Increased speech representation in older adults originates from early and late responses in auditory cortex

Christian Brodbeck, Alessandro Presacco, Stefanie Kuchinsky, Samira Anderson & Jonathan Z. Simon
Puzzle

- Compared to young adults, older adults exhibit:
  - Impaired auditory temporal processing
  - More difficulty comprehending speech, especially in challenging circumstances

- Yet, the speech envelope can be reconstructed more accurately from their cortical responses, recorded with MEG (Presacco et al., 2016)

Different possible explanations, for example...

- Increased cortical gain of bottom-up responses
- Recruitment of additional top-down resources
- Physiological changes, e.g. excitation-inhibition imbalance

This talk

- Localize cortical responses to speech of younger and older adults
  - Anatomy: localization in cortex
  - Time: latency at which information is represented
areas (blobs) with fMRI and then use MEG or EEG to determine the temporal relationships between these areas. However, this approach would work only if the physiological sources of both MEG/EEG and fMRI were identical. Discrepancies between the methods are to be expected to some extent as fMRI reflects neural activity only indirectly, via the BOLD (blood oxygenation-level dependent) signal arising from neurovascular coupling, whereas MEG/EEG pick up signals directly related to the neuronal activity (for time courses, see figure 1). Another apparent difference between the methods, not yet discussed in the literature as far as we know, is that MEG/EEG weights strongly the fastest-conducting pathways, whereas fMRI probably receives its main contribution from neuronal ensembles that are connected via slow and thin fibres, thereby apparently reflecting functionally different brain activations. This difference further emphasizes the complementary nature of these methods.

How could we improve non-invasive time-resolved imaging of the human brain? First of all, the measurements should be performed as close to the neural generators as possible. However, with an intact skull, the distance from the outside of the head to the closest sources in the cortex is at least 1.5 cm, which sets an upper bound for the spatial frequencies (how fast the signals change in space) and thereby for the resolution any MEG or EEG sensor array can provide. In the present-day SQUID-based magnetometers, the sensors are as far as 4–5 cm from the most superficial sources. Since MEG picks up signals mostly from sources in fissural cortex, the sensor-to-source distance can approach even 7 cm. In children, when measured with the current adult-head-optimized devices, these numbers are even larger if both hemispheres are to be measured.
Methods (Presacco et al.)

Design

- 60 s long audiobook excerpts, 3 repetitions each
- 2 excerpts were clean speech
- 8 excerpts with second speaker at different signal to noise ratios (SNRs; +3, 0, -3, -6 dB)

Participants

- 17 young adults (aged 18-27 years)
- 15 older adults (aged 61-73 years)
  - Cognitive screening
  - Clinically normal audiogram

MEG data

- KIT MEG Lab at University of Maryland, 157 axial gradiometers
- Band pass filter 1-8 Hz
Methods (Presacco et al.)

his schoolhouse was a low building of one large room rudely constructed of logs

Speech envelope

"Decoder"

Continuous MEG recording

Time [seconds]
**Results (Presacco et al.)**

Significantly lower values than 350 ms but not than 150 ms \[t(16) / H11005 3.722, P / H11005 0.002 \text{ and } t(16) / H11005 0.973, P / H11005 0.345 \text{ for } 500 \text{ vs. } 350 \text{ ms and } 500 \text{ vs. } 150 \text{ ms, respectively}.\]

Conversely, older adults' ability to track the speech envelope of the foreground is significantly reduced at 350 and 150 ms in both quiet \[t(14) / H11005 0.248, P / H11005 0.807 \text{ and } t(14) / H11005 3.779, P / H11005 0.002 \text{ for } 500 \text{ vs. } 350 \text{ ms and } 500 \text{ vs. } 150 \text{ ms, respectively}.\]

Reconstruction of the unattended speech envelope.

Repeated-measures ANOVA showed a significant correlation\[/H11003 age\]/H11003 interaction across the four noise conditions tested \[F / H11005 (3,90) 2.909, P / H11005 0.039\].

A one-way ANOVA showed significantly higher reconstruction accuracy in older adults at all of the noise conditions tested except \[F / H11001 3; 0; F / H11002 3; \text{ and } F / H11002 6d B, \text{ respectively} \]. All of the reconstruction values were significantly higher than the noise floor (all, \(P / H11021 0.01\)).

**DISCUSSION**

The results of this study provide support for most, but not all, of the initial hypotheses. Behavioral data showed that older adults do have poorer speech understanding in noise than younger adults, despite their normal, audiometric hearing thresholds. In midbrain, noise suppresses the response in younger adults to a greater extent than in older adults, whereas the fidelity of the reconstruction of speech in cortex is higher.
Midbrain (Presacco et al.)

