Greater BOLD Variability in Older Compared with Younger Adults during Audiovisual Speech Perception

Sarah H. Baum[1,2]*, Michael S. Beauchamp[1]

1 Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, Houston, Texas, United States of America, 2 Vanderbilt Brain Institute, Vanderbilt University, Nashville, Tennessee, United States of America

Abstract

Older adults exhibit decreased performance and increased trial-to-trial variability on a range of cognitive tasks, including speech perception. We used blood oxygen level dependent functional magnetic resonance imaging (BOLD fMRI) to search for neural correlates of these behavioral phenomena. We compared brain responses to simple speech stimuli (audiovisual syllables) in 24 healthy older adults (53 to 70 years old) and 14 younger adults (23 to 39 years old) using two independent analysis strategies: region-of-interest (ROI) and voxel-wise whole-brain analysis. While mean response amplitudes were moderately greater in younger adults, older adults had much greater within-subject variability. The greatly increased variability in older adults was observed for both individual voxels in the whole-brain analysis and for ROIs in the left superior temporal sulcus, the left auditory cortex, and the left visual cortex. Increased variability in older adults could not be attributed to differences in head movements between the groups. Increased neural variability may be related to the performance declines and increased behavioral variability that occur with aging.

Introduction

The ability of older adults to understand both auditory-only and audiovisual speech declines with age [1–4]. This decline extends to other important cognitive functions, such as memory, visuospatial abilities, and speed of information processing [5–8]. Interestingly, performance declines with age are not uniform across multiple trials of the same task. Older adults exhibit much greater variability in performance: on some trials, older adults perform as well as younger adults, but on other trials, older adults perform much worse [9–11]. This type of performance decline, referred to as increased intrasubject variability, may be a particularly sensitive measure of age-related changes [12].

We hypothesized that the increased intrasubject variability with age observed in behavioral paradigms should have a neural counterpart. For example, *in vivo* electrophysiology investigations of the visual cortex of experimental animals show increased variability and decreased stimulus specificity with age during viewing of simple visual stimuli [13–15]. Increased neural variability in the auditory brain stem response of healthy older adults has been related to deficits in speech in noise perception [16]. Variability on a cognitive motor task has been tied to individual differences in activation and functional connectivity of the dorsal pre-motor cortex [17]. Wide variability has also been observed in the network activation and functional connectivity of posterior default mode and frontal executive networks in a large cohort (1000BRAINS) of healthy older subjects, which may be

related to differences in neuropsychological tests of cognitive and motor function [18]. Conversely, other investigations of variability in healthy aging have pointed to the possibility of *decreased* variability with age [19–21]. This necessitates continued research to investigate to fully understand the role of response variability in healthy aging.

We used rapid event-related blood-oxygen level dependent magnetic resonance imaging (BOLD fMRI) to measure brain responses to passively-presented audiovisual speech syllables. This allowed us to focus on perceptual processing differences between young and old rather than differences in higher cognitive function and performance. We predicted that repeated presentations of identical speech stimuli would evoke fMRI responses that were more variable in older adults than in younger adults. Neural responses to audiovisual speech were examined using two complementary methods. First, we used a region-of-interest (ROI) analysis focused on the three core areas of the multisensory speech perception network: auditory cortex, visual cortex, and the superior temporal sulcus (STS). Second, we used a voxel-wise analysis to search for differences between older and younger subjects outside of our predefined ROIs.

Materials and Methods

2.1. Ethics Statement

All subjects provided written informed consent and were compensated for their time in accordance with an experimental
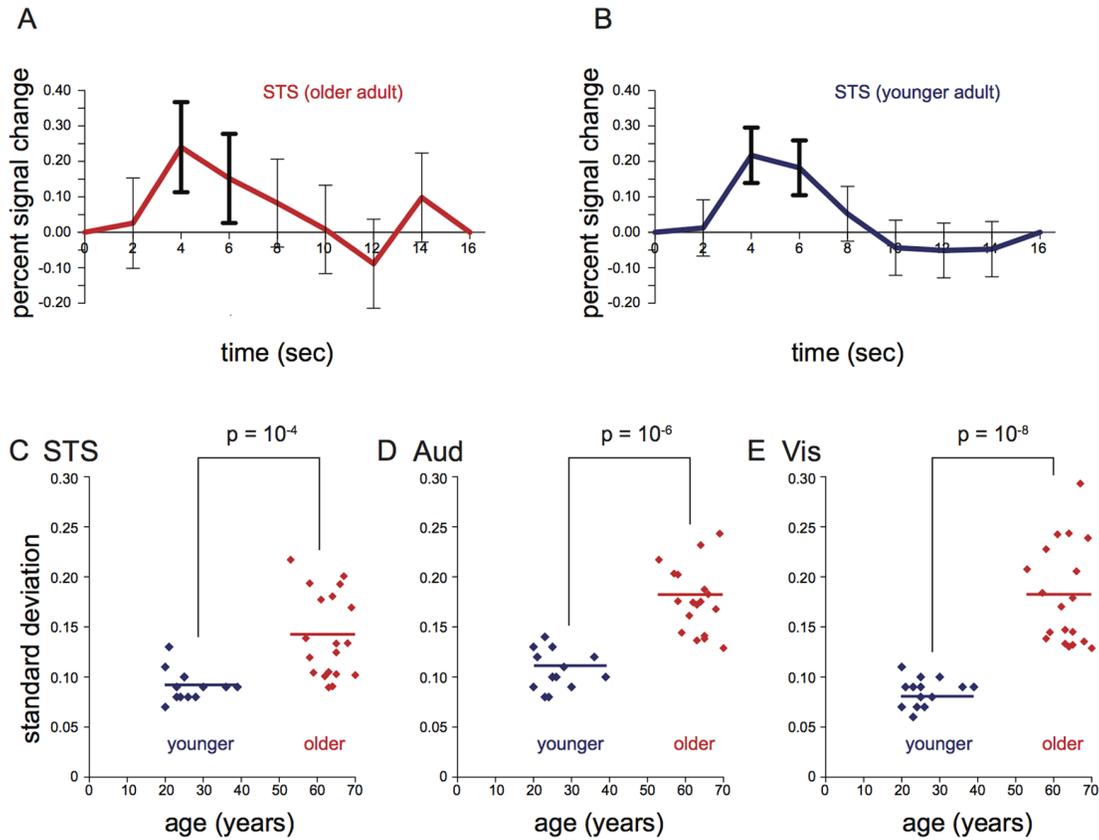
Figure 1. Mean and standard deviation of BOLD responses to audiovisual speech within subjects. A: Hemodynamic response in the left STS of a single older adult (subject JI). Error bars indicate standard deviation of the response within that subject (intrasubject variability) at each time point. The variability at the 4-second and 6-second time points (bold error bars) was used for group analysis. Representative single subject chosen as the subject whose standard deviation was closest to the mean standard deviation for all older subjects. B: Hemodynamic response in the left STS of a single younger adult (subject HU). Representative single subject chosen as the subject whose standard deviation was closest to the mean standard deviation for all younger subjects. C: Scatter plot of age vs. within-subject standard deviation of the STS response. Blue symbols represent younger adults ($n = 14$), red symbols represent older adults ($n = 19$). The lines show the mean of the within-subject standard deviation across each group. The brackets show the results of an unpaired t-test between the within-subject standard deviation in each group. D: Scatter plot of age vs. within-subject standard deviation of the left auditory cortex response. E: Scatter plot of age vs. within-subject standard deviation of the left visual cortex response.
doi:10.1371/journal.pone.0111121.g001
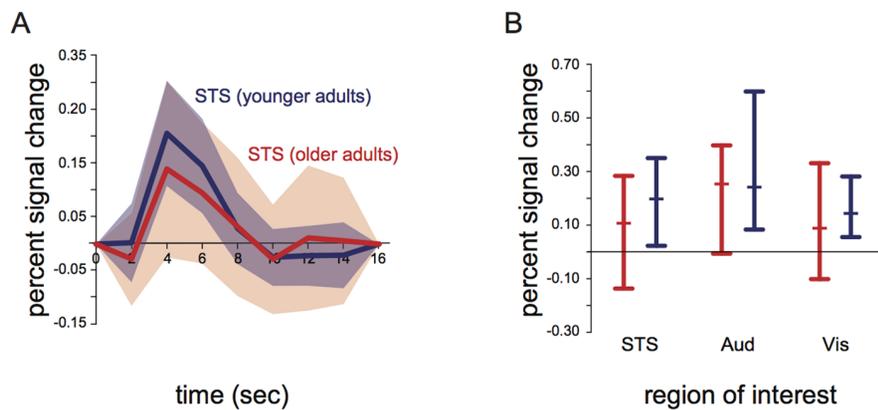
Figure 2. Mean and standard deviation of BOLD responses to audiovisual speech across subjects. A: Average hemodynamic response to audiovisual syllables in the left STS for older adults (red) and younger adults (blue). Shaded region indicates standard deviation of the group response (intersubject variability). B: Response amplitudes in the left STS (STS), left auditory cortex (Aud), and left visual cortex (Vis) across all older adults (red) and younger adults (blue). Error bars show the complete range of data (subjects with maximum and minimum response); middle bar shows median subject.
doi:10.1371/journal.pone.0111121.g002

