

Highlights

- Activity in the lip area of M1 increases when listening to distorted speech.
- Activity is further increased if the distorted speech sounds are lip-articulated.
- The ability to correctly perceive speech relates to excitability in the lip area.
- Activity in M1 lip area associates with individual hearing sensitivity.
- The motor system supports speech perception under challenging listening conditions.

Title: The effect of speech distortion on the excitability of articulatory motor cortex

Helen E. Nuttall¹, Daniel Kennedy-Higgins¹, John Hogan², Joseph T. Devlin², Patti Adank¹

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

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

Correspondence should be addressed to Helen Nuttall, Department of Speech, Hearing and Phonetic Sciences, University College London, Chandler House, 2 Wakefield Street, London, WC1N 1PF, UK. E-mail: h.nuttall@ucl.ac.uk. Tel: +44207 679 4235.

Abstract

It has become increasingly evident that human motor circuits are active during speech perception. However, the conditions under which the motor system modulates speech perception are not clear. Two prominent accounts make distinct predictions for how listening to speech engages speech motor representations. The first account suggests that the motor system is most strongly activated when observing familiar actions (Pickering and Garrod, 2013). Conversely, Wilson and Knoblich's account asserts that motor excitability is greatest when observing less familiar, ambiguous actions (Wilson and Knoblich, 2005). We investigated these predictions using transcranial magnetic stimulation (TMS). Stimulation of the lip and hand representations in the left primary motor cortex elicited motor evoked potentials (MEPs) indexing the excitability of the underlying motor representation. MEPs for lip, but not for hand, were larger during perception of distorted speech produced using a tongue depressor, relative to naturally produced speech. Additional somatotopic facilitation yielded significantly larger MEPs during perception of lip-articulated distorted speech sounds relative to distorted tongue-articulated sounds. Critically, there was a positive correlation between MEP size and the perception of distorted speech sounds. These findings were consistent with predictions made by Wilson & Knoblich (Wilson and Knoblich, 2005), and provide direct evidence of increased motor excitability when speech perception is difficult.

Key words: Speech perception, motor cortex, transcranial magnetic stimulation, motor evoked potentials.

1. Introduction

Human listeners are adept at perceiving speech in a variety of listening conditions, which can differ widely in the difficulty they pose to the listener. Indeed, most of us will have engaged in a conversation affected by a poor telephone connection, an unfamiliar speech style, or a distracting discussion taking place nearby, but despite such limitations we remain remarkably good at extracting the meaning from our interlocutor's speech. It is surprising, therefore, that the neural architecture underlying this success remains little understood. Current models outlining the neural organization of speech processing (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009) propose that the locus of intelligible speech understanding is the temporal lobe within the ventral stream of speech processing. However, the neural pathway of the ventral stream differs in these two models; Rauschecker & Scott suggest that speech processing has its center of gravity in left anterior STS (Superior Temporal Sulcus), while Hickok & Poeppel propose that recognising intelligible speech is bilaterally organized and located both anteriorly and posteriorly to Heschl's Gyrus. Both models also feature a dorsal stream, which is thought to translate acoustic speech signals into articulatory representations for speech production, though the models also posit differences in the integration of dorsal stream function (for details see Hickok and Poeppel 2007; Rauschecker and Scott 2009, and Turkeltaub and Coslett 2010; Adank et al. 2012 for more in-depth reviews).

Evidence suggests that temporal areas in the ventral stream form part of a functional hierarchy for speech processing, where primary auditory cortex is sensitive to the

acoustic features of speech, but higher-order temporal, and frontal, sites (middle and superior temporal gyri, left inferior frontal gyrus) are sensitive to the intelligibility of speech, but insensitive to the acoustic form of the stimuli (Davis and Johnsrude, 2003). Accordingly, it has been demonstrated that when listening to speech in challenging conditions, activity increases in peri-auditory and frontal regions relative to when listening to intelligible speech, and it is thought that such activity may support processing in primary auditory areas and help compensate for the acoustic distortion (Davis and Johnsrude, 2007, 2003; Shahin et al., 2009; Wild et al., 2012a). Concurrently, Wild and colleagues (Wild et al., 2012b) have recently shown that motor regions and left inferior frontal gyrus exhibit elevated responses when attending to degraded speech in the presence of auditory distractors. As such, it is becoming increasingly apparent that speech perception may also involve areas beyond those classic temporal sites already identified.

In particular, recent years have seen renewed interest in the idea that cortical motor systems involved in producing speech may also contribute to perceiving it. Originally proposed by Liberman (Liberman and Mattingly, 1985; Liberman et al., 1967), the Motor Theory of Speech Perception was roundly criticized for its claim that motor cortex, rather than auditory cortex, was the key site for speech comprehension (Diehl et al., 2004; Jusczyk et al., 1981; Scott et al., 2009). Nonetheless, accumulating evidence from Transcranial Magnetic Stimulation studies (TMS) suggests that regions of primary motor cortex (M1), important for accurate control of articulatory gestures, activate during speech comprehension, and are also involved in the precise categorization of complex acoustic signals (D'Ausilio et al. 2009; Möttönen and Watkins 2009; Sato et al. 2010).

These TMS findings resonate with functional Magnetic Resonance Imaging (fMRI) observations of motor cortex activation during speech perception (Osnes et al. 2011; Hervais-Adelman et al. 2012; Szenkovits et al., 2012). Furthermore, this activation is modulated in a somatotopic way, whereby speech articulators in left motor cortex are more responsive when listening to speech produced using the same articulator, compared to when listening to speech produced using a different articulator. Indeed, Fadiga and colleagues (Fadiga et al., 2002) demonstrated that passively listening to words that involve tongue articulation results in the automatic facilitation of the tongue region in primary motor cortex. Such facilitation does not result in overt movement generation, but can be observed as changes in the potentiation of the tongue muscle resulting from increased corticobulbar excitation.

The precise contribution of this observed motor activation, however, is under active debate, and the field is still divided in opinion as to whether the articulatory motor system is essential for speech comprehension. What has been acknowledged is that the motor system may have a modulatory influence on perceptual systems (Hickok et al., 2011). However, the conditions under which the motor system has cause to modulate audition are not clear. Evidence from fMRI argues for a preferential engagement of the motor system when listening to speech that is difficult to understand. In a recent study, Du and colleagues (Du et al., 2014) tested the hypothesis that motor activation contributes to categorical speech perception under adverse, but not quiet, listening conditions. The authors observed a negative correlation between neural activity and perceptual accuracy in left premotor cortex, which contributed to phoneme categorization specifically at moderate-to-adverse signal-to-noise ratios. Using TMS and a highly similar phoneme

categorization tasks to assess categorical perception, Möttönen and Watkins (Möttönen and Watkins, 2009) demonstrated that primary motor cortex makes a significant contribution to phoneme judgments in quiet listening conditions, though specifically at the ambiguous phonetic category boundary. As such, evidence corroborating the activation of motor processes during speech perception is compelling, but our understanding of the listening conditions that preferentially engage the speech motor system is uncertain.

One prominent interpretation of motor activation during perception is provided by motor simulation accounts (Pickering and Garrod, 2013a; Wilson and Knoblich, 2005). These accounts posit that perception of another person's actions results in activation of the corresponding motor plan in the perceiver, leading to covert motor simulation. Although motor simulation accounts are not based specifically on speech, speech production is a form of motor activity and thus these accounts are equally relevant to the processing of speech actions. Under this account, the articulatory plans stored in the speech motor system for production are automatically activated during speech perception, although this activation does not result in overt articulation due to presumed suppression of activity in the subcortical motor system (Baldissera et al., 2001). These motor plans are then used to inform forward models of upcoming articulatory gestures in the incoming speech stream.

However, two different forms of the motor simulation account make dissociable predictions about how the perceiver's motor activity is modulated during action perception. The first proposes greater motor involvement when the observer is familiar

with the perceived action (Pickering and Garrod, 2013a), for instance when observing actions that the observer can also perform or easily understand. Indeed, Calvo-Merino and colleagues (Calvo-Merino et al., 2005) found greater bilateral activation in motor areas when expert dancers viewed movements that they had been trained to perform compared to movements they had not, indicating that the action-observation system integrates observed actions with the motor repertoire of the observer. In line with this possibility, Swaminathan and colleagues (Swaminathan et al., 2013) found that when subjects observed visual speech movements from a known language, motor excitability in the lip area of M1 was higher than when subjects observed speech movements from an unknown language. The authors interpreted these results to suggest that activity in articulatory motor cortex is enhanced when perceiving speech movements that the perceiver is already experienced in producing and perceiving themselves. Similarly, Bartoli and colleagues observed that the effect of TMS to speech motor areas was related to the listener-speaker perceived acoustic distance, such that response times were facilitated for smaller acoustic distances (Bartoli et al., 2015). By this account, activity in M1 speech areas should be greatest when listening to familiar, natural speech, relative to less familiar, motor-perturbed distorted speech, which is difficult to understand, suggesting somatotopic differences would also be most distinct during perception of natural, unperturbed, speech. Under this account, comprehension of degraded/perturbed speech is assumed to be subserved by increased utilization of auditory, but not motor, resources (Pickering and Garrod, 2013a). Notably, Pickering and Garrod also claim that

motor simulation and prediction-by-association, driven by co-occurrence in the auditory input, can also be combined.

In contrast to this view, the second account claims that the motor system is most strongly activated when perception is challenging, predicting greater involvement of M1 speech areas under difficult listening conditions (Wilson and Knoblich, 2005). Under this account, although challenging perceptual conditions would catalyse greater motor activation, the success of the resultant predictive signaling would depend on the degree of similarity between what the observer can perform motorically, and what is being perceived. In turn, this would suggest that articulator-specific effects would be maximally dissociable in terms of M1-activation when listening is difficult. Indeed, TMS in combination with motor evoked potentials (MEPs) has been found to suggest increased motor processing when perceiving spoken sentences in noise (Murakami et al., 2011), although importantly the effect of speech-internal distortion, and somatotopic responsiveness, are unknown. In addition, it has also been shown that TMS to motor areas can significantly affect accuracy (Meister et al., 2007) and response times for speech stimuli in noise (D'Ausilio et al., 2009), but not for speech presented without noise (D'Ausilio et al., 2012). These data support the latter version of the simulation account, and suggest that speech motor activity may be necessary when comprehending speech that is degraded.

