PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

rstb.royalsocietypublishing.org

Research

Cite this article: 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. *Phil. Trans. R. Soc. B* **369**: 20130418. http://dx.doi.org/10.1098/rstb.2013.0418

One contribution of 19 to a Theme Issue 'Mirror neurons: fundamental discoveries, theoretical perspectives and clinical implications'.

Subject Areas:
behaviour, neuroscience

Keywords:
speech perception, transcranial magnetic stimulation, motor theory of speech perception, mirror neurons, motor synergies, tongue

Author for correspondence:
A. D'Ausilio
e-mail: alessandro.dausilio@iit.it

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2013.0418 or via http://rstb.royalsocietypublishing.org.

Royal Society Publishing

Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-Doppler ultrasound imaging

A. D'Ausilio[1], L. Maffongelli[1], E. Bartoli[1], M. Campanella[1], E. Ferrari[1], J. Berry[1] and L. Fadiga[1,2]

[1]Robotics Brain and Cognitive Sciences Department, RBCS, Italian Institute of Technology, IIT, via Morego, 30, Genova 16163, Italy
[2]Section of Human Physiology, Università di Ferrara, via Fossato di Mortara, 17/19, Ferrara 44100, Italy

The activation of listener's motor system during speech processing was first demonstrated by the enhancement of electromyographic tongue potentials as evoked by single-pulse transcranial magnetic stimulation (TMS) over tongue motor cortex. This technique is, however, technically challenging and enables only a rather coarse measurement of this motor mirroring. Here, we applied TMS to listeners' tongue motor area in association with ultrasound tissue Doppler imaging to describe fine-grained tongue kinematic synergies evoked by passive listening to speech. Subjects listened to syllables requiring different patterns of dorso-ventral and antero-posterior movements (/ki/, /ko/, /ti/, /to/). Results show that passive listening to speech sounds evokes a pattern of motor synergies mirroring those occurring during speech production. Moreover, mirror motor synergies were more evident in those subjects showing good performances in discriminating speech in noise demonstrating a role of the speech-related mirror system in feed-forward processing the speaker's ongoing motor plan.

1. Introduction

Actions are central components of social interaction and communication among individuals [1,2]. Among interactive actions, speech is a special case. In fact, speech is the only complex motor behaviour that is inherently human and constitutes a fundamental means to convey intentions and beliefs during social interaction. Speech, defined as the ability to produce sounds by coordinating and moving specific oral motor structures, has indeed shown to elicit mirror-like activities in the listener. The first and most convincing proof on the activation of motor areas during listening of speech sounds came from transcranial magnetic stimulation (TMS). Fadiga *et al*. [3] demonstrated a modulation of tongue motor-evoked potentials (MEPs) while listening to words and pseudo-words containing an alveolar trill (i.e. /birro/ for pseudo-words as opposed to /birra/ for words). The study designed by Fadiga *et al*. [3] had supra-threshold TMS delivered in a time-locked manner to speech stimuli (100 ms after the onset of the critical sound). Watkins *et al*. [4] were able to show that speech, in either auditory or visual form, induced larger *orbicularis oris* MEPs only for the stimulation of the left hemisphere. These pioneering studies demonstrated that the motor system is somatotopically [3] activated by both listening and watching speech stimuli, with a left hemisphere advantage [4].

An additional critical aspect that emerged in these studies was related to the temporal deployment of speech listening-evoked motor resonance. Roy *et al*. [5] explicitly manipulated TMS timing with respect to the presentation of speech sounds (words and pseudo-words) by taking into account word frequency as well. Interestingly, they found that TMS-revealed motor facilitation was related to word frequency. Frequency effects were not present in the first 100 ms,

whereas at 200 and 300 ms, MEPs were found to be much larger for the rare words than for common words and pseudo-words. According to these results, it seems that motor resonance is critical during early phonological parsing for speech processing in general but, at later stages, lexical frequency overrides such a mechanism by eventually exploiting top-down completion.

More recently, D'Ausilio *et al*. [6] recorded tongue cortico-bulbar excitability during phoneme expectation induced by stimulus predictability (the stimulus characteristics could be anticipated with 75% probability). Results showed that motor mirroring is not merely a passive and automatic response to environmental stimuli but that it normally anticipates incoming sensory events by formulating specific feed-forward hypotheses. This feed-forward motor activity is continuously tested against incoming and subtle cues, such as co-articulation features.

Summing up, 20 years after the discovery of mirror neurons [7] and 10 years after the discovery of the same mechanism in speech perception [3], just a handful of papers have been published on this topic. The reason is probably not owing to the lack of interest on the topic. The fact that the listeners' motor system participates in speech perception has important and wide implications for many disciplines. One possible reason why there are few TMS studies on this topic is probably owing to the additional complexity in recording electromyography (EMG) responses from the tongue and the impossibility of discriminating motor synergies from the electric patterns. In fact, using the traditional surface electrode placement tongue muscles cannot be dissociated. The tongue is indeed characterized by many degrees of freedom (at least six, as shown by [8]) supported by a complex muscles anatomy [9]. Furthermore, and more in general, the combined use of TMS and EMG to analyse changes in cortico-bulbar excitability has proved more challenging than the study of the cortico-spinal pathway [10–13].

In this study, we seek to deploy and test the potentiality of a new method to record whole tongue movement synergies evoked by TMS. We used ultrasound tissue Doppler imaging (UTDI) to measure local tongue kinematics during both speech production and perception. UTDI is a standard ultrasound technique, mainly used in cardiology to analyse the local functionality of heart, which employs the Doppler effect to assess structures moving towards or away from the probe, and their relative velocity. To our knowledge, there is only one study using UTDI on speech production, but authors used very different data processing and visualization techniques, which, unfortunately, did not allow full tongue visualization (M-mode; [14]). On the other hand, ultrasound data were successfully used since the 1980s to measure kinematic features of tongue dorsum movements [15]. In those cases, position data had to be tracked frame by frame in order to extract velocity and acceleration profiles from consecutive frames. UTDI data, instead, are better suited for the extraction of velocity information from single snapshots following a given event of interest.

In this study, we first verified the reliability of this approach in recording tongue movements during speech production. Afterwards, we ran a study where subjects were passively listening to some syllables, selected for their mutual distance in terms of motor patterns, while single-pulse TMS was applied to subjects' tongue motor cortex in the exact moment at which the syllables where maximally different during production and a UDTI image was acquired. Subsequently, by using a series of image-processing tools to extract movement features, we were able to demonstrate dissociable patterns in the listeners in agreement with production patterns. Showing the modulation of local patterns of motion during speech listening, specifically matching the patterns of articulation is an additional strong proof that the motor system deploys a specific mirroring of heard speech gestures.

