**. Neural Processing during Older Adults' Comprehension of Spoken Sentences: Age Differences in Resource Allocation and Connectivity. *Cereb Cortex* 20: 773–782, 2010.

Peña EA, **Slate EH**. Global Validation of Linear Model Assumptions. *J Am Stat Assoc* 101: 341, 2006.

Pichora-Fuller MK. Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *Int J Audiol* 47: S72–S82, 2008.

Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time.' *Speech Commun* 41: 245–255, 2003.

Presacco A, **Jenkins K**, **Lieberman R**, **Anderson S**. Effects of Aging on the Encoding of Dynamic and Static Components of Speech: *Ear Hear* 36: e352–e363, 2015.

Presacco A, **Simon JZ**, **Anderson S**. Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *J Neurophysiol* 116: 2346–2355, 2016a.

Presacco A, **Simon JZ**, **Anderson S**. Effect of informational content of noise on speech representation in the aging midbrain and cortex. *J Neurophysiol* 116: 2356–2367, 2016b.

Presacco A, **Simon JZ**, **Anderson S**. Speech-in-noise representation in the aging midbrain and cortex: Effects of hearing loss. *PLOS ONE* 14: e0213899, 2019.

R. Core Team. R: A Language and Environment for Statistical Computing [Online]. R Foundation for Statistical Computing.https://www.R-project.org/.

Rabinovich MI, **Afraimovich VS**, **Bick C**, **Varona P**. Information flow dynamics in the brain. *Phys Life Rev* 9: 51–73, 2012.

Ray S, **Niebur E**, **Hsiao SS**, **Sinai A**, **Crone NE**. High-frequency gamma activity (80–150Hz) is increased in human cortex during selective attention. *Clin Neurophysiol* 119: 116–133, 2008.

Rieke F, **Bodnar DA**, **Bialek W**. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc Biol Sci* 262: 259–265, 1995.

Rosen S, **Carlyon RP**, **Darwin CJ**, **Russell IJ**. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci* 336: 367–373, 1992.

Russo N, **Nicol T**, **Musacchia G**, **Kraus N**. Brainstem responses to speech syllables. *Clin Neurophysiol Off J Int Fed Clin Neurophysiol* 115: 2021–2030, 2004.

Ryhänen T, **Seppä H**, **Ilmoniemi R**, **Knuutila J**. SQUID magnetometers for low-frequency applications. *J Low Temp Phys* 76: 287–386, 1989.

Sahani M, **Linden JF**. How Linear are Auditory Cortical Responses? *MIT Press* 15: 8, 2003.

Särelä J, **Valpola H**. Denoising Source Separation. 40, 2005.

Sarvas J. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys Med Biol* 32: 11–22, 1987.

Schatteman TA, **Hughes LF**, **Caspary DM**. Aged-Related Loss of Temporal Processing: Altered Responses to Amplitude Modulated Tones in Rat Dorsal Cochlear Nucleus. *Neuroscience* 154: 329–337, 2008.

Schmiedt RA, **Mills JH**, **Boettcher FA**. Age-related loss of activity of auditory-nerve fibers. *J Neurophysiol* 76: 2799–2803, 1996.

Schneider BA, **Hamstra SJ**. Gap detection thresholds as a function of tonal duration for younger and older listeners. *J Acoust Soc Am* 106: 371–380, 1999.

Schneider BA, **Pichora-Fuller MK**, **Kowalchuk D**, **Lamb M**. Gap detection and the precedence effect in young and old adults. *J Acoust Soc Am* 95: 980–991, 1994.

Schnupp J, **Nelken I**, **King A**. *Auditory neuroscience: Making sense of sound*. Cambridge, MA, US: MIT Press, 2011.

Sergeyenko Y, **Lall K**, **Liberman MC**, **Kujawa SG**. Age-Related Cochlear Synaptopathy: An Early-Onset Contributor to Auditory Functional Decline. *J Neurosci* 33: 13686–13694, 2013.

Shahin AJ, **Roberts LE**, **Miller LM**, **McDonald KL**, **Alain C**. Sensitivity of EEG and MEG to the N1 and P2 Auditory Evoked Responses Modulated by Spectral Complexity of Sounds. *Brain Topogr* 20: 55–61, 2007.

Shamma S. On the role of space and time in auditory processing. *Trends Cogn Sci* 5: 340–348, 2001.

Shannon CE. A Mathematical Theory of Communication. 55, 1948.

Siedenberg R, **Goodin DS**, **Aminoff MJ**, **Rowley HA**, **Roberts TP**. Comparison of late components in simultaneously recorded event-related electrical potentials and event-related magnetic fields. *Electroencephalogr Clin Neurophysiol* 99: 191–197, 1996.

Silver AH, **Zimmerman JE**. Quantum Transitions and Loss in Multiply Connected Superconductors. *Phys Rev Lett* 15: 888, 1965.

Skoe E, **Kraus N**. Auditory Brain Stem Response to Complex Sounds: A Tutorial: *Ear Hear* 31: 302–324, 2010.

Slee SJ, **Higgs MH**, **Fairhall AL**, **Spain WJ**. Two-dimensional time coding in the auditory brainstem. *J Neurosci Off J Soc Neurosci* 25: 9978–9988, 2005.

Smith JC, **Marsh JT**, **Brown WS**. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr Clin Neurophysiol* 39: 465–472, 1975.