In the present study, we aimed to disambiguate between these two accounts by using TMS to elicit a direct measure of motor excitability during speech perception. To this end, motor evoked potentials (MEPs) were elicited during perception of natural

speech, and speech distorted via a lip and tongue-perturbation during production. Stimulation was thus used to probe the excitability of M1 lip muscle representation to determine whether activation was greater when listening to normal versus distorted speech sounds. In addition, by using speech sounds with two different places of articulation, we tested whether somatotopic facilitation enhanced MEPs in line with predictions made by motor simulation accounts. Lastly, we explored the relationship between individual ability in perceiving distorted speech and motor system activity, to assess the extent to which motor activity is associated with listening performance.

2. Methods

2.1 Subjects

Eighteen subjects took part in this study (six males; average age: 23 years 9 months (\pm SD 3.5 months); age range: 19–30 years). However, data from two subjects were excluded from the analysis; one due to MEP variability and one due to hand MEPs that exceeded 2 SDs from the overall group mean. All subjects were monolingual, native speakers of British English, with normal language function and 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 500, 1000, 2000 and 4000 Hz, bilaterally. All subjects had normal thresholds (≤ 20 dB HL) and

presented no TMS contraindications as assessed by the University College London TMS safety screening form. Subjects 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 currently studying at University level. Experiments were undertaken with the understanding and written consent of each subject, according to Research Ethics Board of University College London.

2.2 Experimental design

All subjects received two blocks of single-pulse TMS to the lip area of M1 in the left hemisphere, as well as two blocks of TMS to the hand area of left M1 as a control site, whilst listening to speech which was either naturally-articulated or distorted (Figure 1). The hand area was chosen as a control site due to its close proximity on the cerebral cortex to M1 lip area, and because previous literature has demonstrated that speech manipulations do not affect hand muscle potentiation (Möttönen and Watkins, 2009; Watkins et al., 2003). At the same time as receiving TMS, subjects were presented with speech stimuli in the following conditions:

(1) Natural Speech: listening to normally articulated vowel-consonant-vowel (VCV) syllables produced in a natural manner, which contained an equal distribution of lip- (/apa/, /aba/) or tongue-articulated (/ata/, /ada/) consonants.

(2) Distorted Speech: listening to motor-distorted VCV syllables obtained by using a

tongue depressor, which obstructed movement of the lips and tongue. Equal numbers of tongue- and lip-articulated stimuli were also used. The tongue depressor was a flat wooden spatula with rounded ends, and was five inches long and one inch wide.

The order of TMS sites and experimental conditions were counterbalanced across participants. During the TMS, subjects were instructed to listen passively to the speech stimuli. After the TMS session, subjects also completed a short identification task to assess their ability to correctly identify the consonant in the distorted VCV syllables.

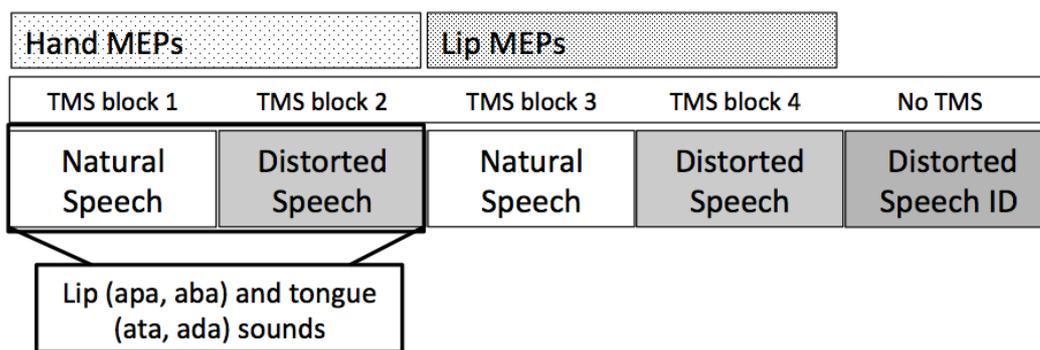


Figure 1. Experimental design. The order in which the M1 lip and hand TMS sites were stimulated and speech types were presented was counter-balanced across subjects. The speech identification task was always completed after the TMS sessions.

Notably, the use of a tongue depressor was intended to impair perception of both tongue-, and lip-articulated sounds due to the tongue depressor restricting both the tongue

and lower lip movement. A tongue depressor was specifically chosen so as to introduce a motor-based distortion into the speech signal, as opposed to an auditory distortion, to relate the speech perception challenge to a speech production difficulty. In a separate pre-test using a different set of subjects, we found that tongue-depressed speech could be identified above chance level (average = 50.8% correct; chance = 25%). Subjects were much less accurate and much slower to respond to the distorted speech compared to the normally articulated, natural speech produced by the same speaker, for which accuracy was at ceiling (see Supplementary Materials for further details of the pre-test). We also found that perception (measured using accuracy and response time) of distorted lip- and distorted tongue-articulated sounds were equally affected by the tongue-depressor manipulation (see Supplementary Materials).

2.3 Speech stimuli

All natural and distorted speech stimuli were recorded by a twenty-seven year old female British English speaker in a sound-attenuated room. Both stimulus types were produced to be approximately the same duration (mean 975.25 ms) but were not synthetically manipulated to be precisely the same length. Stimuli varied by a standard deviation 60.77 ms. Audio digitizing was performed at 44.1 kHz with 16 bits. All syllables were root-mean-square normalized offline using Praat (Boersma, 2001), and then presented using Matlab (R2013a; The Mathworks Inc., Natick, MA) through ultra-

shielded insert earphones (ER-3A; Etymotic Research, Inc., IL), at a comfortable level of approximately 65 dB SPL. For each subject, in each condition, a stimulus list containing five occurrences of /apa/, /aba/, /ata/ and /ada/ stimuli was randomly permuted, and stimuli were presented according to this order for six blocks without cessation. The same stimuli were used for each subject but presented in a different order. This yielded 120 stimulus trials (30 /apa/, 30 /aba/, 30 /ata/ and 30 /ada/ stimuli) for each speech condition (natural and distorted), at each TMS site (lip, hand), resulting in the presentation of 480 stimuli altogether during the TMS session.

2.4 Speech identification task

At the end of the experiment, subjects completed a speech identification task to assess their ability to perceive the distorted speech stimuli. This task was presented on Matlab using custom-written software. Subjects were asked to listen carefully to the distorted speech sounds and to identify the consonant in the middle of the sound as either a 'p', 'b', 't' or 'd' using a key press, as quickly as possible without compromising accuracy. After the stimulus had finished playing, subjects were prompted with a visual cue to enter their response. Subjects were given up to 2000 ms from the onset of the stimulus to make their response, after which the program would present the subsequent trial. Failure to respond during this time period would result in a null response for that particular trial.

2.5 Transcranial magnetic stimulation

Monophasic TMS pulses were generated by a Magstim 200² unit and delivered by a 70mm diameter figure-eight coil, connected through a BiStim² module (Magstim, Dyfed, UK). The coil was placed tangential to the skull such that the induced current flowed from posterior to anterior under the junction of the two wings of the figure-eight coil. The lip area and hand area of M1 were found by using the functional ‘hot spot’ localization method, whereby application of TMS elicits an MEP from the contralateral muscle. Here, the coil position and orientation is adjusted in millimeter movements to ascertain the location on the motor cortex 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 50 μ V (Rossini et al., 1994; Watkins et al. 2003). In this way, we first located the hand area (and associated hand aMT if TMS was to be applied to this site first), and then by moving the coil ventrally and slightly anterior until an MEP was observed in the contralateral lip muscle, identified the lip area ‘hot spot’ and its aMT. The intensity of the stimulator was then set to 120% of each area’s respective aMT for the stimulations applied during the experiment. The mean stimulator intensity (120% aMT \pm SEM) used to elicit lip MEPs in the TMS experiment was 50.6% (\pm 1.9), and 46.4% (\pm 2.0) for hand MEPs. During the presentation of each speech stimulus, Matlab was used to externally trigger the TMS system, which generated a TMS pulse 100 ms

after the onset of the consonant in each stimulus type. All 120 speech stimuli were accompanied by a TMS pulse; therefore, there were no no-TMS trials. There was a 3000 ms inter-stimulus delay between all auditory stimuli and an inter-stimulation delay of between 4500-5000 ms, which included the length of the auditory stimuli as well as the inter-stimulus delay time in between presentations of consecutive auditory stimuli. TMS blocks lasted for approximately 9-10 minutes. Participants were given short breaks in between TMS blocks, during which time the coil was changed to prevent over-heating.

2.6 Electromyography

Electromyographic (EMG) activity was recorded from lip and hand areas using surface electrodes (Ag/AgCl; 10-mm diameter) in a non-Faraday caged, double-walled sound-attenuating booth. For the lips, 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 5 minutes to produce a constant level of contraction (approximately 20% of maximum voluntary contraction) of the lip muscles by pursing, which was verified via visual feedback of the ongoing EMG signal. For the recording of hand EMG, electrodes were attached in a tendon-belly montage with the active electrode placed on the right first dorsal interosseus, the reference electrode on the tendon of the same muscle, and a ground electrode on the wrist. Subjects were also trained to maintain

a constant level of contraction of this muscle during the experimental recordings. Contraction of the lip and hand muscles 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), and analog-to-digital converted using a Micro1401-3 unit (Cambridge Electronic Design, Cambridge). Continuous data were acquired and recorded using Spike2 software (version 8, Cambridge Electronic Design, Cambridge).

2.7 Data analysis

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 mean MEPs were calculated for each TMS site, speech type, and lip or tongue sound combination, in each subject. Individual averages were rectified and the integrated area under the curve (AUC) of the rectified EMG signal of each individual mean MEP was calculated. For lip MEPs, AUC was computed from 8-35 ms, and for hand MEPs 13-40 ms, post-TMS. An earlier window was used for the lip AUC calculations due to the earlier onset latencies for lip MEPs compared to hand. The average height of the pre-TMS baseline EMG activity was also computed 10 ms before

the TMS, and no significant differences were observed between levels of baseline activity between conditions (Lip $t(15) = 0.45$, $p = 0.7$ [mean AUC in the natural condition was $0.16 \text{ mV}\cdot\text{ms}$ (S.D. 0.06), and $0.17 \text{ mV}\cdot\text{ms}$ (S.D. 0.06) in the distorted condition]; Hand: $t(15) = -0.10$, $p = 0.9$ [mean AUC in the natural condition was $0.23 \text{ mV}\cdot\text{ms}$ (S.D. 0.07), and $0.23 \text{ mV}\cdot\text{ms}$ (S.D. 0.07) in the distorted condition]). The overall mean and standard deviations of the AUC across conditions was also calculated in hand and lip MEPs separately. As expected, lip MEPs had lower amplitudes than hand MEPs, thus to enable comparison, the AUC for each condition was standardized to the overall mean of MEPs averaged across each TMS site, yielding a z-score for each condition.

