Title: Speech motor facilitation is not affected by ageing but is modulated by task demands during speech perception

Abbreviated title: Effect of ageing on speech motor cortex

Helen E. Nuttall¹, Gwijde Maegherman², Joseph T. Devlin³, Patti Adank²

1. Department of Psychology, Lancaster University, Fylde College, Fylde Avenue, Lancaster, UK, LA1 4YF

2. Department of Speech, Hearing and Phonetic Sciences, University College London, Chandler House, 2 Wakefield Street, London, UK, WC1N 1PF

3. Department of Experimental Psychology, University College London, 26 Bedford Way, London, UK, WC1H 0AP

Corresponding author: Dr Helen Nuttall, Department of Psychology, Lancaster University, Fylde College, Fylde Avenue, Lancaster, UK, LA1 4YF

Acknowledgements: We gratefully acknowledge Dan Kennedy-Higgins’ contribution to data collection and SRT material development, as well as the time and effort given by all the participants who took part in this study. This work was funded by a Project Grant from The Leverhulme Trust, RPG-2013-254.

Conflict of interest: Authors report no conflict of interest.
Abstract

Motor areas for speech production activate during speech perception. Such activation may assist speech perception in challenging listening conditions. It is not known how ageing affects the recruitment of articulatory motor cortex during active speech perception. This study aimed to determine the effect of ageing on recruitment of speech motor cortex during speech perception.

Single-pulse Transcranial Magnetic Stimulation (TMS) was applied to the lip area of left primary motor cortex (M1) to elicit lip Motor Evoked Potentials (MEPs). The M1 hand area was tested as a control site. TMS was applied whilst participants perceived syllables presented with noise (-10, 0, +10 dB SNRs) and without noise (clear). Participants detected and counted syllables throughout MEP recording. Twenty younger adult subjects (aged 18-25) and twenty older adult subjects (aged 65-80) participated in this study.

Results indicated a significant interaction between age and noise condition in the syllable task. Specifically, older adults significantly misidentified syllables in the 0 dB SNR condition, and missed the syllables in the -10 dB SNR condition, relative to the clear condition. There were no differences between conditions for younger adults. There was a significant main effect of noise level on lip MEPs. Lip MEPs were unexpectedly inhibited in the 0 dB SNR condition relative to clear condition. There was no interaction between age group and noise condition. There was no main effect of noise or age group on control hand MEPs. These data suggest that speech-induced facilitation in articulatory motor cortex is abolished when performing a challenging secondary task, irrespective of age.

Keywords: speech perception; ageing; motor cortex; TMS; MEPs

1. Introduction

It is well-known that age-related auditory and cognitive decline lead to difficulty understanding speech in social settings, particularly when background noise is present (Pichora-Fuller, 2003; Pichora-Fuller et al., 2016). Difficulty communicating in noise leads to reduced social interaction in noisy environments, increasing isolation and decreasing mental wellbeing. Indeed, several studies have found a significant relation between age-related hearing loss and the incidence of cognitive decline and dementia (Gurgel et al., 2014; Thomson, Auduong, Miller, & Gurgel, 2017). The relation between age-related hearing loss and cognitive decline remains after controlling for age, gender, race, education, diabetes, smoking history, and hypertension (Peelle & Wingfield, 2016). It is therefore essential to understand how the neurobiological network subserving speech perception changes with age, so we can best support social communication throughout the ageing process.

Results from past studies indicate that brain areas including and extending beyond primary and association auditory cortices are important for successful speech perception. Cortical regions including, but not limited to, ventral premotor cortex, inferior frontal gyrus, and supplementary and primary motor areas have also been suggested to be involved in speech perception (Adank, Davis, & Hagoort, 2012; Londei et al., 2010; Schwartz, Basirat, Ménard, & Sato, 2012; Skipper, Devlin, & Lametti, 2017; Tremblay, Sato, & Small, 2012). Indeed, it is now largely accepted that articulatory motor areas are active when we perceive speech (Bartoli et al., 2015; Pulvermüller & Fadiga, 2010; Smalle, Rogers, & Möttönen, 2014). The extent of motor activation during speech perception varies depending on the acoustic properties of the stimulus and listening conditions. Murakami et al. (2011) demonstrated that lip motor evoked potentials (MEPs), elicited by transcranial magnetic stimulation (TMS) to the lip area of primary motor cortex (M1) are enhanced when perceiving
speech in noise relative to perceiving speech without noise. This finding has been interpreted to reflect increased excitability in the cortical motor representation of the lips when listening to degraded speech. This effect has been replicated and extended using similar paradigms and different types of stimuli containing internal and external distortions (Nuttall, Kennedy-Higgins, Devlin, & Adank, 2018; Nuttall, Kennedy-Higgins, Hogan, Devlin, & Adank, 2016; Nuttall, Kennedy-Higgins, Devlin, & Adank, 2017). Relatedly, if a listener perceives speech that they believe to be dissimilar to the speech they themselves produce, i.e., due to a different accent, there is greater speech motor activity during perception of these speech utterances (Bartoli et al., 2015). This result indicates that motor recruitment during speech perception is proportional to the variability and unpredictability in the auditory environment, as well as to listening difficulty.