Midbrain

- Older listeners have reduced frequency following response (FFR)
- Increased cortical responses not due to stronger input from midbrain

**Midbrain: younger > older**

(Presacco, Simon, & Anderson, 2016)
Increased cortical gain for bottom-up responses

- Prediction: same origin, more current

Top-down/strategic processing

- Compensate for degraded input from the periphery
- Recruitment of additional frontal and temporal regions for complex sentences (Peelle et al., 2010)
- Prediction:
  - Response enhancement at longer latencies, e.g., 100-200 ms

Low level physiological change: excitation/inhibition imbalance

- Reduction in inhibitory neurons in A1 (de Villers-Sidani et al., 2010)
- Increased firing rates in A1 (Overton & Recanzone, 2016)
- Faster recruitment of higher order regions (Engle & Recanzone, 2013)
- Prediction:
  - Enhanced low latency responses, e.g., 30 ms
  - Potentially involving higher order regions
Methods

Participants
- 17 young adults (aged 18-27 years)
- 23 older adults (aged 61-73 years)

MEG source localization
- Empty room noise covariance
- Minimum norm estimates with depth weighting
- Temporal response functions estimated with coordinate descent algorithm (David et al., 2007)
  - Minimizing $\ell^1$ error
  - Stopping based on cross-validation

Evaluate model predictions:
- At each source element: Pearson correlation $r$(predicted response, measured response)

Bias-correction:
- Compute $r$ of a temporally shuffled model
- Test for better $r$ of the true model

Significance test:
- Mass-univariate $t$-test (Smith & Nichols, 2009)
  - Threshold-free cluster enhancement
  - Max statistic distribution with 10,000 permutations
Temporal response function
Encoding model

his schoolhouse was a low building of one large room rudely constructed of logs

Speech envelope

Continuous MEG source estimates

<table>
<thead>
<tr>
<th>Time [seconds]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1</td>
</tr>
<tr>
<td>2 2</td>
</tr>
<tr>
<td>3 3</td>
</tr>
<tr>
<td>4 4</td>
</tr>
<tr>
<td>5 5</td>
</tr>
</tbody>
</table>
Clean speech: neural localization

Brain activity (MEG source estimate) predicted from acoustic envelope

- Maps of correlation ($r$) between actual and predicted neural time course

Older > Younger

- Ventral to core auditory cortex
- No significant difference between hemispheres

Older

Younger

O > Y

Heschl's gyrus (A1)
Superior temporal gyrus

Hemisphere

Left
Right

$p$
Temporal response function

his schoolhouse was a low building of one large room rudely constructed of logs

Speech envelope

Continuous MEG source estimates

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time [seconds]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Temporal response function (TRF)

- Brain response to an elementary temporal feature in the stimulus
- Time axis: latency between acoustic feature and brain response

~30 ms
- Bottom-up gain (involving non-core area)
- Top-down (early)
  - Consistent with excitation/inhibition imbalance

~180 ms
- Bottom-up gain (no comparable response in younger subjects)
  - Recruiting additional neural resources?
Listening to two speakers (Puvvada & Simon, 2017)

- Early responses track the acoustic signal (~50 ms)
- Later responses track the attended speaker (~100 ms)

~30 ms
- Stimulus-driven
- Consistent with excitation-inhibition imbalance

~120 ms
- Increased attentional modulation
- Consistent with increased task-related processing

~180 - 250 ms
- Continued tracking of mix and attended speaker
- Responses practically absent in younger listeners
Cortical over-representation of speech in older adults:

- Multiple sources of over-representation

~ 30 ms

- Bottom-up cortical gain
  - Main difference outside of core auditory cortex
- Strategic/top-down processing
  - Latency too short
- Low level physiological change; excitation/inhibition imbalance
  - Short latency
  - Fast spread to areas outside core auditory cortex

~ 120 ms

- Bottom-up cortical gain
  - Does not track bottom-up information
- Strategic/top-down processing
  - Increase in task related activity (attention to speech)
- Low level change
  - Effect on task-related activity?

Later responses

- Bottom-up cortical gain
- Enhanced attentional tracking compatible with cognitive effort/compensation
- Persistent stimulus-driven as well as task-related activity
Coauthors

- Alessandro Presacco
- Stefanie Kuchinsky
- Samira Anderson
- Jonathan Z. Simon

Funding: NIH

- Grant P01-AG055365