To assess speech identification ability on the distorted speech task, percent correct identification performance was calculated. For peripheral hearing sensitivity, pure-tone average (PTA) audiometric thresholds were computed across all frequencies measured (500, 1000, 2000 and 4000 Hz) for each ear. The better ear PTA was used in further analyses, and was defined as the ear with the lower PTA (lower PTAs indicate more sensitive peripheral hearing). Raw hearing threshold data were unavailable for two subjects, so these data points were replaced with the mean.

Statistical analyses were carried out using SPSS (version 22.0, IBM). A three-way repeated measures analysis of variance (RM ANOVA) was conducted on the MEP z-scores, with TMS Site (Hand, Lip), Speech Type (Natural, Distorted), and Articulator (Tongue, Lip), as within-subjects factors. Additional two-way RM ANOVAs were conducted separately for Lip and Hand data, with Speech Type and Articulator as within-

subjects factors. Two-tailed paired t-tests and Pearson's correlations were also conducted where appropriate.

3. Results

3.1 Effect of natural and motor-distorted speech on MEPs elicited during speech perception

This experiment used TMS to elicit MEPs to test how different types of speech modulate the reactivity of the motor system during speech perception. Figure 2 shows the grand average lip (Figure 2A) and hand (Figure 2B) MEPs elicited during perception of natural (grey) and distorted speech (black). A visible amplitude difference can be observed between grand average lip MEPs, where the distorted average contains both higher peaks and lower troughs in the MEP complex, indicating greater peak-to-peak amplitude. The hand data do not show any consistent, discernible differences.

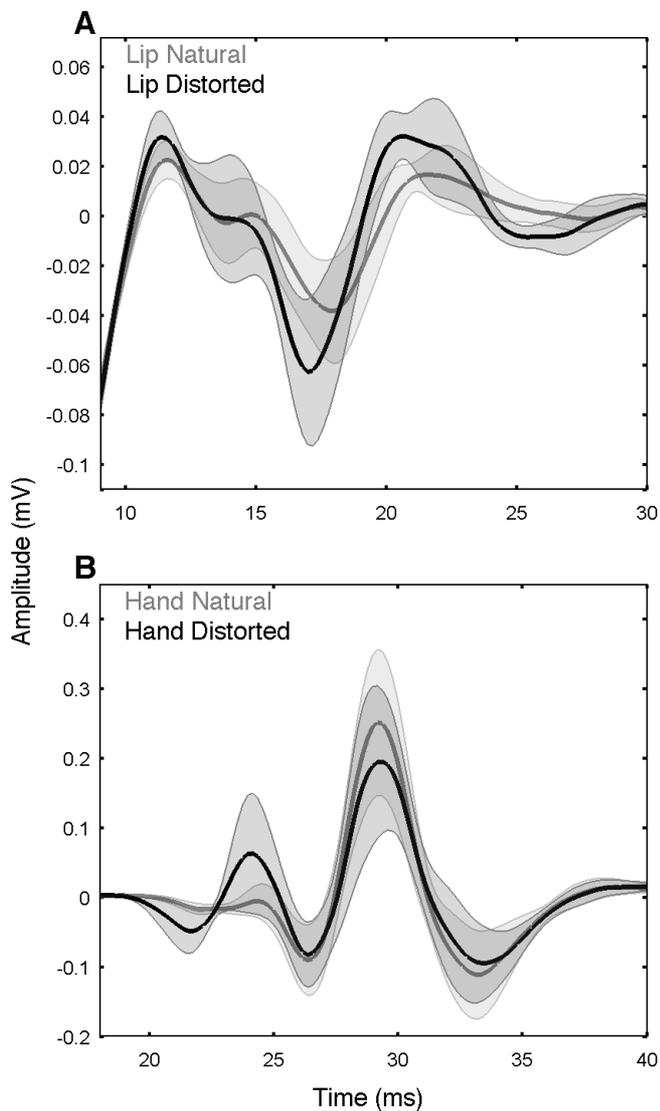


Figure 2. Grand average lip (A) and hand (B) MEPs elicited during perception of natural speech (grey line) and distorted speech (black line) \pm SEM (grey and black shading, respectively).

The difference in lip MEP size as a function of speech type was also reflected in the standardized MEP AUC z-scores (Figure 3A). Here, average AUC z-scores for lip MEPs in the distorted speech condition were greater than those in the natural condition

(Figure 3A). The greatest AUC scores were observed for MEPs elicited during perception of distorted speech articulated by the lips. No differences between speech type or articulator were evident in the hand data (Figure 3B). These observations were confirmed by a three-way RM ANOVA, which indicated a significant three-way TMS site x speech type x articulator interaction ($F(1,15) = 9.77$; $p = 0.007$; $\eta_2 = 0.39$), suggesting that the effect of speech type on lip MEPs, but not hand MEPs, was further modulated by articulator. There was also a significant main effect of speech type ($F(1,15) = 4.96$, $p = 0.046$; $\eta_2 = 0.25$). All other main effects and interactions were non-significant (TMS site: $F(1,15) = 0.00$, $p = 1.00$; articulator: $F(1,15) = 2.50$, $p = 0.14$; TMS site x speech type: $F(1,15) = 3.81$, $p = 0.07$; TMS site x articulator: $F(1,15) = 0.01$, $p = 0.95$; speech type x articulator: $F(1,15) = 0.34$, $p = 0.57$).

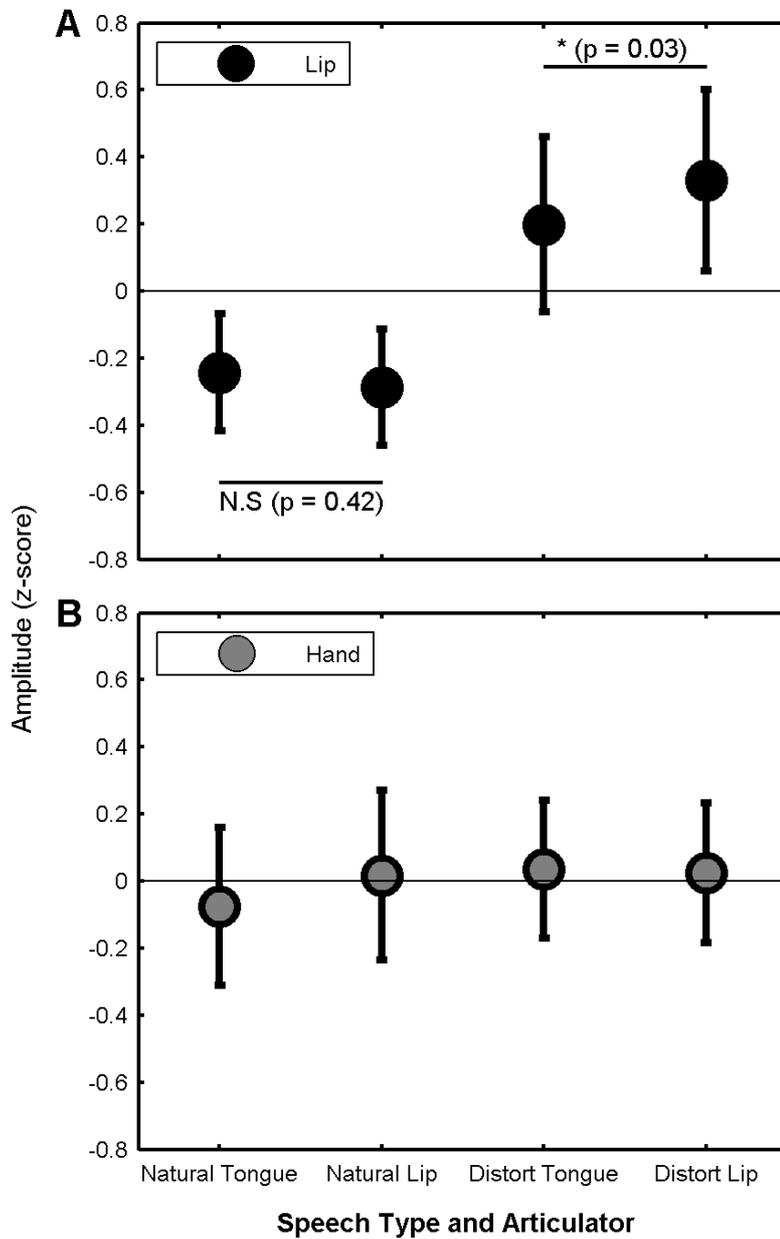


Figure 3. Mean standardized AUC \pm SEM of MEPs measured from the lip (A) and hand sites (B), during the perception of natural and distorted speech articulated by either the tongue or lips.

To explore the three-way interaction, two two-way RM ANOVAs were conducted for the lip data and the hand control data separately. For the lip data, there was a significant main effect of speech type ($F(1,15) = 9.25$; $p = 0.008$; $\eta_2 = 0.38$) indicating that lip MEPs were significantly larger for distorted relative to natural speech. In addition, a significant interaction between speech type and articulator ($F(1,15) = 4.73$; $p = 0.046$; $\eta_2 = 0.24$) indicated that this effect was greatest for lip-articulated speech sounds, and less strong for tongue-articulated stimuli. In contrast, there was no significant main effects or interaction present for the hand data (speech type x articulator: $F(1,15) = 1.96$, $p = 0.18$; speech type: $F(1,15) = 0.11$, $p = 0.75$; articulator: $F(1,15) = 1.14$, $p = 0.30$). These results suggest that speech type had no effect on hand control data, but did have a significant effect on lip MEP size, the size of which effect varied as a function of articulator. This confirmed the observation of greater lip MEPs during perception of distorted speech, as seen in Figures 2A and 3A.