protocol approved by the Committee for the Protection of Human Subjects of the University of Texas Health Science Center at Houston.

2.2. Subjects and exclusion criteria

The young adult cohort consisted of 14 subjects (20–39 years, 6 female, mean age 26.1 years). 24 older subjects were recruited for the study. Five older adult subjects were excluded (two for poor speech identification scores, three for MRI data quality concerns, see details below), leaving 19 total subjects whose data are reported here (53–70 years, 12 female, mean age 63.0 years).

Vision in the older subjects was assessed using a Snellen eye chart. Each eye was tested separately. The range of acuities was 20/20 to 20/70. Hearing in the older subjects were evaluated using a modified Bekesy threshold test at 500 and 2000 Hz [22]. Our subjects had a range of 6.7 dB–30.9 dB for 500 Hz and 7.3–39.4 dB for 2000 Hz, within the normal range (two standard deviations from the mean) for hearing thresholds based on their age [23]. However, speech abilities decline at a different rate than pure audiometric measures later in life [24]. Therefore, we also tested identification of auditory-only and audiovisual syllables. Most subjects scored near ceiling on auditory-only syllable identification (83%−100%, average performance 93%) and audiovisual syllables. Two subjects scored poorly on auditory-only syllable identification (<70%) and were excluded from the analysis. Cognitive function was assessed using the standardized Mini Mental State Examination (MMSE) [25]. All subjects' scores indicated no decline in cognitive function (scores ranged from 26–30, mean MMSE 28.4, scores 25 out of 30 points or greater indicate normal cognitive function). Two older subjects were excluded for large head movements during fMRI data acquisition (standard deviation of motion regressor >3 mm). One older subject was excluded because an initial analysis of the fMRI data showed response amplitudes more than 3 standard deviations greater than the mean.

2.3. Overview of fMRI experiment and analysis

We used two independent methods for fMRI analysis: region-of-interest (ROI) and voxel-wise whole-brain analysis, which give complementary information about brain activity [26,27]. ROI analysis allows us to examine areas for which we have an *a priori* hypothesis and does not require that data be transformed to a brain template, thus allowing for differences in individual anatomy. Furthermore, it limits Type I errors by limiting the number of statistical tests to a handful of ROIs [28]. However,

ROI analyses are blind to effects outside of the predefined ROIs and to functional specialization within ROIs. Therefore, we also performed a voxel-wise analysis to examine activity across the entire brain.

2.4. Block-design localizer

A block-design localizer was used to generate the regions-of-interest (ROIs). Each block contained 10 two-second trials, one word per trial, followed by 10 seconds of fixation baseline. Each trial contained a single word from a bank of digital video recordings of 105 single-syllable words (*e.g.* "view", "door", "make") spoken by a female native English speaker. Words were selected from the MRC Psycholinguistic Database [29]. Auditory-only words consisted of the auditory component of each video with a white visual fixation crosshairs and visual-only words consisted of only the visual component of the video recording.

In older adults, the localizer scan series contained six blocks (two auditory-only, two visual-only and two audiovisual blocks in random order). Each block contained a target trial (the word "press") of the same type (auditory-only, visual-only, or audiovisual) as the other stimuli in the block; subjects were instructed to pay attention to each stimulus and press a response button only during target trials. In younger adults, ten blocks were presented (five auditory-only and five visual-only in random order) with no target trials.