Smith SM, **Nichols TE**. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage* 44: 83–98, 2009.

Stecker GC, **Harrington IA**, **Middlebrooks JC**. Location Coding by Opponent Neural Populations in the Auditory Cortex. *PLOS Biol* 3: e78, 2005.

Steinschneider M, **Nourski KV**, **Kawasaki H**, **Oya H**, **Brugge JF**, **Howard MA**. Intracranial Study of Speech-Elicited Activity on the Human Posterolateral Superior Temporal Gyrus. *Cereb Cortex* 21: 2332–2347, 2011.

Stillman RD, **Crow G**, **Moushegian G**. Components of the frequency-following potential in man. *Electroencephalogr Clin Neurophysiol* 44: 438–446, 1978.

Takahashi GA, **Bacon SP**. Modulation detection, modulation masking, and speech understanding in noise in the elderly. *J Speech Hear Res* 35: 1410–1421, 1992.

Takesian AE, **Kotak VC**, **Sanes DH**. Age-dependent effect of hearing loss on cortical inhibitory synapse function. *J Neurophysiol* 107: 937–947, 2012.

Towle VL, **Yoon H-A**, **Castelle M**, **Edgar JC**, **Biassou NM**, **Frim DM**, **Spire J-P**, **Kohrman MH**. ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. *Brain* 131: 2013–2027, 2008.

Trautner P, **Rosburg T**, **Dietl T**, **Fell J**, **Korzyukov OA**, **Kurthen M**, **Schaller C**, **Elger CE**, **Boutros NN**. Sensory gating of auditory evoked and induced gamma band activity in intracranial recordings. *NeuroImage* 32: 790–798, 2006.

Tremblay K, **Kraus N**, **McGee T**, **Ponton C**, **Otis B**. Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear Hear* 22: 79–90, 2001.

Tremblay K, **Ross B**, **Inoue K**, **McClannahan K**, **Collet G**. Is the auditory evoked P2 response a biomarker of learning? *Front Syst Neurosci* 8, 2014.

Tun PA, **O'Kane G**, **Wingfield A**. Distraction by competing speech in young and older adult listeners. *Psychol Aging* 17: 453–467, 2002.

Varghese L, **Bharadwaj HM**, **Shinn-Cunningham BG**. Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Res* 1626: 146–164, 2015.

de Villers-Sidani E, **Alzghoul L**, **Zhou X**, **Simpson KL**, **Lin RCS**, **Merzenich MM**. Recovery of functional and structural age-related changes in the rat primary auditory cortex with operant training. *Proc Natl Acad Sci* 107: 13900–13905, 2010.

Walton JP, **Frisina RD**, **O'Neill WE**. Age-Related Alteration in Processing of Temporal Sound Features in the Auditory Midbrain of the CBA Mouse. *J Neurosci* 18: 2764–2776, 1998.

Wang H, **Turner JG**, **Ling L**, **Parrish JL**, **Hughes LF**, **Caspary DM**. Age-related changes in glycine receptor subunit composition and binding in dorsal cochlear nucleus. *Neuroscience* 160: 227–239, 2009.

Wang Y, **Zhang J**, **Zou J**, **Luo H**, **Ding N**. Prior Knowledge Guides Speech Segregation in Human Auditory Cortex. *Cereb Cortex* 29: 1561–1571, 2019.

Ward CM, **Rogers CS**, **Van Engen KJ**, **Peelle JE**. Effects of Age, Acoustic Challenge, and Verbal Working Memory on Recall of Narrative Speech. *Exp Aging Res* 42: 97–111, 2016.

Weintraub S, **Dikmen SS**, **Heaton RK**, **Tulsky DS**, **Zelazo PD**, **Bauer PJ**, **Carlozzi NE**, **Slotkin J**, **Blitz D**, **Wallner-Allen K**, **Fox NA**, **Beaumont JL**, **Mungas D**, **Nowinski CJ**, **Richler J**, **Deocampo JA**, **Anderson JE**, **Manly JJ**, **Borosh B**, **Havlik R**, **Conway K**, **Edwards E**, **Freund L**, **King JW**, **Moy C**, **Witt E**, **Gershon RC**. Cognition assessment using the NIH Toolbox. *Neurology* 80: S54-64, 2013.

Williamson SJ, **Kaufman L**. Biomagnetism. *J Magn Magn Mater* 22: 129–201, 1981.

Worden FG, **Marsh JT**. Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalogr Clin Neurophysiol* 25: 42–52, 1968.

Yang X, **Wang K**, **Shamma SA**. Auditory representations of acoustic signals. *IEEE Trans Inf Theory* 38: 824–839, 1992.

Zan P, **Presacco A**, **Anderson S**, **Simon JZ**. Mutual Information Analysis of Neural Representations of Speech in Noise in the Aging Midbrain. *J Neurophysiol* , 2019. doi:10.1152/jn.00270.2019.

Zhu J, **Garcia E**. The Wechsler Abbreviated Scale of Intelligence (WASI). *New York: Psychological Corp* , 1999.

Zhu L, **Bharadwaj H**, **Xia J**, **Shinn-Cunningham B**. A comparison of spectral magnitude and phase-locking value analyses of the frequency-following response to complex tones. *J Acoust Soc Am* 134: 384–395, 2013.