The significant two-way interaction observed for the lip MEP data indicated that articulator modulated the extent of the effect of speech type. To assess this relationship, two-tailed paired t-tests were conducted to test the difference in MEP size during perception of natural speech articulated by the lips or tongue, and between MEPs recorded during perception of distorted speech articulated by the lips or tongue. There was no significant effect of articulator on MEPs during perception of natural speech ($t(16) = 0.83$; $p = 0.42$; Cohen's $d = 0.20$). However, for distorted speech, MEPs elicited during perception of lip and tongue articulated sounds were significantly different ($t(16)$

= -2.29; $p = 0.037$; Cohen's $d = 0.58$), such that MEPs to lip sounds were significantly greater than for tongue sounds (0.34 mV·ms, SD = 0.59).

3.2 Relationship between motor activation and the ability to perceive distorted speech

To assess the relationship between identification of distorted speech and levels of activity in the motor system during speech perception, we tested subjects' identification of distorted speech sounds at the end of the experiment (see Methods section 2.1). Identification scores ranged from 20% to 80% correct (chance performance = 25%) and subjects were able to correctly identify 47.5 % (SD = 19.6) of stimuli on average (average identification of individual syllables: /aba/ = 52.5 ± 33.4 , /ada/ = 42.5 ± 27.0 , /apa/ = 45.0 ± 23.4 , /ata/ = 48.1 ± 28.6). Figure 4 shows the relationship between lip motor activation during natural and distorted speech perception, and identification accuracy of distorted stimuli. Here, we collapsed across both stimulus types as behavioural data obtained in the pre-test suggested that both stimulus types were perceived with equal difficulty (see Supplementary Material for further details). We found that subjects who were better at identifying the distorted stimuli also had greater lip MEPs (Figure 4A and B). Pearson's correlations confirmed this relationship was significant both for MEPs recorded from the lips during perception of natural speech ($r(16) = .65$, $p = 0.007$; Figure 3A), and for lip MEPs recorded during perception of distorted speech ($r(16) = .72$, $p = 0.002$; Figure 3B). Notably, the same association was not observed between speech

perception skills and hand MEPs (natural speech: $r(16) = -.10$, $p = 0.71$; distorted speech: $r(16) = .02$, $p = .95$), confirming that the relationship is specific to the lip representation.

To assess the extent of motor system facilitation during distorted speech perception, we computed the MEP AUC difference score. This score represents the increase in motor activation from natural to distorted speech perception conditions, and is derived by subtracting MEPs recorded during the natural speech perception condition from MEPs recorded during distorted speech perception. Greater scores represent a larger difference between MEP areas in the natural and distorted perception conditions, indicating greater motor facilitation under distorted speech perception, and lower scores vice versa. A negative score would indicate that subjects demonstrated greater MEPs during perception of natural speech compared to MEPs recorded during distorted speech perception. These difference scores are represented along the x-axis of Figure 4C, from which it can be observed that the majority ($n=13$) of subjects demonstrated a greater amount of motor system activity when perceiving distorted speech relative to perceiving natural speech. Here, it can also be observed that subjects who demonstrated greater amounts of facilitation also had higher distorted speech identification scores ($r(16) = .53$, $p = 0.035$; Figure 4C).

It is possible, however, that the relationship between speech identification performance and level of motor cortex activation may be mediated by a third variable, such as listeners' hearing sensitivity. This was measured using PTA at the start of the experiment to ensure all subjects were normally hearing (see Methods section 2.1). PTAs for listeners' best ear ranged from -5 to 18.75 dB HL (mean: 2.7 dB HL, S.D. 5.7). Partial

correlations, controlling for the effect of PTA threshold for the best ear, indicated that unique variance remained between speech identification scores and MEPs elicited during perception of natural ($r(13) = .59, p = 0.022$) and distorted speech ($r(13) = .66, p = 0.008$), but after eliminating the influence of the hearing threshold, the relationship between speech perception and the MEP difference score was no longer significant ($r(13) = .42, p = .12$). These data suggest that individuals who show greater increases in motor activation during perception of distorted speech compared to when perceiving natural speech (heightened facilitation), are better at identifying distorted speech sounds, but that this relationship reflects a shared dependency of both variables upon peripheral hearing sensitivity.

To directly test possible relationships between individual MEP difference scores and PTAs, and speech perception scores and PTAs, additional correlations were performed. This confirmed a significant positive association between MEP difference scores and PTA ($r(16) = .54, p = .03$; Figure 4D). The relationship between PTA and speech perception was not significant ($r(16) = .37, p = .15$). The positive correlation between the MEP difference score and PTA suggests that individuals with higher PTAs, indicative of less sensitive hearing, demonstrate greater motor facilitation when listening to distorted speech relative to clear speech, whereas listeners with good hearing (lower PTAs) demonstrate less facilitation.

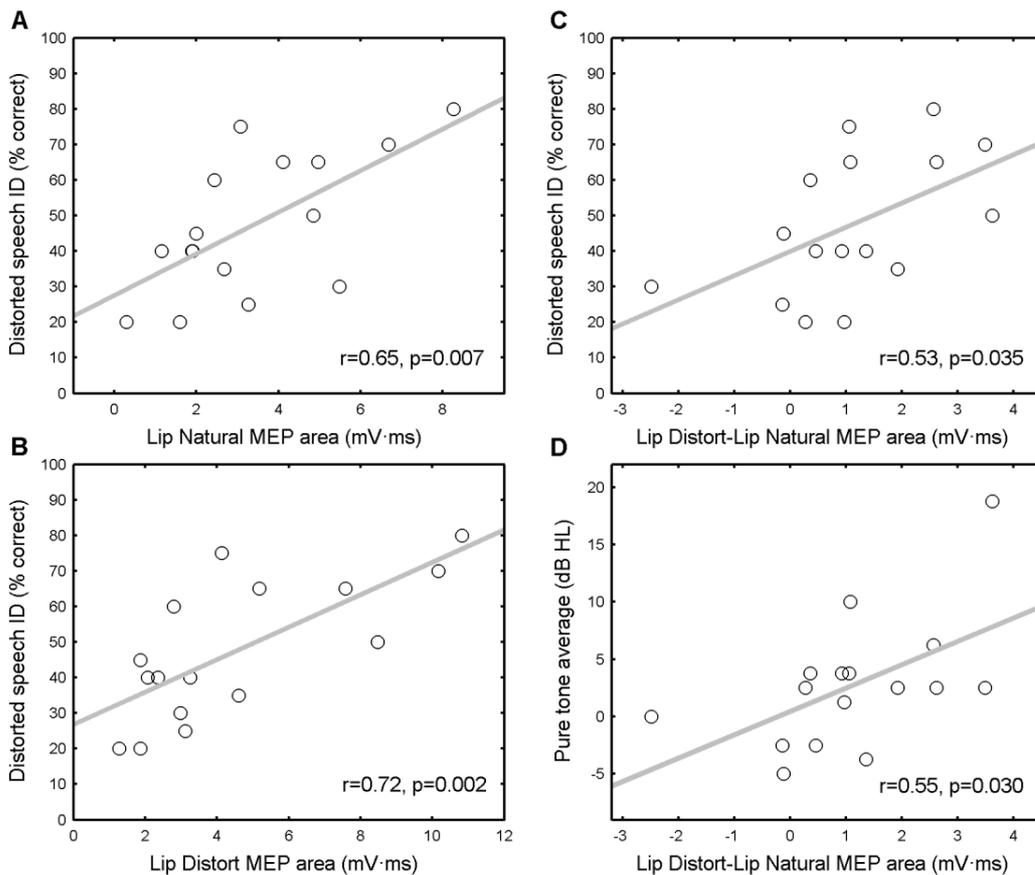


Figure 4. Scatterplots demonstrating relationship between identification accuracy of distorted speech stimuli, and (A) lip MEP area during perception of natural speech, (B) lip MEP area during perception of distorted speech, and (C) the difference in MEP area when MEPs recorded during perception of natural speech were subtracted from MEPs recorded during perception of distorted speech (difference score), and (D) relationship between hearing threshold and MEP difference score.

4. Discussion

This study aimed to disambiguate between two competing action simulation accounts to assess how listening to different types of speech actions engages the motor system during speech perception (Pickering and Garrod, 2013a; Wilson and Knoblich, 2005). To this end, MEPs from M1 lip and hand areas were measured in response to distorted speech sounds produced using a motor perturbation during speech production, and normal, naturally articulated speech sounds, articulated using either the lip or tongue articulators. Our results demonstrate that (1) MEPs from lip muscles were larger in response to distorted speech sounds compared to natural speech; (2) MEPs were further facilitated somatotopically such that MEPs recorded during perception of distorted lip sounds were larger than for distorted tongue sounds; and (3) the extent of excitation in the speech motor system during speech perception was functionally related to identification accuracy of the distorted speech stimuli. MEPs recorded from M1 hand were not modulated by any dimension of the speech stimuli, in line with previous studies that have shown that speech modulates excitability of the motor representation of the lip specifically (Möttönen and Watkins, 2009; Swaminathan et al., 2013). Our finding of increased motor excitability during perception of distorted relative to natural speech is consistent with predictions made by Wilson & Knoblich (Wilson and Knoblich, 2005), who propose that activity in motor areas is maximal when perception is challenging, and complements and extends previous findings from Murakami and colleagues (Murakami et al., 2011). This contrasts with Pickering and Garrod's (Pickering and Garrod, 2013a)

integrated account of perception and production, which postulates that motor activity is greatest when the observer is more familiar with the perceived action (for further detail about simulation accounts and their distinctions, see Pickering and Garrod 2013b). Increased excitability during perception of distorted compared to natural speech is also consistent with non-speech action observation data, where action observation appears to catalyse simulative mapping of observed actions in cortical motor areas (Keysers and Gazzola, 2006).

In addition, the presence of somatotopic facilitation during distorted speech perception also corroborates Wilson & Knoblich's (Wilson and Knoblich 2005) hypothesis. Under their account, articulator-specific effects would be maximally dissociable when listening is difficult, and not during perception of clear speech. This somatotopic finding also provides further support to the view that motor representations of articulators activate during speech perception in an articulatory feature-specific manner (D'Ausilio et al., 2009; Fadiga et al., 2002). Specifically, our data are consistent with previous demonstrations of motor somatotopy in noise (D'Ausilio et al., 2012), and provide novel neurophysiological evidence of this phenomenon. Based on the current data, it does not appear that, at the level of lip muscle potentiation, somatotopic facilitation is evident during natural speech perception. This contrasts with previous findings that have shown somatotopic TMS effects reflected in differences in behavioural performance when using non-distorted speech (Bartoli et al., 2015; Möttönen and Watkins, 2009). The high degree of stimulus predictability and certainty, and/or lack of an active listening task to moderate stimulus engagement may have resulted in the

absence of somatotopic effects during perception of natural speech, and modulation of the motor system more generally (Hickok et al., 2011; Möttönen and Watkins, 2009; Sato et al., 2009).