2.5. fMRI responses to audiovisual speech syllables

For the main experiment, stimuli were presented in two-second trials in a rapid event-related design. Each trial contained a single audiovisual syllable, consisting of McGurk (auditory "ba"+visual "ga", auditory "pa"+visual "ka"), non-McGurk incongruent (auditory "ga"+visual "ba", auditory "ka"+visual "pa"), congruent ("ba", "ga", "da", "pa", "ka" and "ta"), target (audiovisual "press" in older adults, audiovisual "ma" in younger subjects) and fixation trials (fixation crosshairs only). There was only a single exemplar of each audiovisual syllable, meaning that subjects were exposed to identical stimuli repeatedly. This allowed us to isolate the effects of neural variability. Subjects were instructed to respond with a button press only to target trials and to make no response to all other trials. Behavioral data in the scanner was not collected for two younger subjects. Subjects performed near ceiling on this task (18/19 older adults at 100% accuracy; 10/12 younger adults at 100% accuracy) suggesting a high degree of alertness (no significant difference between groups, $t_{29} = 0.6$, $p = 0.57$).

Table 1. Regions of activation in response to audiovisual speech in both older and younger adults.

Label	Area (mm^2)	Peak t-value
R fusiform gyrus, inferior occipital gyrus, middle occipital gyrus (26, −71, −6)	4593	6.2
R fusiform gyrus, inferior occipital gyrus, middle occipital gyrus (55, −21, 2)	3054	4.3
R superior temporal sulcus and gyrus (55, −21, 2)	2361	7.4
L superior temporal gyrus (−46, −33, 16)	2315	5.8
L middle occipital gyrus (−23, −93, 5)	560	4.9
R supramarginal gyrus and subcentral gyrus (49, −6, 44)	219	3.9
Total	13102	

Regions are ranked by area (only clusters greater than 160 mm^2 are reported) and Talairach coordinates following anatomical label in (x, y, z) format are the weighted center of mass of the cluster.
doi:10.1371/journal.pone.0111121.t001

A Average Activation Map
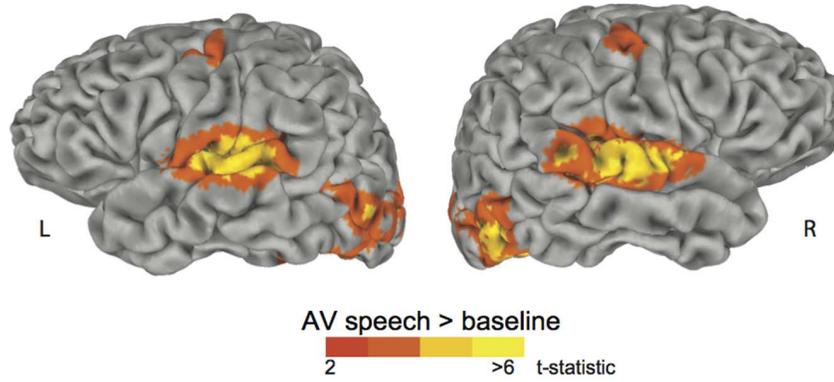
AV speech > baseline

2   >6   t-statistic

B Intrasubject Variability

young > old   old > young

2   >30   t-statistic

C Response Amplitude

young > old   old > young

2   >30   t-statistic

D Intersubject Variability

young > old   old > young

5   >75 χ^2-statistic
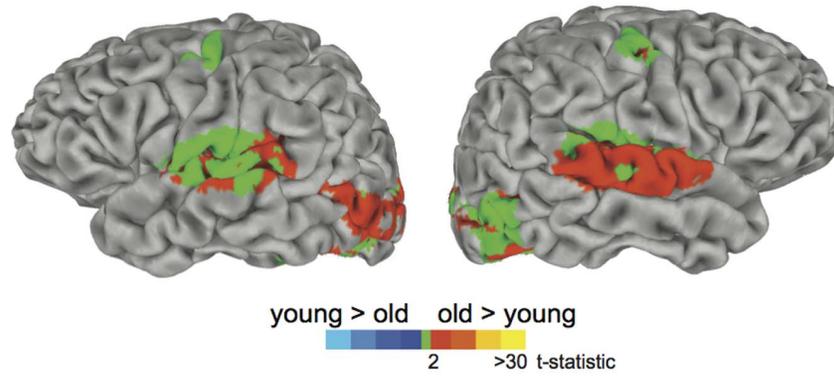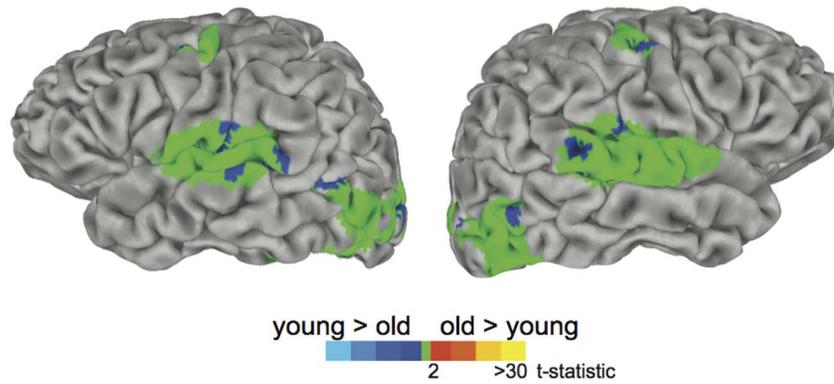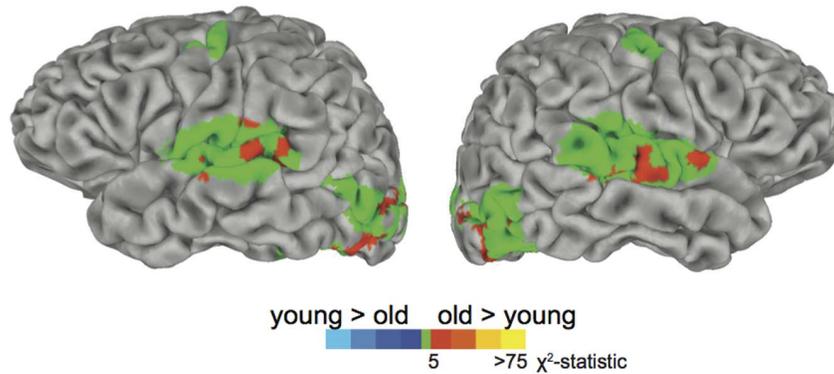
Figure 3. Whole-brain analysis of differences in intrasubject variability, response amplitude, and intersubject variability in older and younger adults. A: Regions that show a significant positive response ($t > 2$ for all audiovisual syllables *vs.* baseline) to audiovisual speech in both older and younger adults. L: left hemisphere, R: right hemisphere. B: Differences in intrasubject variability (variability of the amplitude of the BOLD response to audiovisual speech within each subject) between older and younger subjects, masked by active regions in (A). Orange regions indicate areas with greater intrasubject variability in older adults. C: Differences in response amplitude. Green regions indicate no difference in response amplitude; blue regions indicate areas of greater response amplitude in younger adults. D: Differences in intersubject variability (variability of the amplitude of the BOLD response *across* subjects). Orange regions indicate areas with greater response variability in older adults.
doi:10.1371/journal.pone.0111121.g003

The total length of each video was cropped with digital video editing software (iMovie, Apple Computer) such that each clip started and ended in a neutral, mouth-closed position. Each video stimulus varied in length from 1.7 to 1.8 seconds followed by fixation crosshairs for the remainder of the trial (the crosshairs were always presented in the same screen location as the mouth of the talker visible during other trials in order to minimize eye movements). Prior to the scan, a volume check was conducted for each subject outside the scanner without the presence of scanner noise. Sample videos from the experiment were played and the volume was adjusted so that the volume was "as loud as possible without being uncomfortable or hurting in any way". After each scan series subjects were asked if they could hear the stimuli presented and if any volume adjustments were necessary.