Knowledge of the motor contribution to human speech perception, and of the neurobiological network for human speech perception in general, has been largely derived through studying speech processing in younger adults. It remains unclear if healthy ageing engages a similar neurobiological network, with similar scalable activity across motor areas. It has been hypothesised, however, that in response to age-related sensory decline, there is compensatory recruitment of more general cognitive areas (Du, Buchsbaum, Grady, & Alain, 2016; Erb & Obleser, 2013). Indeed, neuroimaging evidence indicates that prefrontal regions associated with cognitive control, attention, and working memory show increased activation when older adults process speech under challenging listening conditions (Erb & Obleser, 2013; Peelle, Troiani, Grossman, & Wingfield, 2011; Vaden, Kuchinsky, Ahlstrom, Dubno, & Eckert, 2015). It remains unclear whether similar adaptive recruitment also occurs within the neurobiological network for speech processing. The speech motor system may be a particular candidate for such adaptive activity following sensory decline, given speech motor activity scales relative to the variability and unpredictability in the auditory environment, as well as to listening difficulty (Bartoli et al., 2015; Murakami et al., 2011; Nuttall et al., 2017).

One recent study used functional magnetic resonance imaging (fMRI) to investigate such neurobiological adaptations in ageing during speech perception (Du et al., 2016). Speech motor areas showed phoneme-specific activation to different consonant-vowel syllable (/ba/, /ma/, /da/ and /ta/) when listening to speech in noise. However, bilateral auditory cortices showed phoneme-specific activation only when the noise was absent or presented at a very low level. Importantly, a positive correlation under noise masking between activity in speech motor regions and behavioural accuracy was observed in older adults, which indicated a possible compensatory frontal upregulation. Furthermore, this sensorimotor integration occurred at lower task demands in older adults compared with young adults. This increased sensorimotor integration at lower task demands could indicate that older adults engage more speech motor compensation during perception of speech in noise, even at low levels of noise, compared to younger adults.

Relatedly, Panoullieres and colleagues (2018) investigated the excitability of tongue motor cortex during passive speech perception in younger and older listeners, using TMS and tongue MEPs as a real-time quantification of activity in the motor pathway to the tongue muscles. It was observed that the excitability of the tongue motor pathway was similarly facilitated during perception of clear speech in younger and older adults with normal hearing, whereas tongue excitability during perception of speech was reduced in older adults with hearing loss. This study suggests that ageing alone does not enhance speech motor facilitation during passive speech perception. However, the finding that speech motor facilitation is not enhanced contrasts with the findings of Du and colleagues (2016), who found that ageing resulted in enhanced motor activity during active speech perception. One reason for these contrasting results may arise from methodological differences in
the two studies. Du et al. (2016) utilised an active speech perception task where participants had to
detect phonemes, whereas Panoullieres et al. (2018) tested passive speech perception. It therefore
seems likely that the task-demands imposed on the listener can modulate observed motor activity
and are relevant to understanding how neurobiological activity during speech processing changes
with ageing.

Our study aimed to determine how ageing influences speech motor excitability during active
speech perception using TMS to measure MEPs. Participants detected speech syllables under
different levels of background noise. With regards to ageing, we had two alternative hypotheses: 1) If the motor compensation hypothesis is supported, then older adults will show greater lip
excitability compared to younger adults particularly when cognitive load is high, in line with fMRI
data from Du et al (2016). 2) If motor activation is maintained and does not adapt as we age, then
there will be no difference in lip excitability between younger and older adults, in line with MEP data
from Panoullieres et al (2018). We also measured a number of variables to ascertain if subjects
differed in cognitive or speech processing ability as a function of age, including memory, hearing,
cognitive ability, and speech reception threshold.

2. Methods

2.1 Subjects

Forty subjects took part in the study; 20 younger adults (8 males); average age: 20.5, (SD
0.4), and 20 older adults (9 males); average age: 69.2 (SD 2.3). All subjects were right-handed,
monolingual, native speakers of British English, with reportedly normal language function.

Handedness was established via self-report. All participants reported normal hearing. Pure-tone
audiometric hearing thresholds were established using a diagnostic audiometer (AD229b,
Interacoustic A/S, Denmark) in accordance with The British Society of Audiology Recommended
Procedure (The British Society of Audiology, 2011), across 0.25, 0.5, 1, 2, and 8 kHz bilaterally. All
younger subjects had clinically normal thresholds bilaterally across all frequencies of ≤20 dB Hearing
Level (HL). The older adult group had clinically normal thresholds from 250 Hz through to 1 kHz, mild
hearing impairment at 2 and 4 kHz, and moderate impairment at 8 kHz. Pure-tone average (PTA)
audiometric thresholds were computed across the major speech frequencies from 500 Hz to 4 kHz
averaged across both ears, and this value was used for subsequent analyses. Subjects presented no
TMS contraindications, and did not report any neurologic/psychiatric disease, or that they were
under the effect of neuroactive drugs. All subjects had a minimum high school-level education, with
the majority studying to University level. There was no significant difference in educational level
between the two groups (p > 0.2). Younger subjects were recruited via a university research
participation platform, and older subjects were recruited via local community groups. Experiments
were undertaken with the understanding and written consent of each subject, according to the
university’s Research Ethics Board (UREC #0599/001).