Nonetheless, our findings reflect the possibility that speech motor activation can occur at two levels of specificity, which need not be mutually exclusive. These levels may represent 1) general, system-level activation in the speech motor system, or 2) specific, feature-level activation in the speech motor system. Notably, although there was a somatotopic difference in the amount of activation in the distorted condition, activation was significantly increased for both lip- and tongue-based distorted sounds. This could indicate that when listening is challenged, there is both a general increase in motor activity in response to speech globally, which is further enhanced specifically for speech sounds that match somatotopically. Activation may operate at this level of specificity as a function of listening difficulty, when feature-level motoric information may confer a benefit on auditory signal decoding. Indeed, it would have been useful to ascertain if motor activation to natural speech reflects a general, system-level increase in speech motor activity relative to activity associated with a non-speech baseline, to assess if meaningfulness, and the effort associated with deriving meaning from the signal, modulates speech motor resonances. Control conditions employed in previous literature, however, do suggest this to be the case (Murakami, Restle, & Ziemann, 2011), and furthermore, Möttönen, Farmer and Watkins (Möttönen et al., 2010) have demonstrated that motor activity increases when observing communicative hand signals after the observer has learnt that such signals are meaningful.

From a simulation account perspective, the findings of the present study clarify the perceptual conditions that maximally engage motor simulation processes. Importantly, however, we also present data that go beyond the observation of group effects, as we highlight how the relationship between motor activation and perception may operate on an individual level, complementing previously observed relationships between behaviour and the speech motor system (Bartoli et al., 2015; D’Ausilio et al., 2014). Here, we found that individuals were highly varied in their ability to identify distorted speech sounds and that this corresponded to the level of activity in their speech motor system during speech perception. Therefore, it was possible to demonstrate that individuals who were particularly good at perceiving distorted speech also independently demonstrated greater motor activation, relative to poor perceivers. This is consistent with the notion that increased sensorimotor processing may improve speech perception (D’Ausilio et al., 2012; Wilson and Knoblich, 2005). Crucially, the increase in motor activity from natural to distorted speech perception was positively correlated with better identification accuracy of distorted speech. This indicates that individuals who show greater motoric sensitivity to differences in speech quality also demonstrate better perceptual processing when listening is difficult.

However, we also found that this relationship was mediated by peripheral hearing sensitivity. This was surprising given the normally hearing, young adult subject sample used, and the suprathreshold presentation of speech stimuli. Direct correlations established that this was largely driven by a positive relationship between PTA and MEP facilitation during perception of distorted, relative to natural, speech. This indicates that

listeners with less sensitive, but still normal, hearing show greater motor cortex activation when listening is challenging, relative to individuals with better hearing. To our knowledge, the present study is the first to relate motor activation to hearing sensitivity, and suggests that recruitment of motor regions during speech perception may be related not only to the distortion of the speech signal, but the peripheral hearing status of the listener. Despite normal hearing, differences in audiometric configuration at the cochlea are known to impact upon the latency and amplitude of the speech signal that is propagated through the auditory system (Nuttall et al., 2015). Whilst the primary motor cortex does not share anatomical connections with auditory cortex, primary motor activity may be regulated by functional connectivity with auditory cortical regions, wherein the quality of the received speech signal could moderate the extent of processing recruited from non-auditory regions. In addition, it is also known that individual differences in cognitive factors modulate motor activity; in an fMRI study (Szenkovits et al., 2012), it was observed that listeners vary in the degree to which they recruit motor regions as a function of short-term memory ability. Taken together, these findings may indicate that both sensory and cognitive factors may modulate the relationship between speech perception and the motor system.

The findings of the present study can be interpreted to support a complementary involvement of the motor cortex during speech perception, particularly when perception is challenging. This is in line with Wilson & Knoblich's (Wilson and Knoblich, 2005) simulation account, which posits that the motor system contributes to perception by covertly imitating or simulating others' actions during perception of the action, in order

to aid understanding. Wilson and Knoblich hypothesise that covert imitation is part of a perceptual emulation process through which implicit knowledge of motor gestures is used as a model to follow another person's actions (Wilson and Knoblich 2005). Emulation is considered to be particularly useful when perception is challenging, as it assists in both filling in missing information, and predicting the likely course of an ongoing action; the outcomes of which are then fed back to the sensory systems involved in perception. Such a predictive coding scheme for speech perception could be achieved if speech-related activity in auditory areas leads to activation of articulatory motor regions for the purposes of internal, motoric simulation, in a process automatized by Hebbian correlation learning (Braitenberg and Pulvermüller, 1992). For example, auditory-motor co-activation may be effected by the oscillatory activity of neuronal circuits distributed across inferior-frontal and superior-temporal areas, which reach into articulatory motor cortex (Schomers et al., 2014). Perceiving speech then triggers the automatic resonance of these circuits; however, the extent of the oscillatory activity may be modulated by the relative clarity of the speech signal received by auditory cortex. When speech is clear and comprehensible, prediction of the upcoming speech signal can be successfully constrained by the acoustic patterns that match what has already been comprehended. When speech is distorted, the acoustic patterns stored in auditory cortex cannot successfully constrain the possible perceptual options. As such, there is increased signaling to and from M1 to auditory areas, in order to limit the potential interpretations of the degraded auditory input using simulated knowledge of articulatory plans. As a result, activity in cortical representations

of speech muscles is increased, observable as heightened muscle potentiation or greater MEPs.

This possibility is also in line with the Perception-for-Action-Control Theory (PACT), proposed by Schwartz and colleagues (Schwartz et al 2012), which states that speech units are shaped by a double set of articulatory and sensory (audiovisual) constraints, and are characterised by both a motor and a perceptual coherence. Indeed, findings from the present study support the possibility that the motor system is most strongly active when listening is difficult, at which point motor information can enhance specification of possible auditory and visual trajectories and enhance speech scene analysis (Schwartz et al 2012). This proposal concurs with the central principle of Wilson and Knoblich's (Wilson and Knoblich 2005) motor simulation account, whereby motor information can assist action understanding when sensory information is ambiguous.

How auditory and motor areas accomplish this process remains unclear, but the communication pathway(s) involved most likely implicates the arcuate fasciculus, whose reciprocal connections link temporal auditory and frontal articulatory areas. Concurrently, Murakami and colleagues (Murakami et al., 2015) report that M1 excitability during speech perception was reduced by suppression of activity in the posterior superior temporal sulcus, the sylvian parieto-temporal region, and by the combined suppression of pars opercularis of the inferior frontal gyrus and dorsal premotor cortex (PMC): individual suppression of these areas alone did not result in suppression of MEPs. Consistent with this possibility, M1 excitability is thought to be modulated by cortico-cortical inputs from ventral PMC during action observation

(Shimazu et al. 2004) and object perception (Cattaneo et al., 2005). Watkins and Paus (Watkins and Paus, 2004) also found that increased blood flow in ventral PMC (Broca's area) correlated with motor excitability during speech perception in a combined positron emission tomography (PET) and TMS study.

Indeed, research in monkeys has shown that neurons in PMC modulate their firing frequency in relation to motor tasks that require a visual cue, and that firing occurs in PMC at the appearance of the cue well before M1 acts on the cue (Godschalk et al., 1985; Halsband and Passingham, 1985), with similar findings in humans (Cattaneo et al., 2005). If PMC is similarly modulated by auditory information, during speech perception PMC may initiate articulatory motor resonances in readiness of the need for an overt speech-motor response to be executed by M1. This initiation or preparation of action may also be operational under challenging listening conditions, to facilitate covert speech-motor simulation. As such, PMC, likely in concert with pars opercularis of the inferior frontal gyrus, may be at the origin of action simulation, with motor resonances initiated by PMC bringing M1 neurons closer to threshold, resulting in greater neural recruitment for MEPs elicited during perception of distorted speech. Alternatively, we cannot rule out the possibility that TMS to M1 affects M1 inputs into PMC, and resultant changes in cortico-cortical, or cortico-subcortical PMC inputs affects the elicited MEPs. Relatedly, Sato and colleagues (Sato et al., 2009) replicated the PMC TMS study by Meister and colleagues (Meister et al., 2007) using stimuli in quiet only, and found that contrary to Meister's findings, TMS to PMC had little effect on speech perception, and suggested that premotor areas are critical to perceive speech in difficult listening conditions only.

Further work is required to elucidate the connections between these speech motor areas, the listening conditions that modulate them, and the physiological bases of motor simulation processes during speech perception.

Finally, it should be noted that the results of the current study are limited to pre-lexical speech perception, as well as to only motor-based speech distortion. The extent to which these findings generalize to perception of more complex speech stimuli, or speech which is degraded via another means, such as background noise, must be determined to better understand the role of the motor system in speech processing. However, previous work on post-lexical speech has demonstrated increased speech motor activity during auditory perception of connected sentences when compared to the level of speech motor activity during perception of eye and brow movements (Watkins et al., 2003), suggesting that the articulatory motor system may also play a role in post-lexical speech processing. Furthermore, Murakami and colleagues (2011) developed this paradigm by adding background noise and observed motor facilitation for speech in noise, but notably no study has contrasted if and how the type of distortion in the speech signal modulates lip MEPs. However, it is worth acknowledging that the natural speech condition employed in this study was not completely noise-free; due to the online nature of the TMS protocol, perception of both natural and distorted stimulus types was accompanied by the sound of the TMS pulse being discharged. This may have potentially resulted in a small degree of partial simultaneous masking during perception of the speech stimuli. However, unpublished data from our lab suggests that the sound generated by the TMS pulse at 100% stimulation intensity is at least 15 dB SPL less intense than the level of speech

presentation used in the present study. Given that we applied TMS at around half of that intensity, it is likely that the sound of the TMS pulse was approximately 30 dB less intense than the speech stimuli.

In conclusion, the present study tested the dissociable predictions made by two different accounts of motor simulation. Lip MEP data were found to be highly consistent with predictions made by Wilson & Knoblich (Wilson and Knoblich, 2005), who assert that motor activity increases when perception is challenging. This study thus provides direct neurophysiological evidence of a role of motor cortex when speech is hard to comprehend due to a speech-internal distortion, and corroborates the notion that systems for speech perception and speech production are intricately interlinked.