2. Material and methods

This work consists of a *pilot–UTDI study* and a *TMS–UTDI study*. The first one was used to decide the experimental stimuli and the optimal timing of TMS delivery for the TMS–UTDI study. The pilot–UTDI study consisted in the recording of UTDI data during speech production. The TMS–UTDI study was divided into three consecutive sessions: *speech production*, *speech listening* and *speech discrimination* (see Experimental procedure and figure 1*a*).

(a) Subjects

The pilot–UTDI study was run on one right-handed participant (M, age 33). Eleven right-handed subjects participated in the TMS–UTDI study (three males; mean age: 25.7; s.d., 3.9). All subjects had normal hearing abilities and gave informed consent to the experimental procedures, according to the Declaration of Helsinki and the local ethics committee. Subjects were all screened for contraindications to TMS and no immediate or delayed undesired effects of stimulation were produced. Subjects were paid for their participation. The pilot–UTDI study lasted approximately 2 h, and the TMS–UTDI study lasted less than 90 min.

(b) Ultrasound tissue Doppler imaging

A colour Doppler ultrasonic machine (Philips, CX30 CompactXtreme Ultrasound System) was used with a specific TDI transducer (S4–2, Broadband Sector Array Transducer; 4–2 MHz extended frequency range) with TDI acquisition depth of 9 cm and velocity range of $\pm 2.5\ \mathrm{cm\ s^{-1}}$. UTDI acquires data about local motion away or towards the transducer, thus ignoring any component tangential to the probe.

UTDI images could be acquired either in a continuous or triggered mode (figure 1*b*). The triggered mode acquires one single image. Ultrasound or UTDI images acquisition, by definition, requires time. The temporal resolution is limited by the sweep speed of the acoustic beam. And the sweep speed is limited by the speed of sound, as the echo from the deepest part of the image has to return before the next pulse is sent out at a different angle in the neighbouring beam. Our acquisition parameters (depth and number of lines) were constrained by the necessity to record the whole tongue. Therefore, the maximal acquisition frequency was 83 Hz, and thus each image was acquired in about 12 ms, starting from the trigger we provided. Thus, velocity is referred to events happening after 10 ms and no later than the subsequent frame acquisition at 22 ms. Here, we used the continuous mode to record tongue movements during speech productions and the triggered mode during recording of TMS-evoked responses. Continuous acquisitions were performed at 83 frames per second (12.048 ms interval between two consecutive images).

UTDI data acquisition, in the triggered mode, was synchronized by using an Arduino board [16]. A transistor-transistor logic pulse was sent to the UTDI machine (0.5 Hz) as the Philips CX30 can acquire single snapshots following a rhythmic input. The Arduino board also sent triggers to the psychtoolbox script controlling the audio stimuli and to the TMS machine (figure 1*d*). Timing of all triggers, data acquisition and audio output was checked beforehand by using an external I/O board with microsecond precision (Power 1401 CED, Cambridge Electronics, UK).

UTDI raw data were converted in portable network graphics (PNG) files by the Xmedcon software (Free Software Foundation). The PNG images were converted to the Hue Saturation Value

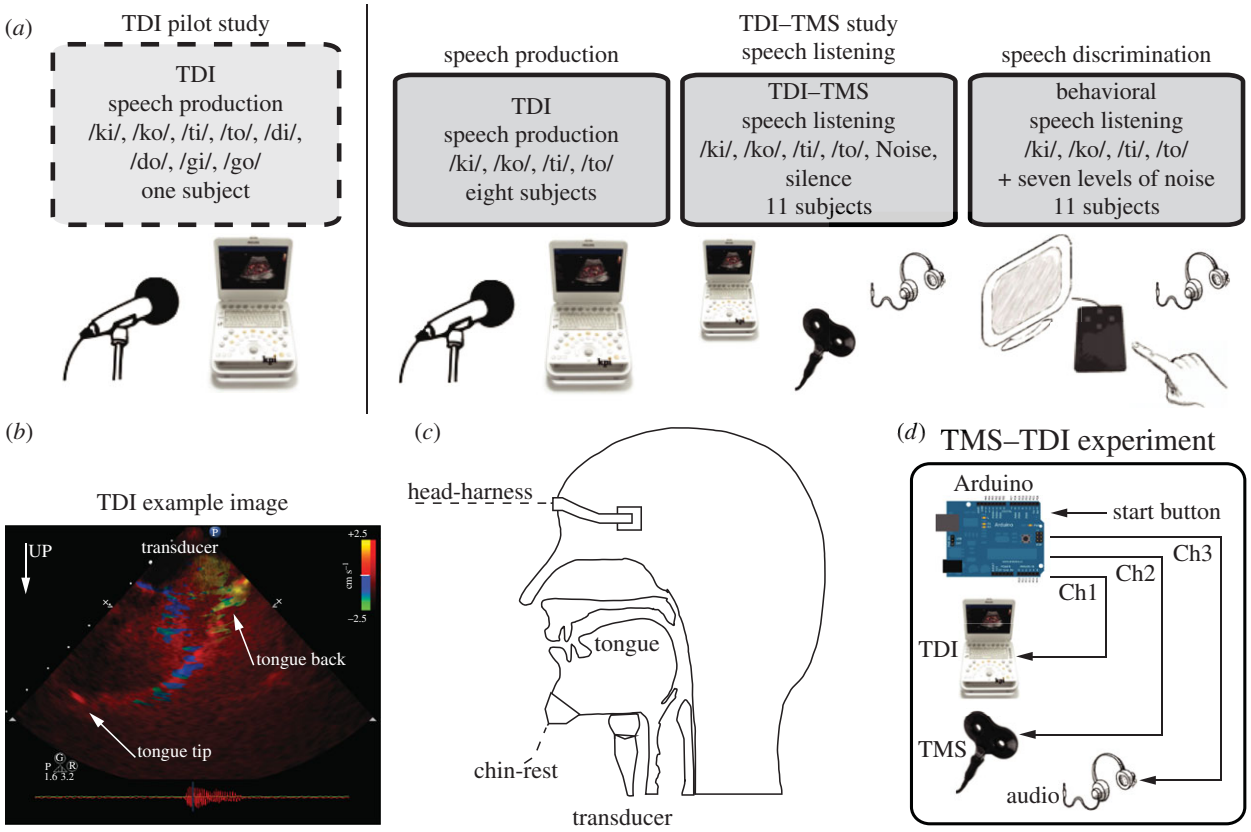

Figure 1. Experimental design and methods. The experiment was divided into a UTDI–pilot study and a UTDI–TMS study, as shown in panel (*a*). The pilot study consisted in recording UTDI speech production data on one subject. The UTDI–TMS study instead was divided into a UTDI recording during speech production and listening followed by a behavioural speech discrimination task. Panel (*b*) shows an example of UTDI image with superimposed colour representation of velocities. Panel (*c*) shows the positioning of the transducer and the systems employed to stabilize subjects' head. Finally, panel (*d*) shows a simplified schematic of how we synchronized the UTDI, TMS machine and audio presentation via an Arduino board. (Online version in colour.)

colour space, which allows colours to be represented with a single value from 0 to 360. Hue values were converted to new arbitrary values ranging between −60 and +60 arb. units, with positive values representing positive velocities (movement towards the transducer) and negative values representing negative velocities (movements away from the transducer).