2.2 Speech stimuli

Stimuli consisted of twenty vowel-consonant-vowel (VCV) syllables containing an equal
distribution of lip- (/apa/, /aba/) or tongue-articulated (/ata/, /ada/) syllables. All stimuli were
recorded in a sound-attenuated room and produced by a twenty-seven year old female British
English speaker with an average vocal pitch of 215 Hz. All stimuli were naturally produced to be of
approximately the same duration (mean 975 ms) but were not synthetically manipulated to be precisely the same length. Stimuli varied by a standard deviation of 61 ms. All stimuli were produced with natural falling intonation, with stress placed on the initial syllable. Audio digitizing was performed at 44.1kHz with 16 bits. All syllables were amplitude root-mean-square normalized offline using Praat (Boersma and Weenink, 2016), and then presented using Matlab (R2013a; The Mathworks Inc., Natick, MA) through ultra-shielded insert earphones (ER-2; Intelligent Hearing Systems, FL), at a comfortable level of around 65 dB SPL. The presentation level was consistent across subjects. For each subject, a stimulus list containing five occurrences of /apa/, /aba/, /ata/ and /ada/ stimuli was randomly permuted, and stimuli were presented according to this order. This was repeated twice without cessation for all conditions (40 stimuli presented in total for each condition). The speech-shaped noise used in the speech-in-noise conditions was created in Matlab using a custom-written script, and contained the same long-term average spectrum as speech, but without amplitude modulation. It is a non-speech noise signal. The noise was presented at three different signal-to-noise ratios, with the noise presented either 10 dB SPL less than the speech, at the same dB SPL as the speech, or 10 dB SPL more than the speech. The experiment took place in a non-Faraday caged, double-walled sound-attenuating booth.

2.3 Design

The experiment was designed to test how activity in lip M1 is differentially modulated in older and younger adults when actively listening to speech presented at different signal-to-noise ratios (SNRs). The noise levels were chosen to cover a range of SNRs from favourable to challenging due to a paucity of data on how speech motor processing interacts with speech-in-noise. Subjects were instructed to listen to the speech stimuli attentively and identify the consonant in the vowel-consonant-vowel sounds presented. The order of experimental conditions was randomised. The following four experimental conditions were tested in blocks:

1) Clear: speech only
2) At an SNR +10 dB SPL: speech stimuli were presented in speech-shaped noise that was 10 dB SPL less intense than the speech
3) At an SNR of 0 dB SPL, where speech and noise were at an equivalent intensity
4) At an SNR of -10 dB SPL: speech stimuli were presented in speech-shaped noise that was 10 dB SPL more intense than the speech

The order of conditions was randomised across participants and all MEP data were collected within a single session. To ensure active engagement with the speech stimuli, subjects were asked to listen carefully to the speech sounds and to identify the consonant in the middle of the sound as either a ‘p’, ‘b’, ‘t’ or ‘d’. At the start of each condition, subjects were instructed to count how many syllables containing consonant ‘x’ were presented (where ‘x’ was either ‘p’, ‘b’, ‘t’, or ‘d’). Which consonant they were asked to listen to and count was randomised across participants and conditions. Subjects were asked to retain this information in memory until the end of the condition, when they could report it verbally to the experimenter. Subjects were not given any aids, writing or otherwise, to assist them with counting the number of stimuli, and were instructed not to overtly articulate any information or use finger-counting. All TMS conditions took place in one session, which lasted approximately two hours. Subjects returned for a second session lasting approximately 1 hour 30 minutes, where PTA, working memory, short-term memory, Montreal Cognitive Assessment (MOCA), and speech reception thresholds (SRTs) were also measured. No aspect of the sessions
were video-recorded, and verbal responses were verified by the experimenter, who was present with the participant throughout both sessions.

2.4 Transcranial magnetic stimulation

Monophasic single TMS pulses were generated by a Magstim 200 unit and delivered by a 70mm diameter figure-of-eight coil, connected through a BiStim2 module (Magstim Co Ltd, Whitland, UK) set to simultaneous discharge mode (inter-pulse spacing of 1 ms). The coil was placed tangential to the skull at a 45 degree angle from the midline such that the induced current flowed from posterior to anterior under the junction of the two wings of the figure-of-eight coil. The lip area of M1 was found by using the functional ‘hot spot’ localization method, whereby application of TMS elicits an MEP from the contralateral muscle, which is under active contraction at approximately 20% of a subject’s maximal contraction. Here, the coil position is adjusted in millimetre movements to ascertain the location on the scalp at which the most robust MEPs are elicited. This location was then marked on a cap and active motor threshold (aMT) determined, which constitutes the intensity at which TMS pulses elicited 5 out of 10 MEPs with an amplitude of at least 200 μV (Möttönen, Rogers, & Watkins, 2014). In this way, we first located the hand area by asking subjects to perform a pinching action where the index finger was held against the thumb to activate the first dorsal interosseous. Following this, participants were asked to purse their lips (see section 3.5 Electromyography for details) and the lip area ‘hot spot’ was identified by moving the coil ventrally and slightly anterior until an MEP was observed in the contralateral lip muscle, and the aMT identified – this method is well documented in MEP literature and is a valid way to localise M1 in the absence of subject-specific MRI scans (Möttönen et al., 2014). In the younger group, the mean aMT for lip was 50% (SD 5.4%), and 40% for the hand (SD 7%). For the older group, mean aMT for lip was 45% (SD 7.4%), and 37% (SD 6.5%) for the hand. A mixed ANOVA with ‘muscle threshold’ as the within-subjects factor (lip, hand), and ‘age group’ as the between-subjects factor (older adults, younger adults) indicated a significant difference between lip and hand motor thresholds (F(1,38) = 40.70, p < 0.001), which is to be expected as hand motor thresholds are known to be lower than speech motor thresholds (Adank, Nuttall, & Kennedy-Higgins, 2016). There was no significant interaction between age group and motor threshold (F(1,38) = .62, p = .44). The intensity of the stimulator was set to 120% of individual participants’ aMT for the stimulation applied during the experiment in order to ensure individual TMS pulses elicited a measurable MEP.