Funding: This work was supported by a Project Grant from The Leverhulme Trust (RPG-2013-254).

Acknowledgements: Our thanks go to the individuals who participated in this study.

References

- Adank, P., Davis, M.H., Hagoort, P., 2012. Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia* 50, 77–84.
doi:10.1016/j.neuropsychologia.2011.10.024
- Baldissera, F., Cavallari, P., Craighero, L., Fadiga, L., 2001. Modulation of spinal excitability during observation of hand actions in humans. *Eur. J. Neurosci.* 13, 190–

4.

- Bartoli, E., D'Ausilio, A., Berry, J., Badino, L., Bever, T., Fadiga, L., 2015. Listener-Speaker Perceived Distance Predicts the Degree of Motor Contribution to Speech Perception. *Cereb. cortex* 25, 281–8. doi:10.1093/cercor/bht257
- Boersma P. 2001., Praat, a system for doing phonetics by computer. *Glott International*, 5:341-345.
- Braitenberg, V., Pulvermüller, F., 1992. Model of a neurological theory of speech. *Naturwissenschaften* 79, 103–17.
- British Society of Audiology (BSA). 2011. Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. Retrieved from <http://www.thebsa.org.uk/resources/>. Last accessed October 2015.
- Calvo-Merino, B., Glaser, D.E., Grèzes, J., Passingham, R.E., Haggard, P., 2005. Action observation and acquired motor skills: an FMRI study with expert dancers. *Cereb. Cortex* 15, 1243–9. doi:10.1093/cercor/bhi007
- Cattaneo, L., Voss, M., Brochier, T., Prabhu, G., Wolpert, D.M., Lemon, R.N., 2005. A cortico-cortical mechanism mediating object-driven grasp in humans. *Proc. Natl. Acad. Sci.* 102, 898–903. doi:10.1073/pnas.0409182102
- D'Ausilio, A., Bufalari, I., Salmas, P., Fadiga, L., 2012. The role of the motor system in discriminating normal and degraded speech sounds. *Cortex* 48, 882–887. doi:10.1016/j.cortex.2011.05.017
- D'Ausilio, A., Maffongelli, L., Bartoli, E., Campanella, M., Ferrari, E., Berry, J., Fadiga, L., 2014. Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-Doppler ultrasound imaging. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369, 20130418. doi:10.1098/rstb.2013.0418
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., Fadiga, L., 2009. The Motor Somatotopy of Speech Perception. *Curr. Biol.* 19, 381–385. doi:10.1016/j.cub.2009.01.017
- Davis, M.H., Johnsrude, I.S., 2007. Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear. Res.* 229, 132–47. doi:10.1016/j.heares.2007.01.014
- Davis, M.H., Johnsrude, I.S., 2003. Hierarchical processing in spoken language

- comprehension. *J. Neurosci.* 23, 3423–31.
- Diehl, R.L., Lotto, A.J., Holt, L.L., 2004. Speech perception. *Annu. Rev. Psychol.* 55, 149–79. doi:10.1146/annurev.psych.55.090902.142028
- Du, Y., Buchsbaum, B.R., Grady, C.L., Alain, C., 2014. Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc. Natl. Acad. Sci. U. S. A.* 111, 7126–31. doi:10.1073/pnas.1318738111
- Fadiga, L., Craighero, L., Buccino, G., Rizzolatti, G., 2002. Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *Eur. J. Neurosci.* 15, 399–402.
- Godschalk, M., Lemon, R.N., Kuypers, H.G.J.M., Van Der Steen, J., 1985. The involvement of monkey premotor cortex neurones in preparation of visually cued arm movements. *Behav. Brain Res.* 18, 143–157. doi:10.1016/0166-4328(85)90070-1
- Halsband, U., Passingham, R.E., 1985. Premotor cortex and the conditions for movement in monkeys (*Macaca fascicularis*). *Behav. Brain Res.* 18, 269–277. doi:10.1016/0166-4328(85)90035-X
- Hervais-Adelman, A.G., Carlyon, R.P., Johnsrude, I.S., Davis, M.H., 2012. Brain regions recruited for the effortful comprehension of noise-vocoded words. *Lang. Cogn. Process.* 27, 1145–1166. doi:10.1080/01690965.2012.662280
- Hickok, G., Houde, J., Rong, F., 2011. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–22. doi:10.1016/j.neuron.2011.01.019
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi:10.1038/nrn2113
- Jusczyk, P.W., Smith, L.B., Murphy, C., 1981. The perceptual classification of speech. *Percept. Psychophys.* 30, 10–23.
- Keysers, C., Gazzola, V., 2006. Towards a unifying neural theory of social cognition. *Prog. Brain Res.* 156, 379–401. doi:10.1016/S0079-6123(06)56021-2
- Lieberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. *Psychol. Rev.* 74, 431–61.
- Lieberman, A.M., Mattingly, I.G., 1985. The motor theory of speech perception revised.

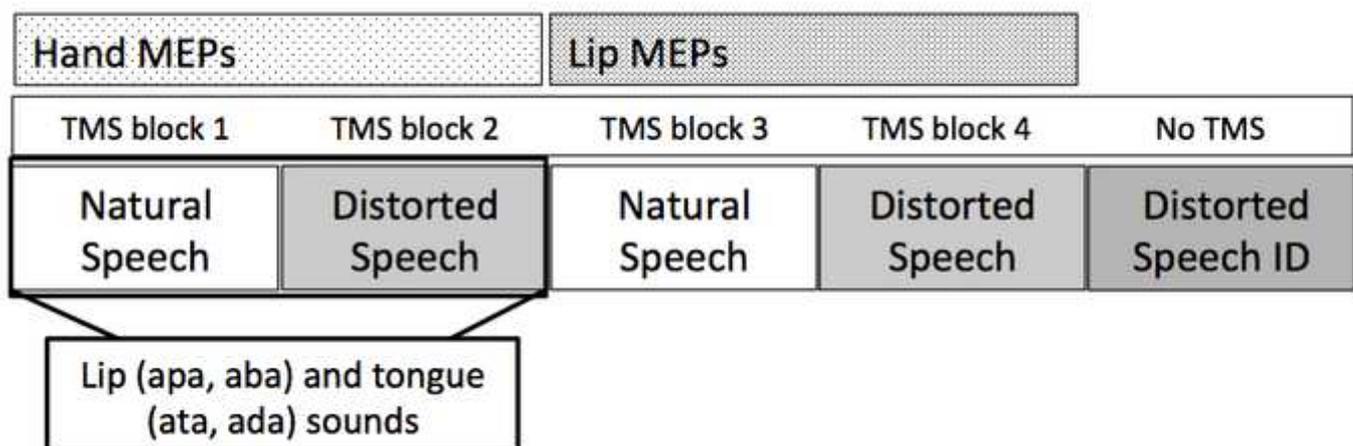
Cognition 21, 1–36.

- Meister, Wilson, Deblieck, Wu, 2007. Report The Essential Role of Premotor Cortex in Speech Perception. *Curr. Biol.* 17, 1692–1696. doi:10.1016/j.cub.2007.08.064
- Möttönen, R., Farmer, H., Watkins, K.E., 2010. Lateralization of motor excitability during observation of bimanual signs. *Neuropsychologia* 48, 3173–3177. doi:10.1016/j.neuropsychologia.2010.06.033
- Möttönen, R., Watkins, K.E., 2009. Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* 29, 9819–9825. doi:10.1523/JNEUROSCI.6018-08.2009
- Murakami, T., Kell, C.A., Restle, J., Ugawa, Y., Ziemann, U., 2015. Left Dorsal Speech Stream Components and Their Contribution to Phonological Processing. *J. Neurosci.* 35, 1411–1422. doi:10.1523/JNEUROSCI.0246-14.2015
- Murakami, T., Restle, J., Ziemann, U., 2011. Observation-execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. *Neuropsychologia* 49, 2045–2054. doi:10.1016/j.neuropsychologia.2011.03.034
- Nuttall, H.E., Moore, D.R., Barry, J.G., Krumbholz, K., de Boer, J., 2015. The influence of cochlear spectral processing on the timing and amplitude of the speech-evoked auditory brain stem response. *J. Neurophysiol.* 113, 3683–91. doi:10.1152/jn.00548.2014
- Osnes, B., Hugdahl, K., Specht, K., 2011. Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *Neuroimage* 54, 2437–45. doi:10.1016/j.neuroimage.2010.09.078
- Pickering, M.J., Garrod, S., 2013a. An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–47. doi:10.1017/S0140525X12001495
- Pickering, M.J., Garrod, S., 2013b. Authors’ response: forward models and their implications for production, comprehension, and dialogue. *Behav. Brain Sci.* 36, 377–92.
- Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–24. doi:10.1038/nn.2331