(c) Ultrasound tissue Doppler imaging–pilot study

The pilot study was necessary to familiarize with the machine set-up, data export, data analyses and image synchronization. In fact, typical UTDI machines are designed for clinical purposes and for such a reason are very user-friendly, although lacking the flexibility of a research instrument. The complexities related to these kind of studies are mostly associated with data analysis and thus, in the pilot, we aimed at optimizing the following *UTDI–TMS study*. In fact, we aimed at delivering TMS-pulses when subjects were listening to stimuli that, from a motor point of view, were maximally different. In the pilot study, we recorded six repetitions for each of eight possible syllables (/ti/, /to/, /di/, /do/, /ki/, /ko/, /gi/, /go/). Three regions of interest (ROI) were examined, centred on the anterior tongue surface (ANT), the posterior tongue surface (POS) and the tongue body (BODY), as shown in figure 2. After extraction of positive and negative motions, we counted the number of positive and negative pixels for each ROI in each image (800 × 600 pixels) for every frame.

(d) Transcranial magnetic stimulation

TMS was delivered through a figure-of-eight coil (70 mm) using a monophasic stimulator (Bistim, Magstim Co., Whitland, UK). The left tongue primary motor cortex was first functionally localized by means of visual online inspection of triggered UTDI images. The best spot was identified by progressively decreasing TMS intensity during the mapping procedure [17] and marked on the scalp. Intensity of stimulation during the experiment was the lowest capable of evoking a detectable UTDI pattern in the tongue muscles, five times out of five consecutive pulses (mean 56.3% of stimulator output maximum intensity; s.d., 4.6). TMS was triggered through the Arduino system [16] controlled by custom-made software. TMS was delivered 200 ms after auditory stimuli onset. Previous studies [5] suggest that TMS stimulation applied between 100 and 300 ms after stimulus onset maximize motor response amplitude and selectivity. Such temporal window is also in agreement with the results of the pilot study (figure 2) showing the maximal difference between tongue positions starting around 200 ms after voice onset.

In the TMS study, the use of the UTDI-triggered mode was motivated by the short average latency of tongue MEPs (starting at 8 ms [3,10]) relative to the sampling rate of UTDI. The continuous mode cannot be triggered and thus, considering an 83 Hz acquisition frequency, the TMS-induced tongue twitch might have occurred between frames. Single images were thus acquired 10 ms after the delivery of the TMS, given the fact that the UTDI machine requires some time to acquire an image, velocity is referred to events happening after 10 ms and no later than 22 ms.

However, MEPs and tongue movements elicited by TMS are coupled and have different latency and duration [18]. In general, the electromechanical delay during voluntary movement is owing to biomechanical properties of the muscles and joints as well as the resistance of the measuring device [19]. In our case, we had no measuring device on the tongue to impede motion. Also, the tongue is characterized by peculiar biomechanical properties (e.g. no elastic component in series, type of motor units, weight/force ratio, etc.) that minimize such a delay. Furthermore, TMS-evoked twitches are generated by the activation