After establishing TMS test intensity, all subjects then received four test blocks of single-pulse TMS to the lip area of M1 in the left hemisphere, and four blocks of single-pulse TMS to the hand area of M1 in the left hemisphere. The muscle order was randomised across subjects, and all blocks of muscle data were collected consecutively. During the TMS test blocks, subjects were presented with the speech stimuli, and were asked to listen and count syllabic information (see Methods 2.3). During the presentation of each speech stimulus, Matlab (R2013a; The Mathworks Inc., Natick, MA) was used to externally trigger the TMS system, such that a TMS pulse was generated 100 ms after the onset of the consonant in each stimulus type. TMS timing was based on previous MEP studies where excitability of the left articulatory motor cortex has been found to occur from 100 ms during speech perception (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Sato, Buccino, Gentilucci, & Cattaneo, 2010). All speech stimuli were accompanied by a TMS pulse; therefore, all 40 stimuli in each block were presented with TMS. There was an inter-stimulation delay of between 4.5-5 s. Participants were given short breaks in between TMS test conditions, during which time the coil was changed to prevent over-heating.
2.5 Electromyography

Electromyographic (EMG) activity was recorded from the lip muscle orbicularis oris using surface electrodes (Ag/AgCl; 10-mm diameter) in a non-Faraday caged, double-walled sound-attenuating booth. Electrodes were attached to the orbicularis oris on the right side of the mouth in a bipolar montage, with an electrode placed at the right temple serving as a common ground. To stabilize background EMG activity, subjects were trained for approximately five minutes to produce a constant level of contraction (approximately 20-30% of maximum voluntary contraction) of the lip muscles by pursing, which was verified via visual feedback of the ongoing EMG signal, in accordance with standard practice for lip aMT thresholding (Möttönen et al., 2014; Watkins, Strafella, & Paus, 2003). Contraction of the lip muscle also facilitates a lower motor threshold relative to when the muscle is at rest, enabling the use of lower levels of stimulation during the experiment. The raw EMG signal was amplified by a factor of 1000, band-pass filtered between 100–2000 Hz, and sampled at 5000 Hz online using a 1902 amplifier (Cambridge Electronic Design, Cambridge, UK), and analog-to-digital converted using a Micro1401-3 unit (Cambridge Electronic Design, Cambridge, UK). Continuous data were acquired and recorded using Spike2 software (version 8, Cambridge Electronic Design, Cambridge, UK).

2.6 Speech processing ability

Participants’ ability to perceive speech in noise was assessed by comparing their performance on the Speech Reception Threshold (SRT) task (Plomp & Mimpen 1979a, 1979b). In this task, sentences are presented in speech-shaped noise with the signal-to-noise ratio (SNR) varying adaptively depending on individual participant performance. The task is designed to determine the threshold at which participants can recognise about 50% of keywords in a sentence, and is similar to clinically available speech perception tests.

The first sentence was presented at a favourable SNR, for example +20 dB. Correct repetition of three or more keywords resulted in a reduction of 10 dB on subsequent trials, until participants were unable to correctly repeat more than two keywords. At this point the SNR increased in steps of 6dB until another reversal occurred (i.e., participants’ correct repetition of three or more of the keywords) with all subsequent changes occurring in steps of 4 dB. A reversal refers to the shift in direction of SNR change from one trial to the next, for example, if a participant repeated more than three keywords for four sentences in a row, then the SNR will reduce after each sentence making the subsequent sentence on each occasion harder to perceive. If on the fifth sentence the participant was unable to repeat at least three of the keywords, the SNR will increase making the subsequent sixth trial easier to understand. Such a change in direction from decreasing to increasing (or vice versa) SNR represents a ‘reversal’. Participants’ SRTs were computed by taking the mean SNR (dB) from all trials where a reversal occurred tracking at 50% (Plomp & Mimpen, 1979a, 1979b).