Subjects viewed audiovisual stimuli presented in a rapid event-related design with slight variations in the number of trials, as follows: older subjects, $n = 6$:150 audiovisual syllables, 40 target trials; older subjects, $n = 13$:160 audiovisual syllables, 50 target trials; younger subjects, $n = 5$:220 audiovisual syllables, 70 target trials; younger subjects, $n = 9$:200 audiovisual syllables, 80 target trials.

2.6. MRI and fMRI analysis

Two T1-weighted MP-RAGE anatomical MRI scans were collected at the beginning of each scanning session with a 3 Tesla whole-body MR scanner (Phillips Medical Systems). The two anatomical scans were aligned to each other and averaged in order to provide maximal gray-white matter contrast. These scans were then used to create a cortical surface model using FreeSurfer [30,31] for visualization in SUMA [32]. For the fMRI scan series, T2* weighed images were collected using gradient echo-planar imaging (TR = 2000 ms, TE = 30 ms, flip angle = 90°) with in-plane resolution of 2.75×2.75 mm. Auditory stimuli were presented through MRI-compatible in-ear headphones (Sensimetrics, Malden, MA) which were covered with ear muffs to reduce the amount of noise from the scanner. Visual stimuli subtending approximately 20×30 degrees of visual angle were presented on a projection screen with an LCD projector and viewed through a mirror attached to the head coil. Responses to the target trials were collected using a fiber-optic button response pad (Current Designs, Haverford, PA). Analysis of the functional scan series was conducted using Analysis of Functional NeuroImages (AFNI) [33].

2.6.1. fMRI analysis: response amplitude and variability. In order to generate whole brain maps of the amplitude and standard deviation measures at each voxel, we carried out a voxel-wise analysis using the AFNI function *3dDeconvolve,* which uses maximum-likelihood estimation in the context of the generalized linear model (GLM). TENTzero functions were used to estimate the individual hemodynamic response function (using the option – *iresp*) and standard deviation of each response function (using the option – *sresp*) in each voxel for each stimulus type, beginning at stimulus onset and ending 16 seconds later for single syllables and 26 seconds later for blocks of words (the response was constrained to begin and end at zero amplitude). The model functions consisted of independent, piecewise linear impulse response functions (also known as stick

functions) that independently estimated the amplitude of the hemodynamic response at each time point following stimulus presentation. This methodological point is important because it allowed us to estimate the amplitude and standard deviation from the actual response in each individual voxel, not from a fixed hemodynamic response function such as a gamma variate. The use of a fixed function could introduce a confound because of differences in hemodynamic response functions; for instance, if older people had slightly broader hemodynamic response functions, then their deviation from a fixed function would be greater, unrelated to trial-to-trial variability.

For single syllables, we estimated the amplitude of the response as the mean of the response at 4 seconds and 6 seconds after stimulus onset (the peak of the hemodynamic response function). To estimate BOLD variability within each subject for single syllables, the standard deviation at the 4-second and 6-second time points of each impulse response function were averaged to produce a single value per voxel. The brain response to all audiovisual syllables (both response amplitude and variability) was similar, so they were combined for further analysis and only the average across stimulus types (excluding target trials) is reported.

2.6.2. Region-of-interest selection. Data from the whole-brain voxel-wise analysis (*2.6.1*) was first grouped using regions of interest created for each subject individually in native image space. ROIs were selected to target brain areas that are reliably active during multisensory speech perception [34]. A combination of anatomical and functional criteria was used. The anatomic parcellation of the cortical surface was constructed from each individual subject's structural scans with FreeSurfer [35,36]. Functional criteria were constructed from the independent localizer runs (see section 2.4 for details), eliminating bias [37].

We considered three contrasts when constructing the three regions of interest: auditory words vs. fixation baseline, visual words vs. fixation baseline, and audiovisual words vs. fixation baseline. The STS ROI was defined by finding all voxels in the posterior half of the anatomically parcellated STS that showed a significant response ($t > 2$, $p < 0.05$) during the localizer ($t > 2$ for auditory-only word blocks *vs.* baseline and t > 2 for visual-only word blocks *vs.* baseline). For 5 out of 19 older adults, no voxels in the left STS met this criterion, so an alternative criterion was used ($t > 2$ for audiovisual word blocks *vs.* baseline). The auditory cortex ROI was defined by finding voxels in the anatomically parcellated transverse temporal gyrus, lateral superior temporal gyrus and planum temporale that were significantly active during the auditory-only blocks ($t > 2$ for auditory-only word blocks *vs.* baseline). The extrastriate visual cortex ROI was defined by finding voxels in the anatomically parcellated extrastriate lateral occipitotemporal cortex that were active during the visual-only blocks ($t > 2$ for visual-only word blocks *vs.* baseline).

2.6.3. Whole-brain analysis. For the whole-brain voxel-wise analysis, subjects' individual data were first aligned to the N27 atlas brain [38] using the AFNI function *auto_tlrc*. Blurring kernels of approximately 3–6 mm have been found to be the most sensitive for detecting activation clusters [39]. We chose a 3×3×3 mm FWHM Gaussian kernel to minimize blurring between adjacent ROIs.

Table 2. Regions with greater intrasubject variability in older adults compared to younger adults.