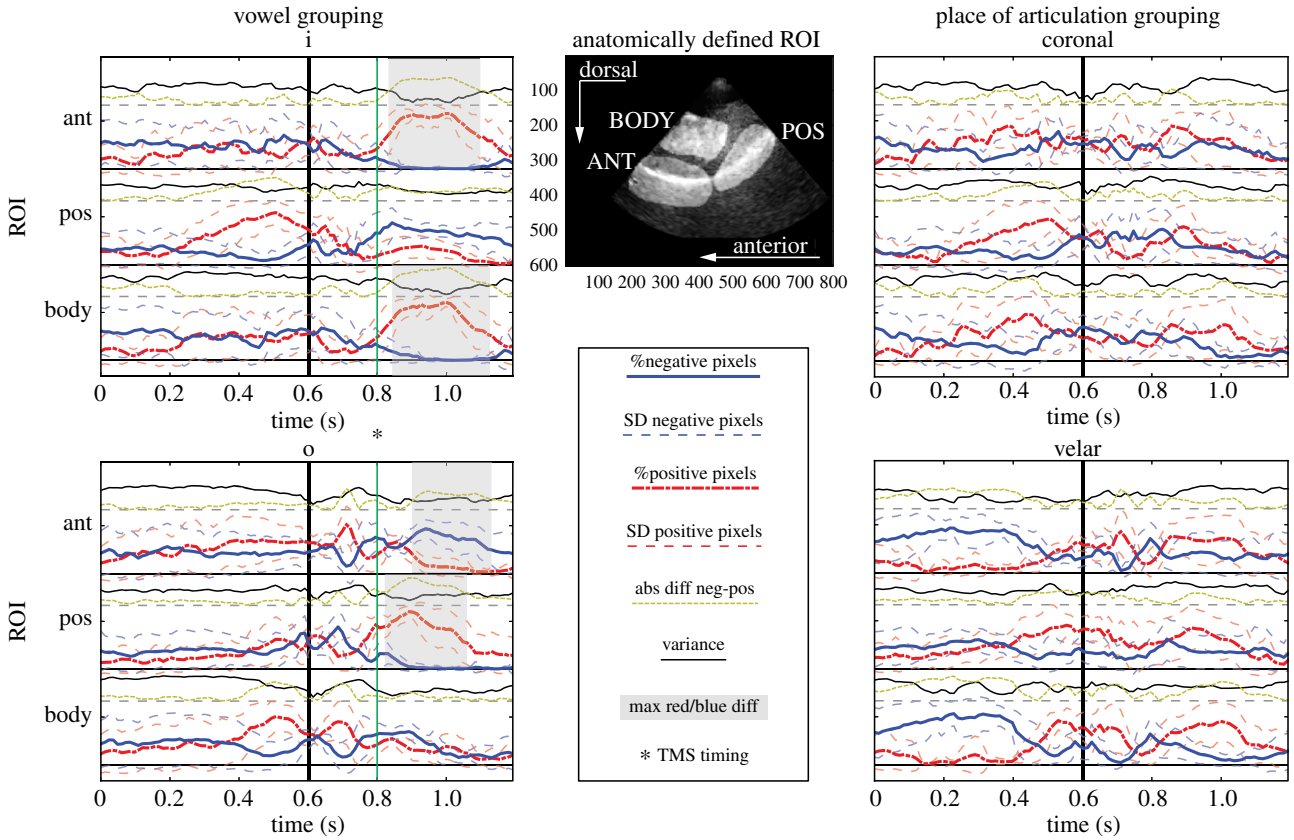

Figure 2. UTDI–pilot study. The UTDI–pilot study data were analysed separating three ROIs based on anatomy, as shown in the central panel. In four plots, we present the averaged UTDI speech production acquired in the pilot study. Pixel counts for negative motions are shown as a solid or blue line, for positive motions as a red or dash-dotted line. The standard deviations are shown in dashed lines. The solid black line shows the instantaneous total variance and the yellow or dotted line is the instantaneous absolute difference between the mean positive and negative pixel counts. The graph is more reliable when variance (solid black) is low and negative/positive (solid or blue/dash-dotted or red) difference is high (dotted or yellow), as shown by the grey shading. Data were grouped according to vowels (/i/ and /o/) on the left side and place of articulation (coronal and velar) on the right side. Green vertical lines, marked by an asterisk, show the choice of TMS timing. (Online version in colour.)

of many upper motor neurons in an almost synchronous manner, leading to a contraction that is faster than voluntary movements. Therefore, in our specific case, such an electromechanical delay is minimal, in the range of few milliseconds and, if we add that the UTDI machine itself needs some computing time to extract velocity data, 10 ms was the optimal time interval between triggers to TMS and UTDI.

Further confirmation that such a delay was in the range of few milliseconds come from two additional tests we run. In the first one, we measured MEPs co-registered with accelerometric data of the tongue (figure 3*a*), whereas in the second, high-speed tongue tip position data and MEPs (figure 3*b*). In the first test, we recorded the tongue MEPs together with tongue motion via an accelerometer in one subject. The three-axis analogue accelerometer was custom built by our electronic laboratory facility to be extremely light (only 20 g, 8 × 8 × 1.5 mm, to reduce the problems with measuring true electromechanical delays). We recorded 25 trials with both EMG (ZeroWire wireless system, sampling at 5 Khz) and synchronized accelerometer data (same sampling). We computed the vector norm of the accelerations on the three axis and then subtracted the average pre-TMS values. Please note, in figure 3*a* the negligible delay between MEP onset and accelerometric data onset. In the second test, we recorded the tongue MEPs together with tongue kinematics in another subject. We used an Optotrak Certus system (NDI, Inc., sampling at 1.2 Khz) with active markers. We recorded 25 trials with both EMG (ZeroWire wireless system, sampling at 2 Khz) and synchronized position data. We computed the vector norm of the positions on the three axis and then subtracted the average pre-TMS values. Please note, in figure 3*b* the negligible delay between MEP onset and position data. In both tests, TMS (monophasic stimulator; Magstim Co., Whitland, UK) was applied on the tongue motor area at 120% of the resting motor threshold. These tests were conducted on laboratory members according to international safety and ethical standards.

(e) Syllables stimuli

The stimuli (/ki/, /ti/, /ko/, /to/) are characterized by different points of articulation. The critical articulator is the tongue, as the velar phoneme /k/ requires a more posterior realization than the coronal phoneme /t/. When considering vowels, the realization of /i/ is more anterior than /o/. Their combinations produce a large variation in tongue movements, which was confirmed by the large differences visible in the pilot data (figure 2). During the *speech production* session, subjects were asked to utter the selected syllables. In the *speech listening* session, they had to passively listen to the same syllables. The stimuli were recorded from a male speaker with a professional microphone. The recordings (350 ms) were processed to reduce background noise and intensity-normalized using the freeware software AUDACITY (http://audacity.sourceforge.net/). In the *speech discrimination* session, subjects had to listen and recognize the same stimuli embedded in seven levels of grey-noise (5, 20, 35, 50, 65, 80 and 95%; see Experimental procedure).

(f) Experimental procedure

The TMS–UTDI study consisted of three sessions, which were conducted during the same day. In the *speech production session*, participants read aloud the four syllables while recording continuous UTDI data (see the electronic supplementary material, figure S1 for