After presentation of each sentence, participants were asked to repeat verbatim what they heard. Responses were scored online immediately after each trial using a graphical user interface (GUI) on a standard computer screen that was not visible to participants. Each sentence contained five keywords upon which scoring was based, for example: “‘The MEAL was COOKED BEFORE the BELL RANG’” (keywords in uppercase letters). Keywords were also judged to be correct if participants changed the grammatical number of presented words e.g. ‘Meals’ (plural) instead of ‘Meal’ (singular). All other digressions variations were scored as incorrect with no feedback given.
Orders of sentence list were counterbalanced using a Latin-square technique. All sentences were pseudo-randomly ordered such that the order of presentation was different between participants, but each sentence was only presented once per participant.

2.7 Digit span tasks

To determine forward and working memory spans, participants were presented with auditory strings of digits of increasing length, which they were asked to repeat. At the start of the task, participants were made aware they would be asked to repeat strings of digits in the forward order (the order in which they had heard the digits), and then strings of digits in the reverse order, depending on experimenter instruction. An example was given using a digit string of two, e.g. “8, 3”. Digits were read by the experimenter in a neutral voice and tone and at a rate of approximately one digit per second. Participants were told that the first 6 strings of digits should be repeated in the forward order, while the second 6 strings of digits should be repeated in the reverse order – this change from forward to reverse order was also indicated verbally by the experimenter each time during the task. The starting string was two digits long. If participants recalled at least five out of six digits correctly, the test continued by increasing the number of digits to three, and so on. If participants recalled fewer than five out of six digits correctly, the test was concluded. If participants recalled five out of six digits for the forward but not the backward repetition, the backward repetition task was concluded but the forward repetition task continued, and vice versa.

2.8 Data analysis

The difference score regarding syllables presented relative to syllables detected (syllables presented-syllables detected) was calculated for each individual in each condition as a measure of their active listening score. A score of zero would indicate perfect performance. A score greater or less than zero would indicate less accurate performance. For the TMS data, individual EMG sweeps starting 40 ms before the TMS pulse and ending 40 ms post-stimulation were exported offline from the recording software into Matlab, where average MEPs were calculated for each condition in each participant. MEPs exhibit substantial variability for reasons that are not fully understood, but may reflect biological variation associated with the phase in the Mu rhythm at the time the MEP was elicited, differences in neural recruitment, and differences in circadian rhythms (e.g., Keil et al., 2014; Wassermann, 2002). Individual averages were rectified and the integrated area under the curve (AUC) of this rectified EMG signal of each individual mean MEP was calculated from 8-35 ms post-stimulation, which captured the time at which lip MEPs are expected to occur (Devlin & Watkins, 2007). Deriving the area under the curve from averages in this manner was useful to determine the systematic structure of the time-locked MEP, which removed subjective bias associated with identifying the MEP window for each individual trial. The area under the curve of the MEP was then converted into standard scores within-subjects (not standardised to group mean), to control for inter-subject variability. The standardized AUCs of MEPs were used in the statistical analyses. We included all MEP trials in the analysis and did not selectively pick trials with MEPs exceeding a certain criterion in order to avoid bias. Raw (non-standardised) Lip and Hand AUCs were tested (two-tailed) for age-group differences in AUC means and standard deviations. No significant differences were observed (p > 0.1).

Statistical analyses were carried out using SPSS (version 28.0, IBM). Separate two-tailed independent t-tests were conducted for the dependent variables of PTA, MOCA, SRT score, and
short-term, and working memory, with age group as the grouping factor. Mixed-factor ANOVAs were separately conducted on standardised lip MEP area under the curve (AUC) data, hand MEP AUC data, and active listening score data as the dependent within-subjects variables (4 levels: Clear, +10 dB SNR, 0 dB SNR, -10 dB SNR), with age group as the between-subjects variable. To test our hypotheses, we conducted planned contrasts comparing the control condition (clear) to the three different noise conditions, to investigate whether speech motor processing differed depending on if the noise level was high (-10 dB SNR condition), medium (0 dB SNR condition), or low (+10 dB condition).

3. Results

3.1 Participant characteristics

Several cognitive variables were measured in order to determine whether the older and younger adults demonstrated differences in cognitive and speech processing abilities. Older adults had slightly lower scores on the test measuring cognitive ability, i.e., the MOCA, relative to the younger adults, but this pattern did not reach significance ($F(1,38) = 3.95, p = 0.054$; see Table 1 for participant characteristic data). The younger adults’ short-term memory, as measured using the forwards digit span, was found to be significantly greater in capacity than the older adults’ short-term memory ($F(1,38) = 5.44, p = 0.025$). There was no significant difference between older and younger adults’ working memory capacities, measured using the backwards digit span ($F(1,38) = .975, p = 0.33$).

The results of the pure-tone audiometry averaged across both ears for all the tested frequencies are plotted in Figure 1A. The younger adults had normal thresholds at all tested frequencies. For the older adults, pure-tone thresholds were in the normal range ($\leq 20$ dB HL) across frequencies ranging from 250 Hz-1 kHz but indicated mild hearing impairment at 2 and 4 kHz, and moderate hearing impairment at 8 kHz. As expected, the pure tone average (PTA) across major speech frequencies 500 Hz to 4 kHz differed significantly between older and younger adults ($F(1,38) = 37.84, p < 0.001$). Older adults had significantly less sensitive hearing than the younger adults (higher hearing thresholds), in line with expected age-related hearing loss that occurs as a natural part of healthy ageing. There was an inter-aural PTA difference of 1.5 dB HL for older adults and 0.4 dB HL for younger adults, but the groups were not significantly different ($t(38) = -1.85, p = .34$).