Label	Area (mm^2)	Peak t-value
R superior temporal gyrus (54, −19, 1)	2341	5.0
L middle and inferior occipital gyri (−38, −69, −3)	1943	5.8
L planum temporale (−47, −29, 13)	1697	4.9
R inferior occipital gyrus (37, −61, −13)	882	4.2
R superior temporal sulcus (45, −56, 5)	163	5.3
Total	7026	

Regions are ranked by area (only clusters greater than 160 mm^2 are reported) and Talairach coordinates following anatomical label in (x, y, z) format are the weighted center of mass of the cluster.
doi:10.1371/journal.pone.0111121.t002

To conduct a voxel-wise search for any differences in response amplitude, the average response amplitude (average of the response to all non-target audiovisual speech stimuli relative to fixation baseline at the 4 and 6 second time points) was calculated in each voxel in each subject. *3dttest++* was used to perform an unpaired t-test for every voxel in standard space between the old and young adult groups. The results were mapped from the MRI volume to the cortical surface with *3dSurf2Vol* and masked with the group t-statistic ($t > 2$ for the contrast of all audiovisual syllables *vs*. baseline). After the voxel-wise t-test we preformed a clustering technique [40]. This finds only voxels that are significantly active above a particular threshold and spatially contiguous. The probability of finding two voxels above a particular threshold *and* being adjacent is much smaller than the chance of a single voxel above that threshold [41]. Using the AFNI program *slow_surf_clustsim.py*, we estimated that a cluster with a size of 160 mm^2 would have a corrected p-value of 0.045. A clusterizing filter on the surface (*SurfClust*) was applied and only regions larger than 160 mm^2 (and $t > 2$ for the contrast of all audiovisual syllables *vs*. baseline) are reported.

To conduct a voxel-wise search for differences in intersubject variability, the MATLAB function *vartestn* was used to perform a Bartlett's multiple sample test for equal variances on the response amplitudes, followed by clusterizing.

To conduct a voxel-wise search for differences in intrasubject variability, an unpaired t-test between groups was performed on the standard deviation of the response at each voxel (*3dttest++*) followed by clusterizing.

2.6.4. Motion correction. Functional data for each subject was first aligned to the averaged anatomical dataset for that subject and then motion-corrected using local Pearson correlation with the AFNI script *align_epi_anat.py* [42]. For each volume, an estimate of the amount of motion in each direction, relative to the reference, was produced. These estimates were used as regressors of no interest in the fMRI analysis. To capture a single value describing the amount of head motion in each subject, the standard deviation of each motion direction across time was averaged across motion directions.

In addition to the standard motion correction steps described above, we also performed the "motion scrubbing" procedure developed by Power and colleagues [43] to further investigate if motion was a potential confound in our main finding of increased intrasubject variability in older adults. First, motion estimates at each time point were calculated in each of the six motion directions (rotational measures: roll, pitch, yaw, and displacement measures: superior, left, and posterior directions). Rotational displacement measures were converted to millimeters using the formula $d = R*(\pi/180)*r$, where R is the rotation in degrees and r is the radius (we used r = 50 mm as prescribed by Power). To express the total amount of motion for each time point in a single value, the absolute value of the displacement in each direction was summed, where the total displacement at the *i*th data point was $D_i = |d_\alpha| + |d_\beta| + |d_\gamma| + |d_x| + |d_y| + |d_z|$. Then the framewise displacement for the *i*th time point was calculated as $FD_i = D_{(i-1)} - D_i$ to express instantaneous head motion. The scrubbing threshold was half of the smallest voxel dimension, as recommended by Power (EPI volumes were collected using a 2.75 mm isotropic voxel, therefore we used a threshold = 1.375 mm). The GLM analysis was then completed a second time, excluding the data points with a framewise displacement exceeding this threshold.

Results

3.1 Responses in the ROIs to audiovisual speech

Our initial analysis focused on three ROIs implicated as critical nodes in the network for multisensory speech perception: the left

Table 3. Regions with greater response amplitude in younger adults compared to older adults.

Label	Area (mm^2)	Peak t-value
R occipital pole (13, −94, 5)	302	3.3
R superior temporal sulcus (47, −36, 4)	273	3.8
L occipital pole (−17, −93, 5)	223	4.3
L subcentral sulcus (−53, −21, 13)	202	3.6
Total	1000	

Regions are ranked by area (only clusters greater than 160 mm^2 are reported) and Talairach coordinates following anatomical label in (x, y, z) format are the weighted center of mass of the cluster.
doi:10.1371/journal.pone.0111121.t003

Table 4. Regions with greater intersubject variability in older adults compared to younger adults.

Label	Area (mm²)	Peak χ²
L inferior occipital gyrus (−30, −75, −12)	1819	29
R inferior occipital gyrus (24, −81, −8)	1552	27
R superior temporal gyrus (62, −8, 2)	438	15
R transverse temporal gyrus (44, −24, 10)	200	7.8
R fusiform gyrus (34, −49, −19)	194	9.7
L superior frontal gyrus (−1, 0, 56)	164	27
Total	4367	

Regions are ranked by area (only clusters greater than 160 mm² are reported) and Talairach coordinates following anatomical label in (x, y, z) format are the weighted center of mass of the cluster.

superior temporal sulcus (STS), the left auditory cortex, and the left extrastriate visual cortex.

3.1.1. Variability of the hemodynamic response in ROIs within subjects. The left STS showed a robust hemodynamic response to audiovisual syllables that was similar in amplitude in individual older and younger subjects. An important behavioral difference in many behavioral paradigms between older and younger subjects is the variability across trials *within* individual subjects (intrasubject variability). To search for a neural correlate of this phenomenon, we calculated the hemodynamic response function in each subject in every voxel and then measured the standard deviation across trials at each time point in the hemodynamic response within each subject (plotting them as error bars around the mean at each time point). As shown in Figure 1A and 1B, while the response amplitudes were similar for individual older and younger subjects, the variability at each time point of the response was much larger in the older subject.

To quantify this difference, we averaged the standard deviation from the peak of the response (4 and 6 seconds after stimulus onset) to produce a single number for intrasubject variability for each ROI for each subject (Figure 1C–E). For each ROI, there was much greater within-subject standard deviation in older subjects (STS: 0.14% in older adults *vs.* 0.09% in younger adults, $t_{31} = 4.2$, $p = 2 \times 10^{-4}$; auditory cortex: 0.18% *vs.* 0.11%, $t_{31} = 5.9$, $p = 10^{-6}$; visual cortex: 0.18% *vs.* 0.08%, $t_{31} = 7.0$, $p = 10^{-8}$).

3.1.2. Mean and standard deviation of the hemodynamic response in ROIs across subjects. An unpaired t-test with percent signal change in the left STS as the dependent measure revealed slightly greater amplitude of mean response in younger adults (0.19% in younger adults *vs.* 0.12% in older adults, $t_{31} = 2.1$, $p = 0.048$). There were no significant differences in auditory cortex (0.26% *vs.* 0.24%, $t_{31} = 0.5$, $p = 0.63$) or visual cortex (0.16% *vs.* 0.10%, $t_{31} = 1.5$, $p = 0.15$).