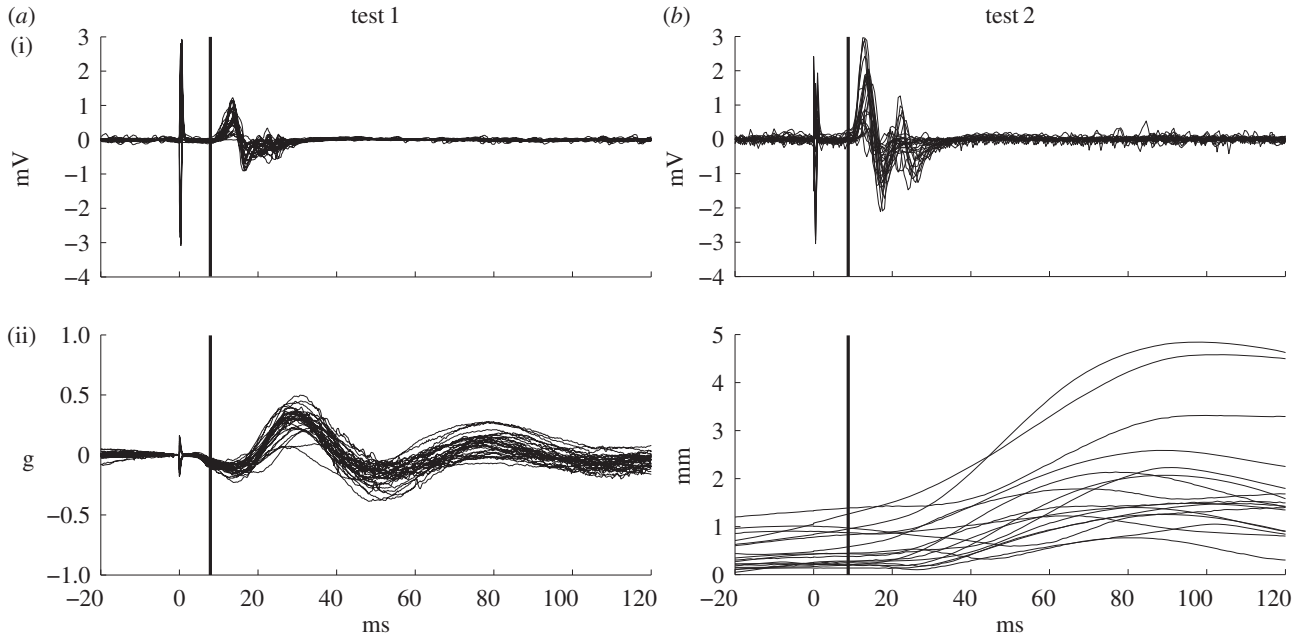

Figure 3. Pilot on tongue electromechanical delay. In test 1, we recorded in one subject the tongue MEPs together with tongue motion via an accelerometer. In (*a*), we show all MEPs (i) and all accelerometer data (ii). Please note the negligible delay between MEP onset and accelerometric data onset. In test 2, we recorded in another subject the tongue MEPs together with tongue position. In (*b*), we show all MEPs (i) and all vector norms of the position data (ii). Please note the negligible delay between MEP onset and movement onset.

an example of UTDI data during speech production). We recorded only eight subjects out of 11 because of technical problems with the equipment. Each syllable was repeated six times in a self-paced manner, with a rate of about 0.5 Hz.

In the *speech listening session*, subjects had to passively listen to auditory presented stimuli while tongue movements evoked by TMS were recorded through UTDI single snapshots mode. The stimuli consisted of six different audio files: four syllables /ti/, /ki/, /to/, /ko/ and two non-speech-related files as control conditions (white noise and silence). Audio stimuli were presented through headphones by using the psychtoolbox functions in MATLAB. Each trial consisted of the presentation of only one audio file, randomly repeated four to eight times with an inter-stimulus interval of 2 s. The TMS-pulse was delivered on the last repetition to avoid subjects' anticipation of TMS stimulation and at the same time being able to anticipate which was the syllable for that specific trial. Each condition (four syllables, noise and silence) was repeated 15 times for a total of 90 trials. Inter-trial interval ranged from 10 to 20 s. On 10% of the trials, they were asked to perform a one-back task (i.e. to repeat which was the last sound heard) in order to control for attention. When recording UTDI, subjects were leaning their chin on a chinrest and their forehead on an adjustable head harness (figure 1*c*). The transducer was placed externally along the inferior midline of the mandible, just anterior to the hyoid bone (figure 1*c*).

In the *speech discrimination session*, we measured subjects' ability to discriminate syllables in noise. The auditory syllables (/ko/, /ti/, ki/, ti/) were embedded in seven levels of increasing grey-noise (see Syllables stimuli). Audio files were randomly presented through earphones, by using the psychtoolbox functions of MATLAB (Mathworks, Inc.; 280 trials composed by 4 syllables × 7 noise levels × 10 repetitions). The task was to identify as fast and as accurately as possible the last presented syllable by pressing, with the right index finger, one of the four buttons associated with the auditory presentation of the syllables (button-stimuli association was counterbalanced across subjects). We measured reaction times (RTs) and accuracy. In the first training session, subjects learned how to use the response pad, supported by feedback about correctness (all four syllables at 5% of the grey-noise level, repeated three times each, for a total of 12 trials). The second training session was devised to familiarize with different noise levels in the stimuli, without feedback about the correctness. Finally, subjects did the task with the two extremes of noise levels (5 and 95%, for each of the four syllables, with two repetitions and a total of 16 trials).

(g) Analyses

(i) Transcranial magnetic stimulation–ultrasound tissue Doppler imaging

Each image collected during the *speech listening* session was preprocessed (see Ultrasound tissue Doppler imaging) to obtain numerical matrices (600 × 800 pixels) with positive or negative values coding for the intensity of the pixelwise velocity with respect to the transducer. Images displaying less than 1% of pixels with non-zero values were excluded from subsequent analysis (29 images excluded from a total of 990 images). Those images were characterized by almost no motion, which could have been induced by technical problems. At a single-subject level, each condition was averaged. Images for 'noise' and 'silence' conditions were averaged and used as a baseline. The baseline was then subtracted from the average images of the four syllables. For each of these resulting images, three features were extracted: mean velocity, negative percentage and positive percentage. Mean velocity is the average value of the non-zero pixels. Negative and positive percentage is the number of pixels with either negative or positive values respect to the total number of pixels in the image.

In the second step, each image was divided into four clusters on the basis of the spatial distribution of pixels in the whole image. Clustering was performed by using the *k*-mean function with the Hartigan–Wong algorithm. The four clusters were labelled as anterior, posterior, dorsal and ventral and coordinates of the geometrical centroids were calculated (figure 5*a*). Mean velocity, percentage of negative and positive pixels were calculated for each cluster.

All variables were analysed by means of a repeated measure analysis of variance, using as within-subject factors vowel (two levels: 'I', 'O') and consonant (two levels: 'K','T'). All analyses were run using the R statistical package [20].

(ii) Transcranial magnetic stimulation–ultrasound tissue Doppler imaging and speech discrimination

RTs were *z*-transformed at a single-subject level and analysed by means of a repeated measures analysis of variance, by using as within factors vowel (two levels: 'I', 'O'), consonant (two levels: 'K','T') and noise (seven levels: '5', '20', '35', '50', '65', '80' and '95%'). This analysis was run to assess the presence of an effect of the factor noise, which could have caused a decrease in the identification speed. As expected, the main effect for noise was significant ($F_{6,60} = 3.72$, $p < 0.01$), owing to an increase in RTs with increasing levels of noise (mean *z*-score $\pm$ s.e.: 5%, -0.286 ± 0.19; 20%, -0.274 ± 0.11; 35%, -0.228 ± 0.10; 50%, -0.047 ± 0.13; 65%, 0.106 ± 0.11; 80%, 0.341 ± 0.10; 95%, 0.387 ± 0.16). The main effect for vowel was also significant ($F_{1,10} = 15.18$, $p < 0.01$), caused by longer RTs for the identification of /i/ sounds (0.31 ± 0.07) with respect to /o/ sounds (-3.1 ± 0.07). No other main effects or interactions were found.