Speech perception skills were also tested by measuring speech reception thresholds (SRTs; see Methods for details). The SRTs for the older adults were significantly higher than the younger adults, indicating that older adults were not able to tolerate as much noise as the younger adults and could not report words as accurately in more intense levels of noise ($F(1,38) = 6.47, p = 0.015$).

Table 1. Participant characteristics of older and younger subject groups. Mean values represent mean raw scores (plus standard error [SE] for demographic variables measured. P-values represent significance of two-tailed independent t-tests used to compare younger and older groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (SE)</th>
<th>p-value</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Older</th>
<th>Younger</th>
<th>P-value</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MOCA (cognitive ability)</strong></td>
<td>26.9 (0.3)</td>
<td>27.8 (0.3)</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td><strong>PTA (hearing sensitivity)</strong></td>
<td>20.7 (2.3)</td>
<td>5.8 (0.9)</td>
<td>&lt;.001*</td>
<td></td>
</tr>
<tr>
<td><strong>Digit span forwards (short-term memory)</strong></td>
<td>6.0 (0.2)</td>
<td>6.7 (0.2)</td>
<td>0.025*</td>
<td></td>
</tr>
<tr>
<td><strong>Digit span backwards (working memory)</strong></td>
<td>4.9 (0.2)</td>
<td>5.3 (0.3)</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td><strong>SRT (speech reception threshold)</strong></td>
<td>-2.6 (0.5)</td>
<td>-4.1 (0.2)</td>
<td>0.017*</td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 1.** Bilateral audiometric threshold averages across all frequencies for older and younger populations. Error bars represent 1 SE.

3.2 TMS behavioural data: Effect of age on active listening

During the TMS conditions, participants were asked to actively listen to the syllables and detect and count how many syllables of a certain type were presented, and to maintain this information in memory until the end of the condition. This was implemented to ensure participants...
were engaging with the stimuli. At the end of each condition, we computed participants’ active listening scores, by subtracting the number of syllables detected by the participant from the number of syllables actually presented in the condition. The active listening scores are shown in Figure 2 as a function of listening condition.

An active listening score of zero would indicate perfect performance (same number of syllables detected and counted as were presented). Any deviation from zero, therefore, represents an error in active listening. Specifically, a positive score would be indicative of stimuli being missed altogether (more stimuli presented than reported), whereas a negative score would be indicative of stimuli being misperceived (more stimuli reported than presented). Data in Figure 2 indicate that active listening ability varied as a function of listening condition in the older adults (black bars) but was reasonably consistent across listening conditions in the younger adults (white bars).

![Figure 2](image)

Figure 2. Mean active listening scores expressed as difference between syllables presented-syllables detected per condition, in both older adults (black bars) and younger adults (white bars). To achieve an active listening score of zero would indicate perfect performance (same number of syllables detected as presented). Any deviation from zero represents an error in syllable detection. Error bars represent 1 SE.

The mixed ANOVA confirmed a significant interaction between active listening scores and age group (F(3,114) = 5.29, p = 0.002). Planned contrasts indicated a significant difference in active listening scores in the Clear condition versus the 0 dB SPL condition (F(1,38) = 4.7; p = 0.04), and the Clear condition versus the -10 dB SNR condition (F(1,38) = 11.4; p = 0.003 [mean score Clear: -0.5 [SE 0.4]; mean score 0 dB SNR: -2.1, [SE 0.6]]; mean score SNR -10 dB SNR: 1.8 [SE 0.4]) in the older adults. There was no significant difference between Clear and +10 dB SNR in the older adults (p > 0.5). There were no significant differences in any planned contrasts in the younger group (all p > 0.4).

In order to evaluate whether the significant interaction between age group and listening scores was driven by hearing acuity, we ran an exploratory analysis that including PTA (see Table 1)
as a covariate. The original study was powered for a 2 x 4 ANOVA, so it is likely that this exploratory analysis may lack sufficient power. The ANCOVA with the PTA covariate added did not eliminate the significant age group x listening score interaction ($p = 0.002$). It also did not impact the main effect of listening condition ($p = 0.002$). There was also no significant interaction with the covariate ($p > 0.1$). Based on these findings, the interaction between age group and listening score appears robust.

These findings indicate that the younger group made a similar (low) number of errors in both the Clear condition and the noise conditions. The older group, however, made significantly more errors in the SNR 0 dB and -10 dB conditions compared to the Clear condition. The poorer performance of the older group in the SNR 0 dB condition suggests that they were more affected by the background noise compared to the younger adults, such that they likely misperceived the stimuli presented (more stimuli were reported than were presented). In the SNR -10 dB condition they were also more adversely affected than the younger adults, and this resulted in stimuli being undetected and thus missed entirely (relative to misperceived as in the SNR 0 dB condition).

Taken together, these results indicate that the active listening task employed during the TMS conditions did measure active listening ability in an age-sensitive manner. Specifically, active listening was adversely affected at lower levels of noise in the older adult group, who first showed decrements in the SNR 0 dB SPL condition.