The standard deviation of the response across subjects was also similar between groups (Figure 2; left STS: SD of 0.08% for younger adults *vs.* 0.12% for older adults, Bartlett's multiple sample test for equal variances $\chi^2_1 = 2.5$, $p = 0.12$; left auditory cortex: 0.14% *vs.* 0.12%, $\chi^2_1 = 0.7$, $p = 0.41$; left visual cortex: 0.08% *vs.* 0.12%, $\chi^2_1 = 2.3$, $p = 0.13$).

3.2. Whole-brain analysis

Our first set of analyses was limited to our three *a priori* ROIs created using block-design localizers. To overcome this limitation, and to prevent any biases introduced by slight differences in the localizers between old and young subjects, our second set of analyses examined the entire brain. First, we selected all voxels that showed a significant positive response ($t > 2$ for all audiovisual

syllables *vs.* baseline) to audiovisual syllables across old and young subjects (Table 1 and Figure 3A).

3.2.1. Variability of the whole-brain hemodynamic response within subjects. We calculated the variability of the response *within* each subject at each voxel, and then compared the two groups. Many brain regions showed greater intrasubject variability in older adults (Table 2 and Figure 3B), including left auditory cortex and bilateral extrastriate visual cortex (total area on the cortical surface = 7026 mm²; peak t-statistic = 5.8, $p = 2 \times 10^{-6}$). There were no regions where intrasubject variability was greater in younger adults.

3.2.2. Mean and standard deviation of the whole-brain hemodynamic response across subjects. We compared the amplitude and standard deviation of the response across subjects. Younger adults had greater amplitude of response in right and left visual cortex and right superior temporal sulcus (Table 3 and Figure 3C; total area = 1000 mm²; peak t-statistic = 4.3, $p = 2 \times 10^{-5}$; there were no regions where response amplitude was greater in older subjects). Older adults had greater standard deviation of response in bilateral visual cortex and right superior temporal cortex (Table 4 and Figure 3D; total area = 4367 mm²; peak Bartlett's $\chi^2_1 = 29$, $p = 9 \times 10^{-8}$; no regions with greater standard deviation in younger adults).

3.3. Potential confound: differences in head movements between younger and older adults

Our initial analysis used standard techniques for estimation and correction of head motion in the functional data, and our GLM included motion estimates as regressors of no interest. Estimated head motion was small in both groups but greater in older than younger adults (0.48 mm *vs.* 0.32 mm, $p = 0.004$). The finding of increased BOLD signal variability in older adults remained unchanged after applying a "motion scrubbing" procedure [43]: STS: 0.13% *vs.* 0.09%, $t_{31} = 3.4$, $p = 0.002$; auditory cortex: 0.18% *vs.* 0.10%, $t_{31} = 5.3$, $p = 8 \times 10^{-6}$; visual cortex: 0.17% *vs.* 0.09%, $t_{31} = 5.5$, $p = 1.6 \times 10^{-5}$. There was no correlation in either group between amount of head motion and intrasubject variability (older adults: r = 0.04, p = 0.88; younger adults: r = 0.16, p = 0.59). An ANCOVA with age group as one factor, head motion as a covariate, and standard deviation of the fMRI response as the dependent variable and revealed no interaction between age group and head motion in any ROI (p > 0.68), and the finding of increased intrasubject variability in older adults remained significant (STS: $F_{1,29} = 13$, $p = 0.001$; auditory cortex: $F_{1,29} = 25$, $p = 3 \times 10^{-5}$; visual cortex: $F_{1,29} = 34$, $p = 2 \times 10^{-6}$). A whole-brain ANCOVA that included the amount of head motion in each

subject as a covariate gave results nearly identical to the analysis without the head motion covariate. Discarding the six older adults with the greatest amount of head motion rendered group differences in head movements insignificant (0.39 mm *vs.* 0.32 mm, p = 0.10), but left the main finding of increased intrasubject variability intact (left STS: 0.14% *vs.* 0.09%, $t_{25} = 3.7$, p = 0.001; left auditory cortex: 0.18% *vs.* 0.11%, $t_{25} = 5.4$, p = 10^{-4}; visual cortex: 0.17% *vs.* 0.08%, $t_{25} = 6.7$, p = 5×10^{-7}).

Discussion

We compared brain responses to repeated presentations of identical audiovisual speech in healthy older and younger adults using fMRI. The most important finding was greater intrasubject variability in the older adults: across multiple presentations of identical stimuli, older adults had greater variability in their brain responses than younger adults. This was true across all of the brain areas that responded to audiovisual speech and was confirmed with two independent types of analysis (ROI and whole-brain). We also observed two less robust effects: older adults had smaller mean response amplitudes than younger adults, and the older adult group had a greater standard deviation of the response amplitude (intersubject variability) than younger adults.

What is the relationship between variability and perception/processing?

Across a variety of behavioral tasks, older adults have worse performance and increased intrasubject variability compared with young adults [11,44–47]. Older adults with mild dementia show more intrasubject variability than healthy age-matched controls [48] and healthy older adults with more trial-to-trial variability showed greater cognitive declines over time [12].

While a link between increased neural (BOLD fMRI) variability and increased behavioral variability is sensible on its face, the precise link between the two is a matter of speculation. With increased neural variability the distribution of responses in a population of neurons to a given stimulus would become wider, which in turn would make it harder for the brain to differentiate between the possible stimuli that evoked the response [49]. Consistent with this idea, decreased stimulus specificity has been observed in single cell recordings of older cats and non-human primates [13–15]. Older adults are particularly impaired in perceiving speech if it is embedded in artificially generated noise [2,4,50,51]. The increased noise in the stimulus could exacerbate the effects of increased neural variability, which could be considered "neural noise". We did not test older subjects using noisy audiovisual speech, the type of speech on which they are most impaired. Therefore, we could not directly compare the increased neural variability we observed in older subjects with the decreased performance of recognizing speech in noise that they are known to have. Future studies using noisy stimuli would be expected to produce poorer performance and reveal differences between subjects correlated with BOLD variability.

Neural variability at early stages of cortical sensory processing might be compounded by neural variability at decision layers higher in the cortical hierarchy. Speech perception involves categorical judgments about the identity of each syllable. Neuronal variability could impair these decisions, an effect that may be even more important than added sensory noise [52]. Neuronal variability may also differentially affect the ability to make both fine and coarse discriminations. Low levels of neuronal variability favor fine discriminations performed at locations in stimulus space in which neuronal selectivity changes rapidly, while high levels of neuronal variability favor coarse discriminations performed at

locations in stimulus space where neuronal responses are maximal [53]. Therefore, increased neuronal variability with aging might impair fine discrimination while leaving coarse discriminations relatively intact. Within individual trials, variability in neural responses can allow a population of neurons to encode multiple stimulus attributes. For instance, the mean of the population response may encode the stimulus estimate, while the variability of the population response encodes the uncertainty [54,55].