The central scope of this experiment was to measure individual differences in coping with different levels of noise when actively discriminating the experimental stimuli. In order to obtain an individual measure of the increase in RTs depending on noise level, we fitted a linear model using the *z*-transformed RTs as a dependent variable and noise as the independent variable. The linear fit was applied at the single-subject level and for each vowel and consonant, separately. The estimated slope resulting from the linear model fit was considered as an indicator of the degree of change of RTs depending on noise. This measure was then used to seek for correlations with TMS–UTDI results.

3. Results

(a) Transcranial magnetic stimulation–ultrasound tissue Doppler imaging

Every syllable was repeated a variable number of times (four to eight) and magnetic stimulation was delivered only during the last repetition. During no-TMS trials, images were acquired for each audio stimulus with the same timing. We extracted one image per trial, for all subjects, while listening to the syllable preceding and no TMS stimulation. The mean percentage of active pixels for each subject, while passively listening to syllables and no TMS, was always lower than 0.3% and on average 0.11%. Please note that such amount of activity is one order of magnitude smaller than our instrument artefact rejection criteria (Subject1: 0.017%; S2: 0.252%; S3: 0.051%; S4: 0.087%; S5: 0.294%; S6: 0.136%; S7: 0.113%; S8: 0.039%; S9: 0.028%; S10: 0.159%; S11: 0.126%).

For the TMS trials, the mean velocity of tongue showed no significant interaction ($F_{1,10} = 0.002$, $p = 0.96$) but significant consonant ($F_{1,10} = 6.498$, $p < 0.05$) and vowel main effects ($F_{1,10} = 5.011$, $p < 0.05$), with a prevalence of negative velocities for /k/ (-0.11 ± 0.15 arb. units) and /o/ sounds (-0.18 ± 0.11) and a prevalence of positive velocities for /t/ (0.06 ± 0.12) and /i/ sounds (0.13 ± 0.17) (figure 4*a*).

The percentage of negative pixels showed no significant interaction ($F_{1,10} = 1.81$, $p = 0.21$) but a significant consonant main effect ($F_{1,10} = 5.838$, $p < 0.05$) and no significant vowel main effect ($F_{1,10} = 3.786$, $p = 0.08$), with a larger percentage of negative pixels during the listening of /k/ sounds ($7.78 \pm 0.34\%$) with respect to /t/ sounds ($7.33 \pm 0.40\%$) together with a slight, but not significant, increase during the listening of /o/ sounds ($7.97 \pm 0.37\%$) with respect to /i/ sounds ($7.15 \pm 0.46\%$) (figure 4*c*).

The percentage of positive pixels showed no significant interaction ($F_{1,10} = 0.171$, $p = 0.69$), no significant consonant main effect ($F_{1,10} = 4.33$, $p = 0.064$) as well as no significant vowel main effect ($F_{1,10} = 2.5$, $p = 0.14$; figure 4*d*).

Analyses on the four regions identified by the *K*-means clustering analysis showed dissociable patterns of UTDI signals specific to these regions. In the posterior cluster, mean velocity showed no significant interaction ($F_{1,10} = 0.682$, $p = 0.43$) but a significant consonant main effect ($F_{1,10} = 5.628$, $p < 0.05$) and no significance for the vowel main effect ($F_{1,10} = 0.337$, $p = 0.57$), owing to a prevalence of negative velocities for the /k/ sound (-0.20 ± 0.16 arb. units) with respect to a substantially neutral balance between positive and negative motions in /t/ (0.03 ± 0.21), as shown in figure 5*e*.

In the ventral cluster, mean velocity showed no significant interaction ($F_{1,10} = 1.833$, $p = 0.21$) but a significant consonant main effect ($F_{1,10} = 5.683$, $p < 0.05$) and no significance for the vowel main effect ($F_{1,10} = 4.413$, $p = 0.062$) owing to a prevalence of positive velocities for the /t/ (0.21 ± 0.15 arb. units) and /i/ sounds (0.27 ± 0.19) with respect to a neutral balance between positive and negative motions in /k/ (-0.09 ± 0.15) and a negative prevalence in /o/ (-0.15 ± 0.15), as shown in figure 5*d*.

In the dorsal cluster, the percentage of positive pixels showed no significant interaction ($F_{1,10} = 0.721$, $p = 0.42$), no significant consonant main effect ($F_{1,10} = 1.142$, $p = 0.31$) but a significant vowel main effect ($F_{1,10} = 5.536$, $p < 0.05$), showing greater percentage of positive pixels during the listening of /i/ ($2.34 \pm 0.14\%$ of pixels) with respect to /o/ sounds ($1.97 \pm 0.15\%$), as shown in figure 5*c*.

Moreover, the distance between the centroid of the posterior cluster and the most anterior portion of active tongue muscle on the antero-posterior axis was calculated and analysed, showing no significant interaction ($F_{1,10} = 0.842$, $p = 0.38$), no significant consonant main effect ($F_{1,10} = 0.95$, $p = 0.35$) but a vowel main effect ($F_{1,10} = 4.994$, $p < 0.05$). The effect was owing to a greater distance between these points (as shown in figure 5*b*), thus showing a more anterior extension for the tongue, during the listening of syllables containing the /i/ (281.2 ± 13.1 pixels) with respect to the /o/ sounds (276.6 ± 12.7 pixels).

(b) Transcranial magnetic stimulation–ultrasound tissue Doppler imaging and speech discrimination during speech discrimination

The slope values obtained from the linear fit between RTs and noise levels at individual-subject level were correlated with the variables that showed a modulation in the TMS–UTDI experiment. No correlations were found between slope values and mean velocity or the percentage of negative pixels. A strong positive correlation between slope and the percentage of positive pixels was found for /o/ ($r = 0.75$, $p < 0.01$) and /k/ sounds ($r = 0.75$, $p < 0.001$), to a lesser extent for /i/ sounds ($r = 0.62$, $p < 0.05$) and *no* significant correlation for /t/ sounds ($r = 0.55$, $p = 0.08$). The presence of these correlations means that the more the participants were affected by noise in speech discrimination the greater was the percentage of positive pixels evoked by TMS for the same stimulus.