3.3 TMS lip data: Effect of cognitive load but not ageing on lip MEPs

Figure 3 shows the effect of listening condition on lip MEP data, and indicates that MEPs were inhibited in the 0 dB SNR condition relative to the other conditions. The mixed ANOVA confirmed a significant effect of listening condition on MEP data ($F(3,114) = 3.92; p = 0.01$). Planned contrasts indicated that there was a significant difference between MEPs in the Clear condition and MEPs in the 0 dB SNR condition ($F(1,38) = 8.45, p = 0.006$). Z-scores indicate that MEPs were inhibited (smaller) in the SNR 0 dB SPL condition relative to the Clear condition (SNR 0 dB SPL: -0.4 [SE 0.1]; Clear: 0.2 [SE 0.1]), suggesting that speech motor excitability was reduced in the SNR 0 dB SPL condition compared to the Clear condition. There were no significant differences between the Clear condition and the other noise conditions (SNR +10 dB SPL, $F(1,38) = 0.22, p = 0.64$; SNR -10 dB SPL, $F(1,38) = 0.07, p = 0.79$). The interaction between listening condition and age group was not significant ($F(3,114) = 1.37, p = 0.25$).
Figure 3. Mean lip MEP per condition in both older adults (black) and younger adults (white). MEP AUCs are expressed as z-scores, where a positive score indicates a larger response and a negative score indicates a smaller response, relative to the mean. Error bars represent 1 SE.

3.4 TMS hand data: No effect of cognitive load or ageing on hand MEPs

The control hand MEP data are shown in Figure 4. The mixed ANOVA confirmed that there was no significant effect of listening condition on hand MEP data (F(3,114) = 0.89; p = 0.45). The interaction between listening condition and age group was not significant (F(3,114) = 0.57, p = 0.64).

Taken together, these results indicate that excitability of the lip motor pathway during active speech perception was modulated by listening condition, and this modulation was not affected by age. Notably, the direction of this modulation was not as expected. Lip MEPs did not show evidence of the noise facilitation effect that others have found previously during passive speech perception (Nuttall et al., 2016; 2017). Instead, we found a novel observation that during active listening the noise facilitation effect is absent, and in the SNR 0 dB condition where stimuli were most frequently incorrectly misperceived according to the behavioural data (leading to a negative difference score), lip MEPs are instead inhibited. Importantly, this was not the case in the -10 SNR dB condition, where stimuli were most frequently missed altogether (leading to a positive difference score) and MEPs were not affected. Hand control data did not show evidence of any modulation, suggesting that this effect in the lip data does not represent a domain-general effect of active listening demands on the motor system, but is specific to lip motor cortex.
The present study aimed to establish whether and how speech motor facilitation during speech perception is affected by age. We found that behavioural data from the active listening task involving syllable perception was significantly affected by age. Older adults made more errors in perception at medium noise levels, and missed many stimuli altogether at higher levels of noise, compared to younger adults who performed similarly across all noise levels. Interestingly, in contrast, we found no effect of age on the lip MEP data during perception of the stimuli. We also did not observe the expected noise facilitation effect in the MEP data. Instead, lip MEP data were suppressed in the medium noise condition (0 dB SNR) in both groups. We speculate that this novel and unexpected finding is a consequence of the active listening task that we used to ensure participants were actively engaging with the speech stimuli during MEP recording, as opposed to listening passively. One possibility is that engagement in this active listening task blocked the expected motor facilitation effect in the noise conditions. Furthermore, rather than totally abolish the motor facilitation effect, in the 0 dB condition the lip MEPs were indeed inhibited, showing evidence of a reversal of the predicted noise-facilitation effect.

Data from vision-based automatic imitation experiments shed important light on our finding of inhibited motor facilitation associated with the active listening task. Crucially, visuo-motor experiments find that attention plays a central role in the motor resonances that typically occur in the motor pathways involved in the execution of the action being visually perceived (Cracco et al., 2018; Heyes, 2011). Puglisi and colleagues (2018) note that adding a demanding perceptual secondary task, as we have done by asking participants to 1) selectively detect syllables and 2) count and maintain the total number detected in memory, can have a profound influence on the coding of the perceived action, resulting in the modification (Catmur, 2016) or even in the elimination (Bach, Peatfield, & Tipper, 2007; Chong, Cunnington, Williams, & Mattingley, 2009; Gowen, Bradshaw,
Galpin, Lawrence, & Poliakoff, 2010) of the resulting motor resonances. Our data seem to be consistent with this explanation. The active listening requirement of our task, which was used to ensure participants engaged with the stimuli, effectively added additional attentional demands during speech perception that modulated speech motor resonances.