Previous fMRI studies showing greater inter/intrasubject variability in older adults

There have been a number of fMRI studies comparing young and older adults in other tasks, and the preponderance of these studies have reported more variability in older adults, as we observed in our dataset. D'Esposito et al. [56] measured BOLD responses in motor cortex to a bilateral button press cued by the appearance of a briefly presented white circle. Greater intrasubject variability in older adults compared to younger adults was observed; there were no differences in response amplitude and only a small increase in intersubject variability. Huettel et al. [57] measured responses in visual cortex to checkerboard stimuli with no behavioral task. Greater intrasubject and intersubject variability in older adults was observed, with no difference in amplitude of response. Samanez-Larkin et al. [58] found that healthy older adults exhibited suboptimal decision-making on a financial investment task compared to younger adults and also exhibited greater temporal variability in the nucleus accumbens. In contrast to these studies that reported more variability in older adults, Garrett et al. [21] reported *less* neural variability in older adults using a variety of complex cognitive tasks. One possible explanation for this result is that the analysis of Garrett et al. used a measure derived from multivariate voxel pattern analysis (MVPA) to measure variability across all brain voxels. MVPA analyses and traditional univariate analyses such as ours may give conflicting or even contradictory results. The explanation for these discrepancies is a matter of debate [59].

Potential Confounds

A potential confound in BOLD fMRI studies of older populations is vascular changes with age [60]. However, studies that directly measure neuronal activity also find age-related increases in variability. Anderson et al. [61] presented auditory syllables to healthy older adults and measured the auditory brain stem response, an electrophysiological measure of neuronal activity that is not influenced by the vasculature, and found greater variability in older adults. In a study of non-human primates, single neuron responses in V1 and MT of older monkeys had greater variability than in younger monkeys [62]. These findings suggest that the variability differences in our study have a neuronal component in addition to any possible vascular sources.

Intergroup differences in fMRI studies may also be driven by differences in head movements, especially in studies of resting state functional connectivity [43,63]. In a task-based study such as ours, averaging the response to multiple stimuli reduces movement effects, since head movements and stimulus presentation are independent. We used five different methods to account for differences in head motion, and found no effect on our main results. Consistent with these analyses, two previous studies (Huettel et al. [57] and D'Esposito et al. [56]) did not find a correlation between head motion and BOLD signal variability.

Conclusion

Our most robust finding was of greater intrasubject variability in older adults. This finding was true in both the ROI and whole-brain analysis. Among the physical changes to the aging brain, a decrease in myelination has been observed [64,65]. These decreases in white matter integrity could lead to increases in neuronal variability by preventing neurons from firing consistently even with the same sensory input. Better understanding the neural sources of this variability and its behavioral consequences may help in designing strategies to ameliorate declines in speech perception, one of our most important cognitive functions.