As the second step, we focused on the significant effects shown in the clustered data. We correlated RT slope values

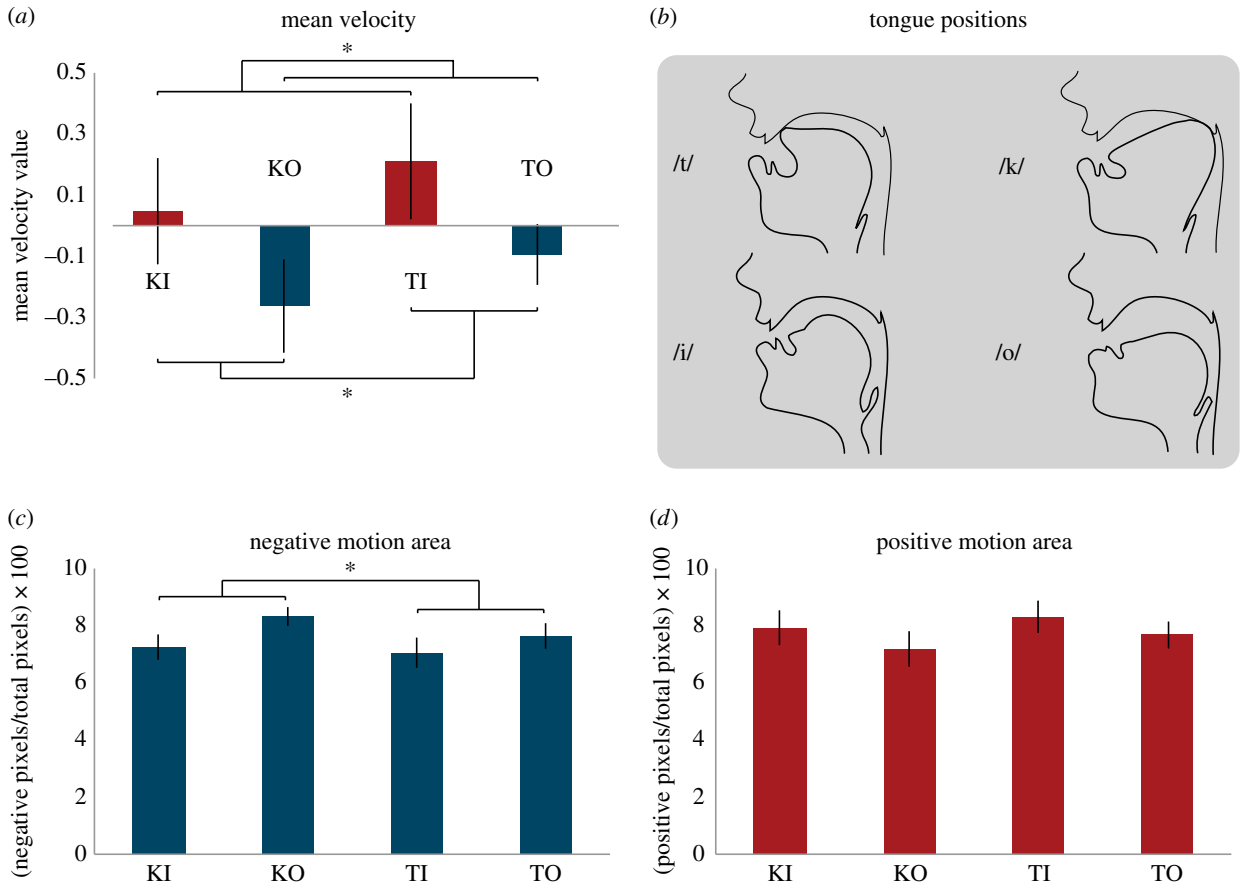

Figure 4. UTDI–TMS results. The figure shows the UTDI–TMS results by measuring the mean velocity (*a*). Panel (*b*) shows the schematic of tongue positions for each of the two consonants (/t/ and /k/) and two vowels (/i/ and /o/). The lower part of the figure contains the negative (*c*) or positive (*d*) motion. Area of motion is defined as the percentage positive or negative pixels over the total number of pixels. Asterisks denote significant comparisons and bars represent the standard error of the mean. (Online version in colour.)

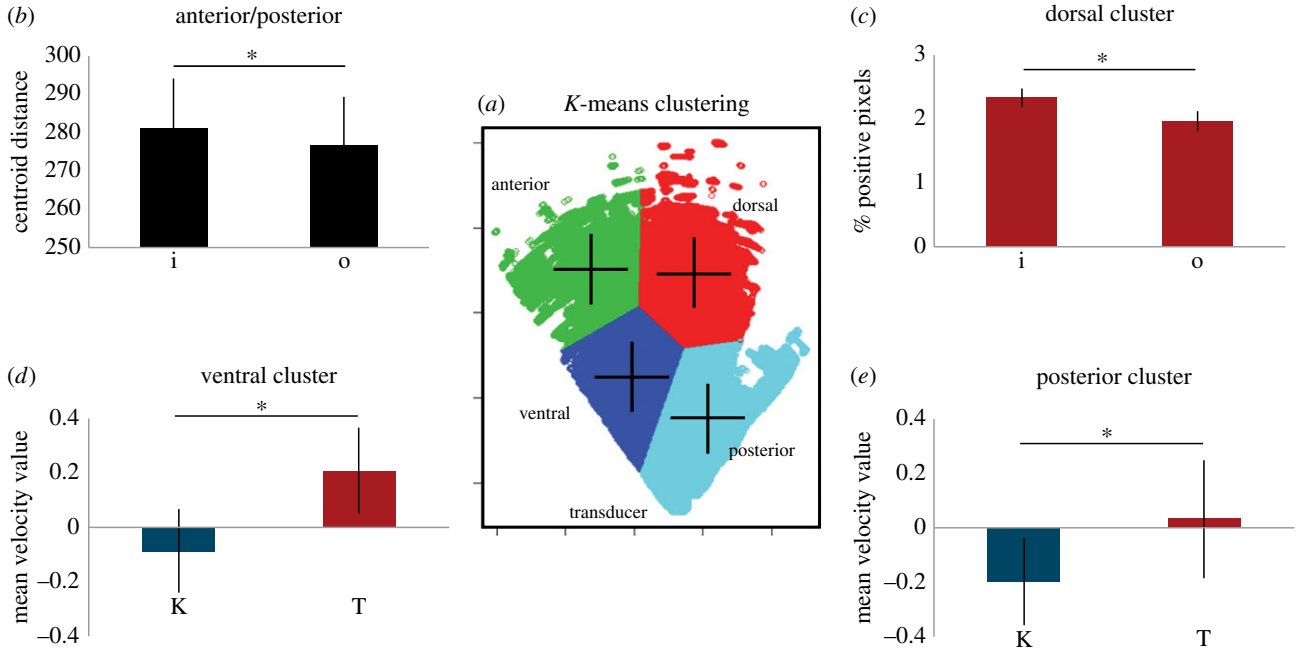

Figure 5. Tongue regional kinematic differences. Panel (*a*) shows the clustering results in an exemplar UTDI–TMS image. *K*-means clustering separated the images into a dorsal, ventral, anterior and posterior clusters based only on spatial distribution of data. Crosses represent the centroid of each cluster. Note that here the tongue image has been rotated. Panel (*b*) shows the distance between the posterior cluster and the most anterior activity in the anterior cluster for the two vowels /i/ and /o/. Panel (*c*) shows the percentage of positive pixels in the dorsal cluster for the two vowels. Panels (*d*,*e*) show the mean velocity in the ventral and posterior clusters, respectively, for the two consonants (/t/ and /k/) separately. Asterisks denote significant comparisons and bars represent the standard error of the mean. (Online version in colour.)