Relatedly, neuroimaging studies have shown that facilitation of motor cortex during action observation is not an automatic event, but instead the effect scales relative to what attentional resources are available to the observer at the time of perception (Bach et al., 2007; Chong et al., 2009; Muthukumaraswamy & Singh, 2008; Perry & Bentin, 2010; Woodruff & Klein, 2013). This highlights the possibility that motor resonances do not deploy in a binary fashion. Such a possibility has been observed in facilitated (larger) speech MEPs during perception of speech in noise relative to speech without noise, during passive speech perception. Whilst this may be the case during passive speech perception, as has been repeatedly demonstrated, the elimination and suppression of the motor facilitation effect during active speech processing underscores the complexity of the motor resonance response. Indeed, data from the present study indicate that motor resonances can be modulated by top-down influences associated with task demands. We did not include a non-speech condition in our design, and therefore cannot verify the speech facilitation effect by comparing speech data to non-speech data. However, the notion of speech motor facilitation is well-documented in previous studies across several labs (Bartoli et al., 2015; Murakami, Kell, Restle, Ugawa, & Ziemann, 2015; Nuttall et al., 2016; Nuttall, Kennedy-Higgins, Devlin, & Adank, 2018), and thus the inclusion of a non-speech condition is not essential for the interpretation we are proposing here.

One interpretation for these results is that attention modulates early auditory processing, which affects subsequent later sensorimotor integration and resulted in the modulatory effect of the present study’s task demands on speech motor resonances during speech processing. Indeed, electrophysiological evidence from auditory cortex has shown that selective attention decreased neural responses to distractors relative to auditory targets (Schwartz & David, 2018). This suppression enhanced neural target detection thresholds, suggesting that limited attention resources serve to focally suppress responses to distractors that interfere with target detection. It would therefore be interesting to know if the motor responses to the syllables that participants had been told to selectively attend to were enhanced relative to the other distractor syllables. However, as the present experiment was not designed to specifically test the effect of selective attention on the speech motor system, this analysis is not possible as there are too few selectively attended trials within each condition to perform any meaningful analysis. It is an important goal of our future work, however, to explore this effect, and to test how it scales with background noise.

It is also interesting to note that the MEP inhibition effect is present specifically in the 0 dB SNR condition, and not in the most acoustically adverse -10 dB SNR condition. The behavioural data may offer some insight into this pattern in the results. Active listening scores in the 0 dB SNR condition were less accurate and led to a negative score for older adults especially, indicating that a greater number of syllables were reported as detected than were actually presented. This performance was in contrast to the -10 dB SNR condition, where active listening scores were also less accurate particularly for the older adults, but the scores were positive, indicating that a smaller number of syllables were reported as detected than were actually presented. It is possible that the noise level in the -10 dB SNR was simply too adverse for the older participants to perform the task. Hence, the stimuli were not sufficiently detectable to show any evidence of the neurophysiological response being modulated by the divided attentional demands of the task, as it is not clear that the task was performed as intended, as many of the stimuli were missed on average. Our behavioural
results reflect previous findings of speech perception ability in adult listeners, who often perform at floor when tested on speech recognition in -10 dB SNR (Holder, Levin, & Gifford, 2018). With respect to the original hypotheses that this paper set out to address, however, we did not find any differences in speech motor excitability between younger and older adults during active speech perception. This is in contrast to findings from Du et al (2016) and Panouilleres et al. (2018). However, absence of evidence does not imply evidence of absence, and it is possible that the conditions we employed, such as divided attentional demands, obscure any ageing differences. Hence, we cannot conclude from these data alone that there are no differences in speech motor activity between younger and older adults whatsoever. Panouilleres et al’s (2018) TMS paradigm did not require participants to use any active listening strategies, and Du et al’s (2016) fMRI study used selective but not divided attention. Importantly, the lack of a neurophysiological ageing effect is not due to a lack of task sensitivity, nor a lack of sensory decline in the ageing group: task performance was affected by age as older adults’ task performance was more negatively affected by less adverse SNRs compared to younger adults. Additionally, older adults’ hearing was less sensitive compared to younger participants. The conflation of attentional resources required to perform the task here may have reduced what compensatory motor resources were available to draw upon, ultimately obscuring any group differences that may exist. Further studies that manipulate attentional load and task difficulty are warranted to test this possibility. It should be acknowledged that it is a limitation of the current study that we did not systematically investigate the role of hearing and its impact on the experimental MEP manipulation. Future work should evaluate the contribution of hearing, along with other factors known to be affected by ageing, on cognition or speech perception. This study used a mixed ANOVA approach to investigate the effect of age group and listening condition on the dependent variables. For demographic characterisation and transparency, we collected data on our sample with regards to common variables that differ as a function of age. A larger study that employs, for example, a linear mixed effects approach could model how these variables interact with age and listening condition, which would build on the present findings and provide useful insights into the relationships explored in the present study.

In summary, whilst we found evidence to indicate that ageing reduces cognitive and sensory functioning, we did not find evidence to suggest that the speech motor cortex compensates for this reduced function. Interestingly, however, we did observe the surprising and unexpected finding that speech motor cortex is modulated by task demands, i.e., the active listening requirement of the task used to control participants’ engagement with stimuli. Speech perception often occurs in the context of competing information, whereby the brain must strategically allocate resources to multiple tasks simultaneously (Gennari, Millman, Hymers, & Mattys, 2018). How the neurobiological network for speech perception controls resource allocation during difficult listening conditions is not well-understood. Our data shed new light on this topic and indicate that auditory-motor modulation adapts based on the attentional resources available. Future work should seek to clarify the relationship between sensorimotor integration and attention and extend this to concurrent auditory and visual processing to better understand the role of the speech motor system in resource allocation during speech perception.

References