References

1. Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. Journal of the Acoustic Society of America 26: 212–215.
2. Dubno JR, Dirks DD, Morgan DE (1984) Effects of age and mild hearing loss on speech recognition in noise. The Journal of the Acoustical Society of America 76: 87–96.
3. Grant KW, Seitz PF (2000) The use of visible speech cues for improving auditory detection of spoken sentences. Journal of the Acoustical Society of America 108: 1197–1208.
4. Gosselin PA, Gagne JP (2011) Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. International journal of audiology 50: 786–792.
5. Peich MC, Husain M, Bays PM (2013) Age-related decline of precision and binding in visual working memory. Psychology and aging 28: 729–743.
6. Cerella J, Hale S (1994) The rise and fall in information-processing rates over the life span. Acta psychologica 86: 109–197.
7. Jenkins L, Myerson J, Joerding JA, Hale S (2000) Converging evidence that visuospatial cognition is more age-sensitive than verbal cognition. Psychology and aging 15: 157–175.
8. Hedden T, Gabrieli JD (2004) Insights into the ageing mind: a view from cognitive neuroscience. Nature reviews Neuroscience 5: 87–96.
9. Bielak AA, Cherbuin N, Bunce D, Anstey KJ (2013) Intraindividual Variability Is a Fundamental Phenomenon of Aging: Evidence From an 8-Year Longitudinal Study Across Young, Middle, and Older Adulthood. Developmental psychology.
10. Vandermorris S, Murphy KJ, Troyer AK (2013) Age-related elevations in intraindividual variability on associative memory tasks. Neuropsychology, development, and cognition Section B, Aging, neuropsychology and cognition.
11. Lovden M, Schmiedek F, Kennedy KM, Rodrigue KM, Lindenberger U, et al. (2013) Does variability in cognitive performance correlate with frontal brain volume? NeuroImage 64: 209–215.
12. Lovden M, Li SC, Shing YL, Lindenberger U (2007) Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: longitudinal data from the Berlin Aging Study. Neuropsychologia 45: 2827–2838.
13. Schmolesky MT, Wang Y, Pu M, Leventhal AG (2000) Degradation of stimulus selectivity of visual cortical cells in senescent rhesus monkeys. Nature Neuroscience 3: 384–390.
14. Hua T, Li X, He L, Zhou Y, Wang Y, et al. (2006) Functional degradation of visual cortical cells in old cats. Neurobiology of aging 27: 155–162.
15. Liang Z, Yang Y, Li G, Zhang J, Wang Y, et al. (2010) Aging affects the direction selectivity of MT cells in rhesus monkeys. Neurobiology of aging 31: 863–873.
16. Anderson S, Parbery-Clark A, Yi, Kraus N (2011) A neural basis of speech-in-noise perception in older adults. Ear and hearing 32: 750–757.
17. Stewart JC, Tran X, Cramer SC (2014) Age-related variability in performance of a motor action selection task is related to differences in brain function and structure among older adults. Neuroimage 86: 326–334.
18. Caspers S, Moebus S, Lux S, Pundt N, Schutz H, et al. (2014) Studying variability in human brain aging in a population-based German cohort-rationale and design of 1000BRAINS. Front Aging Neurosci 6: 149.
19. Grady CL, Garrett DD (2014) Understanding variability in the BOLD signal and why it matters for aging. Brain Imaging Behav 8: 274–283.
20. Garrett DD, Kovacevic N, McIntosh AR, Grady CL (2011) The importance of being variable. J Neurosci 31: 4496–4503.
21. Garrett DD, Kovacevic N, McIntosh AR, Grady CL (2012) The Modulation of BOLD Variability between Cognitive States Varies by Age and Processing Speed. Cerebral Cortex 23: 684–693.
22. Price LL (1963) Threshold testing with Bekesy audiometer. Journal of speech and hearing research 6: 64–69.
23. Brant LJ, Fozard JL (1990) Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging. The Journal of the Acoustical Society of America 88: 813–820.
24. Divenyi PL, Stark PB, Haupt KM (2005) Decline of speech understanding and auditory thresholds in the elderly. The Journal of the Acoustical Society of America 118: 1089–1100.
25. Folstein MF, Folstein SE, McHugh PR (1975)"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. Journal of psychiatric research 12: 189–198.
26. Saxe R, Brett M, Kanwisher N (2006) Divide and conquer: a defense of functional localizers. Neuroimage 30: 1088–1096; discussion 1097–1089.
27. Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RN (2006) A critique of functional localisers. Neuroimage 30: 1077–1087.
28. Poldrack RA (2007) Region of interest analysis for fMRI. Social cognitive and affective neuroscience 2: 67–70.
29. Wilson M (1988) MRC psycholinguistic database: Machine-usable dictionary, version 2.00. Behavioral Research Methods, Instruments, & Computers 20: 6–10.
30. Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9: 179–194.
31. Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9: 195–207.
32. Argall BD, Saad ZS, Beauchamp MS (2006) Simplified intersubject averaging on the cortical surface using SUMA. Human Brain Mapping 27: 14–27.
33. Cox RW (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research 29: 162–173.
34. Nath AR, Beauchamp MS (2011) Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. Journal of Neuroscience 31: 1704–1714.
35. Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. NeuroImage 53: 1–15.
36. Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, et al. (2004) Automatically parcellating the human cerebral cortex. Cerebral Cortex 14: 11–22.
37. Kriegeskorte N, Simmon WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nature Neuroscience 12: 535–540.
38. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, et al. (2001) A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philosophical Transactions of the Royal Society B: Biological Sciences 356: 1293–1322.
39. Skudlarski P, Constable RT, Gore JC (1999) ROC analysis of statistical methods used in functional MRI: individual subjects. NeuroImage 9: 311–329.
40. Xiong J, Gao J, Lancaster J, Fox P (1995) Clustered pixels analysis for functional MRI activation studies of the human brain. Human Brain Mapping 3: 287–301.
41. Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, et al. (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine/Society of Magnetic Resonance in Medicine 33: 636–647.
42. Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, et al. (2009) A new method for improving functional-to-structural MRI alignment using local Pearson correlation. NeuroImage 44: 839–848.
43. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage 59: 2142–2154.
44. Murphy KJ, West R, Armilio ML, Craik FI, Stuss DT (2007) Word-list-learning performance in younger and older adults: intra-individual performance variability and false memory. Neuropsychology, development, and cognition Section B, Aging, neuropsychology and cognition 14: 70–94.
45. MacDonald SW, Karlsson S, Rieckmann A, Nyberg L, Backman L (2012) Aging-related increases in behavioral variability: relations to losses of dopamine D1 receptors. The Journal of neuroscience : the official journal of the Society for Neuroscience 32: 8186–8191.
46. Hultsch DF, MacDonald SW, Dixon RA (2002) Variability in reaction time performance of younger and older adults. The journals of gerontology Series B, Psychological sciences and social sciences 57: P101–115.

47. West R, Murphy KJ, Armilio ML, Craik FI, Stuss DT (2002) Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. Brain and cognition 49: 402–419.

48. MacDonald SW, Nyberg L, Backman L (2006) Intra-individual variability in behavior: links to brain structure, neurotransmission and neuronal activity. Trends in Neurosciences 29: 474–480.

49. Li SC, Lindenberger U, Sikstrom S (2001) Aging cognition: from neuromodulation to representation. Trends in cognitive sciences 5: 479–486.

50. Stephan KE, Marshall JC, Friston KJ, Rowe JB, Ritzl A, et al. (2003) Lateralized cognitive processes and lateralized task control in the human brain. Science 301: 384–386.

51. Jones AK, Friston K, Frackowiak RS (1992) Localization of responses to pain in human cerebral cortex. Science 255: 215–216.

52. Beck JM, Ma WJ, Pitkow X, Latham PE, Pouget A (2012) Not noisy, just wrong: the role of suboptimal inference in behavioral variability. Neuron 74: 30–39.

53. Butts DA, Goldman MS (2006) Tuning curves, neuronal variability, and sensory coding. PLoS biology 4: e92.

54. Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nature Neuroscience 9: 1432–1438.

55. Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. Trends in Neurosciences 27: 712–719.

56. D'Esposito M, Zarahn E, Aguirre GK, Rypma B (1999) The effect of normal aging on the coupling of neural activity to the bold hemodynamic response. NeuroImage 10: 6–14.

57. Huettel SA, Singerman JD, McCarthy G (2001) The effects of aging upon the hemodynamic response measured by functional MRI. Neuroimage 13: 161–175.

58. Samanez-Larkin GR, Kuhnen CM, Yoo DJ, Knutson B (2010) Variability in nucleus accumbens activity mediates age-related suboptimal financial risk taking. The Journal of neuroscience : the official journal of the Society for Neuroscience 30: 1426–1434.

59. Jimura K, Poldrack RA (2012) Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia 50: 544–552.

60. Fang HCH (1976) Observations on aging characteristics of cerebral blood vessels, macroscopic and microscopic features. In: Terry RD, Gershon S, editors. Neurobiology of aging. New York: Raven Press. 155–166.

61. Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N (2012) Aging affects neural precision of speech encoding. The Journal of Neuroscience 32: 14156–14164.

62. Yang Y, Liang Z, Li G, Wang Y, Zhou Y (2009) Aging affects response variability of V1 and MT neurons in rhesus monkeys. Brain Research 1274: 21–27.

63. Van Dijk KR, Sabuncu MR, Buckner RL (2012) The influence of head motion on intrinsic functional connectivity MRI. NeuroImage 59: 431–438.

64. Lu PH, Lee GJ, Raven EP, Tingus K, Khoo T, et al. (2011) Age-related slowing in cognitive processing speed is associated with myelin integrity in a very healthy elderly sample. Journal of clinical and experimental neuropsychology 33: 1059–1068.

65. Kerchner GA, Racine CA, Hale S, Wilheim R, Laluz V, et al. (2012) Cognitive processing speed in older adults: relationship with white matter integrity. PLoS one 7: e50425.