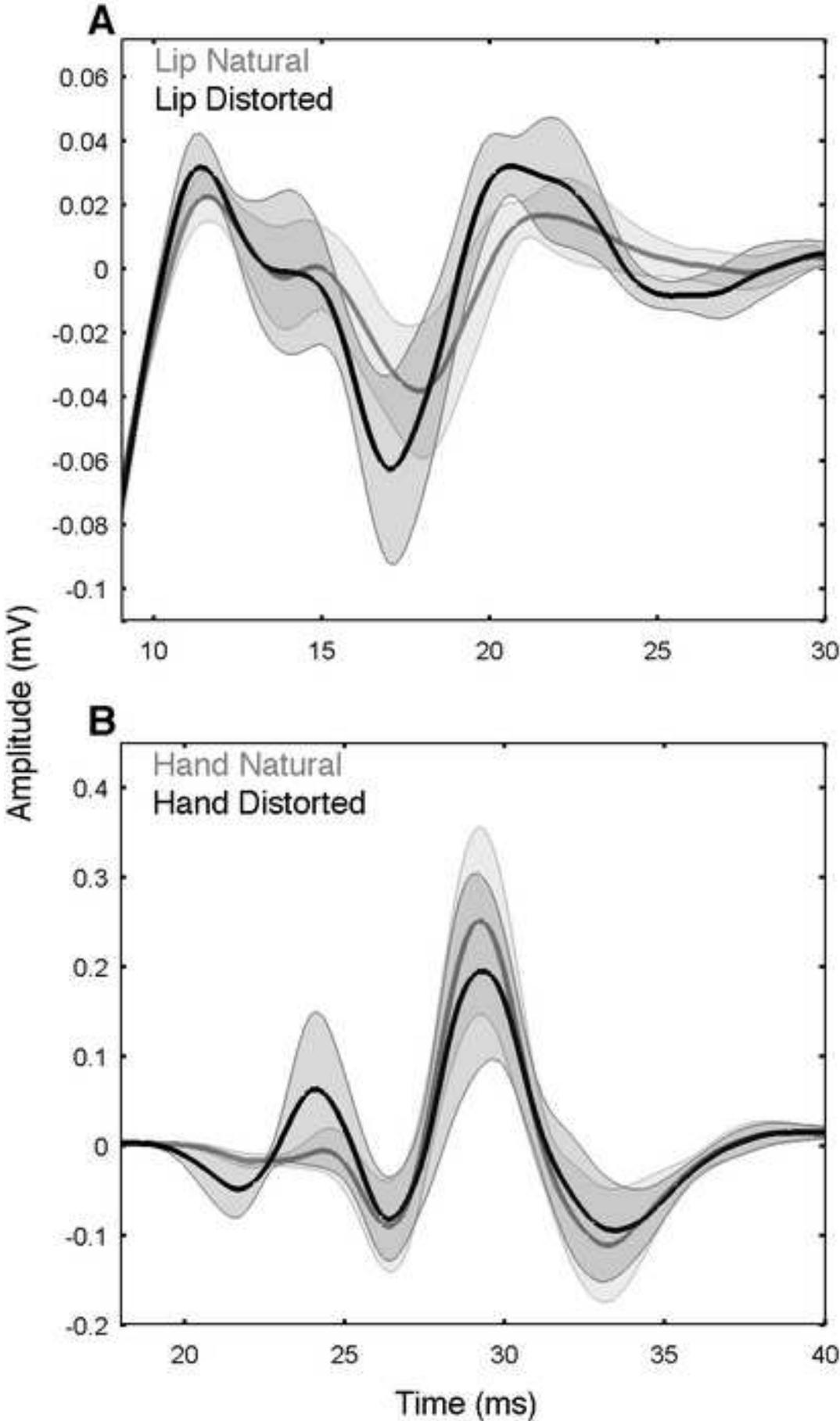
- Rossini, P.M., Barker, A.T., Berardelli, A., Caramia, M.D., Caruso, G., Cracco, R.Q., Dimitrijevic, M.R., Hallett, M., Katayama, Y., Lucking, C.H., 1994. Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an IFCN committee. *Electroencephalogr. Clin. Neurophysiol.* 91, 79-92.
- Roy, A.C., Craighero, L., Fabbri-Destro, M., Fadiga, L., 2008. Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. *J. Physiol. Paris* 102, 101–105. doi:10.1016/j.jphysparis.2008.03.006
- Rösler K.M., Magistris M.R., 2008. The size of motor-evoked potentials: Influencing parameters and quantification. In: Wassermann E.M., Epstein C.M., Ziemann U., Walsh V., Paus T., Lisanby S.H., editors. *The Oxford Handbook of Transcranial Stimulation*. 1st ed. Oxford (UK): Oxford University Press. p 77-89.
- Sato, M., Buccino, G., Gentilucci, M., Cattaneo, L., 2010. On the tip of the tongue: Modulation of the primary motor cortex during audiovisual speech perception. *Speech Commun.* 52, 533–541. doi:10.1016/j.specom.2009.12.004
- Sato, M., Tremblay, P., Gracco, V.L., 2009. A mediating role of the premotor cortex in phoneme segmentation. *Brain Lang.* 111, 1–7. doi:10.1016/j.bandl.2009.03.002
- Schomers, M.R., Kirilina, E., Weigand, A., Bajbouj, M., Pulvermüller, F., 2014. Causal Influence of Articulatory Motor Cortex on Comprehending Single Spoken Words: TMS Evidence. *Cereb. Cortex.* doi:10.1093/cercor/bhu274
- Schwartz, J.-L., Basirat, A., Ménard, L., Sato, M., 2012. The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *J. Neurolinguistics* 25, 336–354. doi:10.1016/j.jneuroling.2009.12.004
- Scott, S.K., McGettigan, C., Eisner, F., 2009. A little more conversation, a little less action--candidate roles for the motor cortex in speech perception. *Nat. Rev. Neurosci.* 10, 295–302. doi:10.1038/nrn2603
- Shahin, A.J., Bishop, C.W., Miller, L.M., 2009. Neural mechanisms for illusory filling-in of degraded speech. *Neuroimage* 44, 1133–43. doi:10.1016/j.neuroimage.2008.09.045
- Shimazu, H., 2004. Macaque Ventral Premotor Cortex Exerts Powerful Facilitation of Motor Cortex Outputs to Upper Limb Motoneurons. *J. Neurosci.* 24, 1200–1211. doi:10.1523/JNEUROSCI.4731-03.2004

- Swaminathan, S., MacSweeney, M., Boyles, R., Waters, D., Watkins, K.E., Möttönen, R., 2013. Motor excitability during visual perception of known and unknown spoken languages. *Brain Lang.* 126, 1–7. doi:10.1016/j.bandl.2013.03.002
- Szenkovits, G., Peelle, J.E., Norris, D., Davis, M.H., 2012. Individual differences in premotor and motor recruitment during speech perception. *Neuropsychologia* 50, 1380–92. doi:10.1016/j.neuropsychologia.2012.02.023
- Turkeltaub, P.E., Coslett, H.B., 2010. Localization of sublexical speech perception components. *Brain Lang.* 114, 1–15. doi:10.1016/j.bandl.2010.03.008
- Watkins, K., Paus, T., 2004. Modulation of Motor Excitability during Speech Perception: The Role of Broca's Area. *J. Cogn. Neurosci.* 16, 978–987. doi:10.1162/0898929041502616
- Watkins, K.E., Strafella, a. P., Paus, T., 2003. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41, 989–994. doi:10.1016/S0028-3932(02)00316-0
- Wild, C.J., Davis, M.H., Johnsrude, I.S., 2012a. Human auditory cortex is sensitive to the perceived clarity of speech. *Neuroimage* 60, 1490–502. doi:10.1016/j.neuroimage.2012.01.035
- Wild, C.J., Yusuf, A., Wilson, D.E., Peelle, J.E., Davis, M.H., Johnsrude, I.S., 2012b. Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–21. doi:10.1523/JNEUROSCI.1528-12.2012
- Wilson, M., Knoblich, G., 2005. The case for motor involvement in perceiving conspecifics. *Psychol. Bull.* 131, 460–73. doi:10.1037/0033-2909.131.3.460

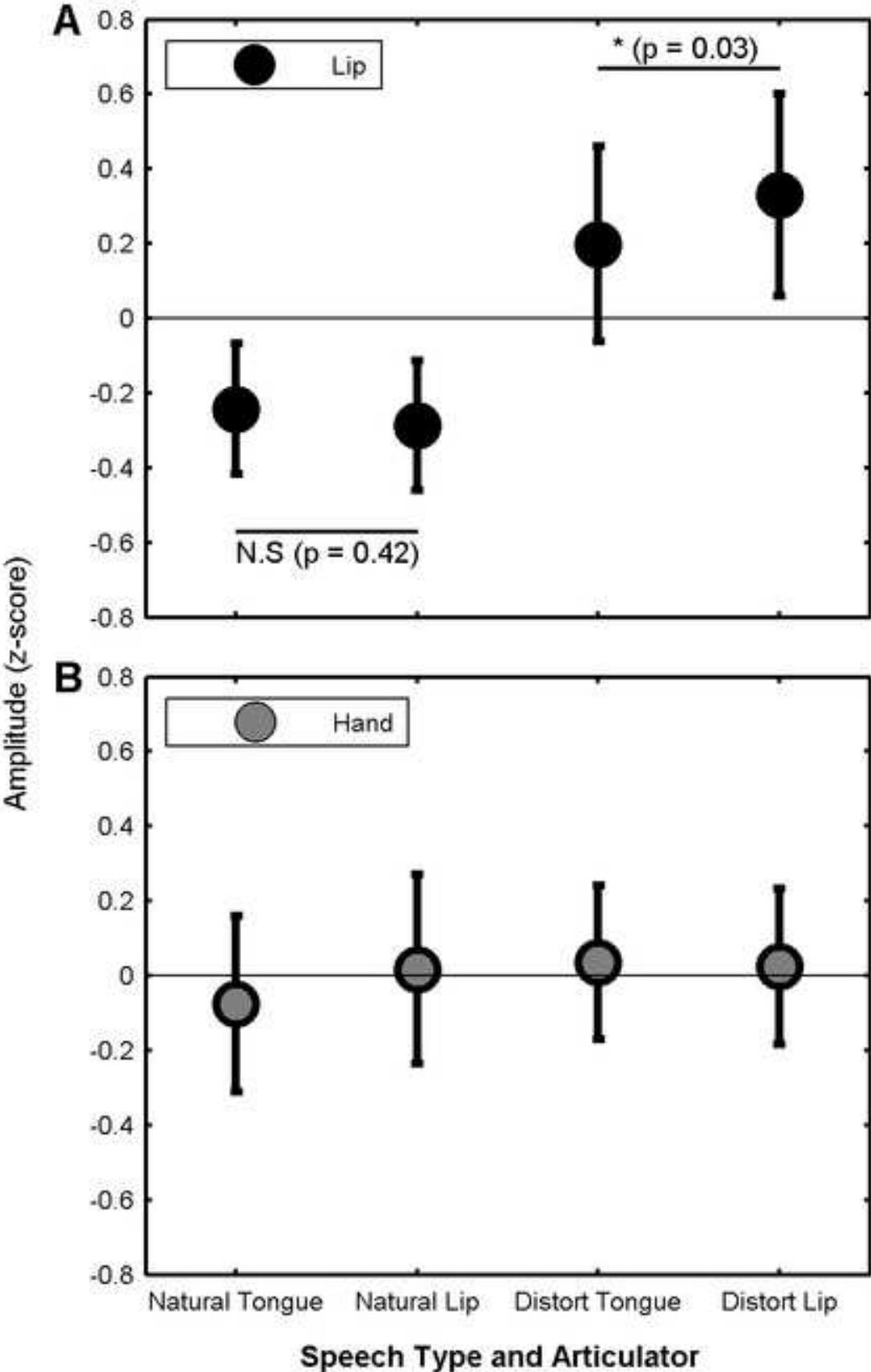
9. Figure
[Click here to download high resolution image](#)



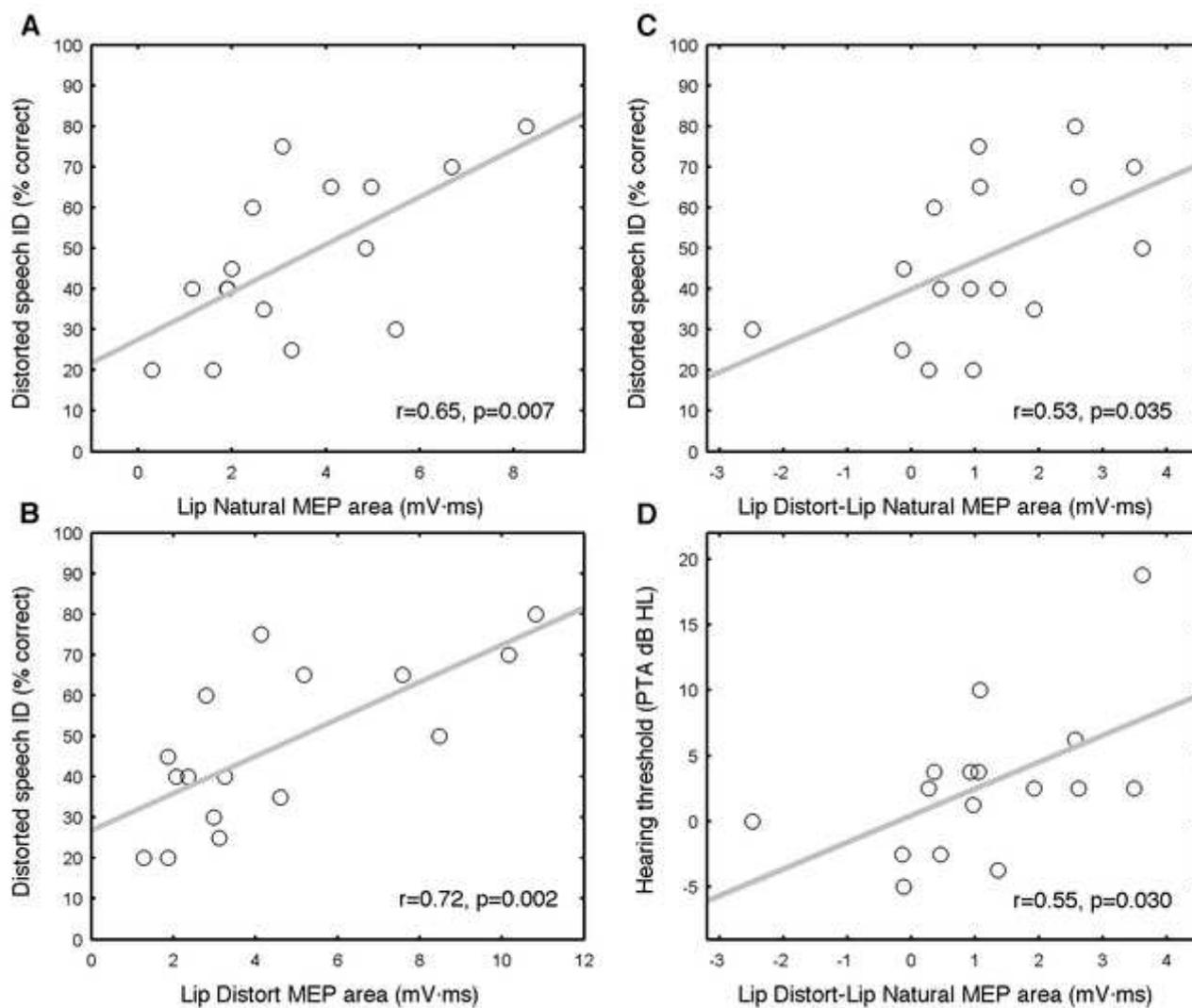
9. Figure
[Click here to download high resolution image](#)



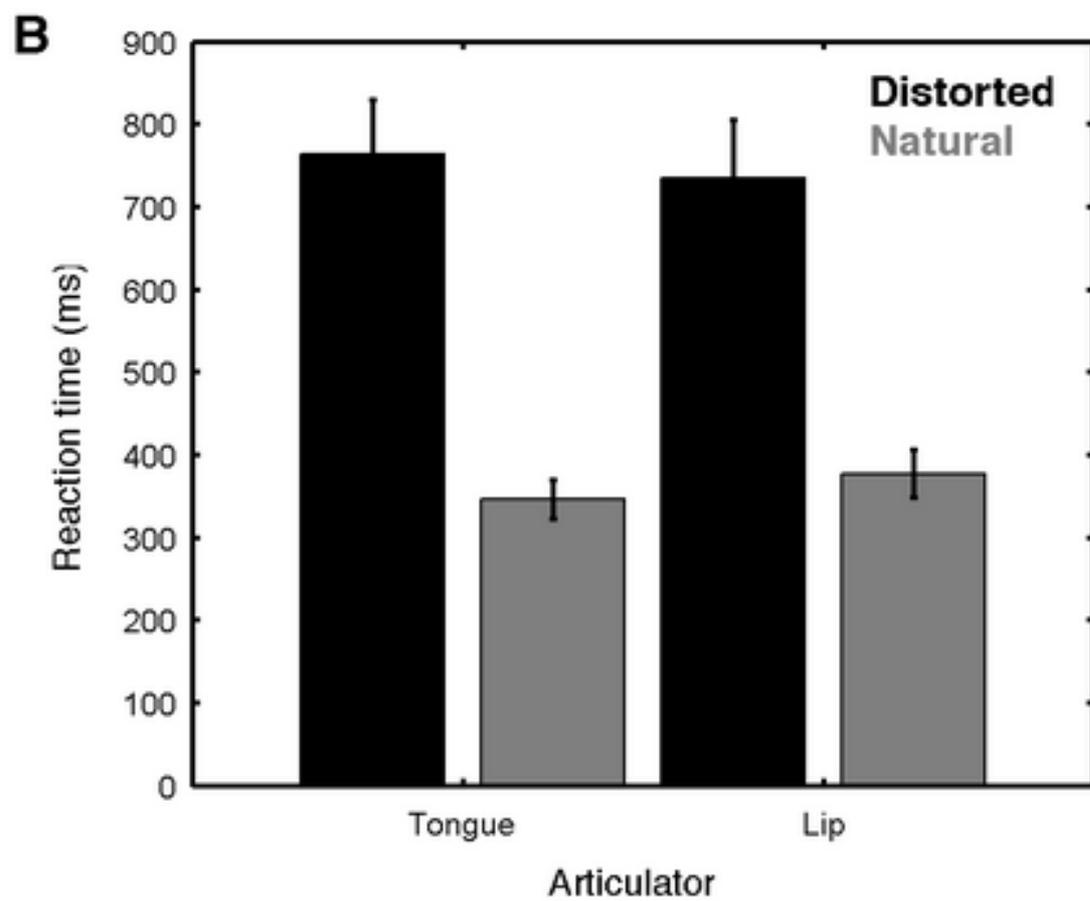
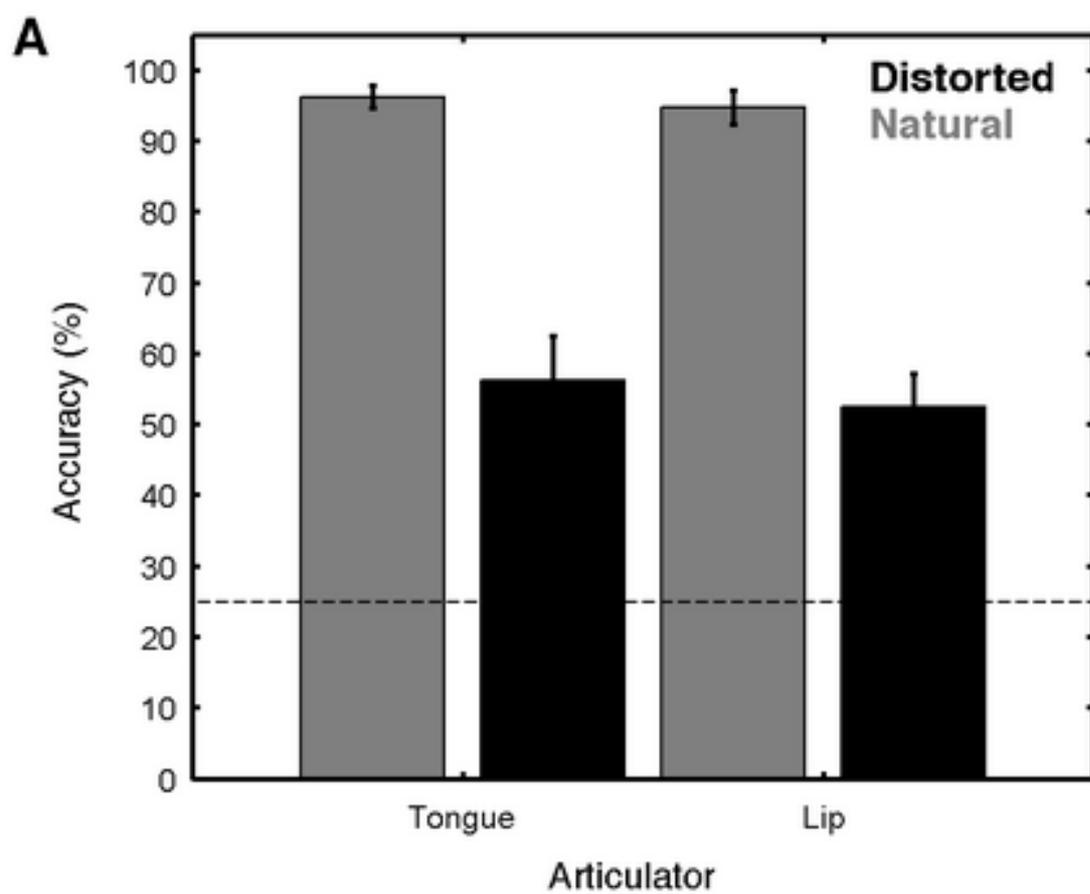
9. Figure
[Click here to download high resolution image](#)



9. Figure
[Click here to download high resolution image](#)



9. Figure
[Click here to download high resolution image](#)



10. Supplementary Material

[Click here to download 10. Supplementary Material: NeuroImage_SupplementaryMaterial_HENuttall_Revision.docx](#)