with mean velocity in the posterior and ventral clusters. No correlation could be detected in the posterior cluster, whereas the ventral one showed a significant positive correlation for /k/ sounds ($r = 0.70$, $p < 0.05$); figure 6*b*. Individuals affected by noise in speech discrimination tended to have more positive velocities during the listening of /k/ sounds in the TMS experiment. This reflects the tendency to cancel the differential pattern between /k/ and /t/ sounds.

Finally, the RT slope values correlated with the percentage of positive pixels in the dorsal cluster. A significant correlation was found for /o/ sounds ($r = 0.60$, $p < 0.05$) and no significance was found for /i/ sounds ($r = 0.58$, $p = 0.06$); figure 6*a*. Again, this correlation points to a decreased TMS–UTDI pattern specificity for those participants who were less resistant to noise in the discrimination experiment.

4. Discussion

The idea that listener's motor representations contribute to speech perception has been the central prediction of several theoretical models [21–23]. By contrast, several competing theories have suggested that a purely sensory analysis is sufficient for speech classification [24]. The discovery of mirror neurons [7] has provided the direct neural demonstration that the motor system becomes activated during perception. This evidence has been shown valid for speech as well by TMS studies on motor facilitation during speech perception, as revealed by tongue EMG [3,25].

Here, we show a new method that combines TMS and UTDI to gain a new level of detail when studying speech-induced motor mirroring. In fact, previous TMS recording of corticobulbar excitability could only show a rather coarse picture of the activation in the motor system [3–6,26,27]. UTDI, instead, can visualize local tongue motion with great spatial detail. We were thus able to reliably separate patterns of movements evoked by TMS during listening to four different syllables (figure 4*a*) varying in place of articulation (coronal versus velar) and position in vowel space (frontal versus posterior).

The pattern of motor-evoked activities was in agreement with the expected direction of velocities. In fact, the pilot-UTDI study (figure 2) showed movements towards the transducer (red shifts in two out of three ROIs) for the /i/, as opposed to that of the /o/ vowel (blue (solid) line shift in one ROI, red (dash-dotted) line shift in one ROI) starting at 200 ms after speech onset. Accordingly, TMS applied at 200 ms evoked more positive (red or dash-dotted) activity during the listening of syllables containing the /i/ as opposed to negative (blue or solid) activity evoked by the /o/ sounds. Interestingly, we could elicit a pattern that is specific both in terms of motion type and temporal deployment, suggesting that the motor system, during passive speech listening, employs a mirroring strategy that is extremely accurate.

When looking at the different clusters, a larger positive motion for /i/ with respect to /o/ was present in the dorsal area (figure 5*c*) suggesting that the global movement towards the transducer was mostly accounted by the postero-dorsal aspect of the tongue. This is in agreement with tongue position during the vocalization of /i/ where the anterior part moves anteriorly and the tongue back is lowered (figure 4*b*). Here, for the /i/ sound, we observe the tongue back moving towards the transducer, whereas motion towards the alveolar ridge is perpendicular to the recording probe and thus invisible to

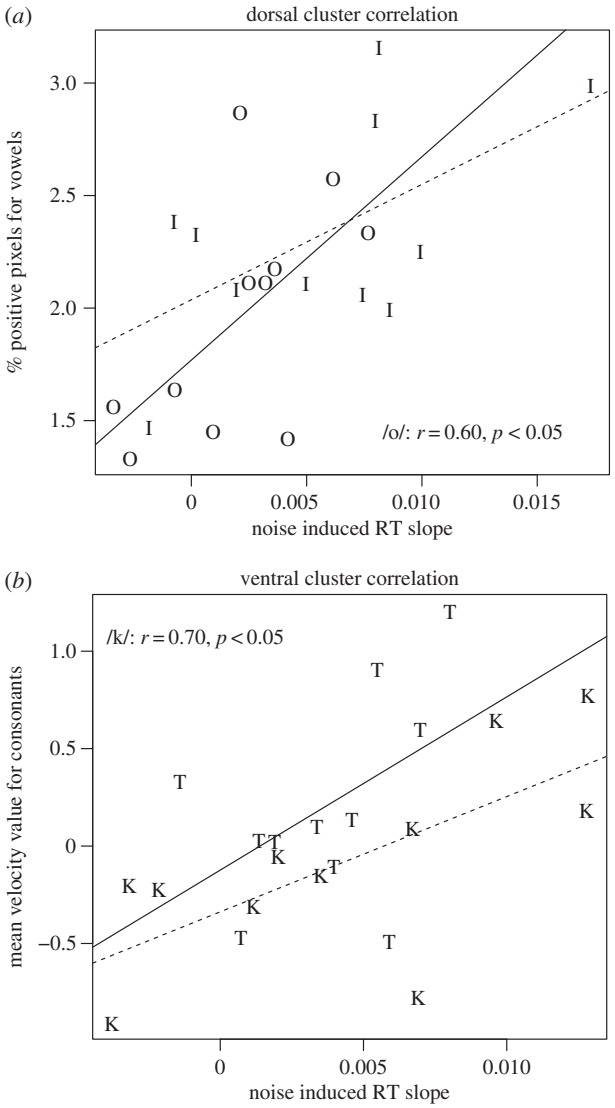

Figure 6. Correlation with speech discrimination. The figure shows the correlation between individual RT data from the speech discrimination study and significant features extracted from the UTDI–TMS study. RT noise-dependent slopes (*a*) correlated with the percentage of positive pixels in the dorsal cluster for the /o/ vowel. Panel (*b*) shows instead the significant correlation with the mean velocity values for the consonant /k/.

the machine. However, the antero-posterior extension of the tongue could be derived by measuring cluster centroid distances. Such a measure was significantly larger when listening to /i/ as opposed to /o/ (figure 5*b*).

The differences between coronal and velar sounds were on larger negative values for /k/ as opposed to /t/ sounds (figure 4*a*,*c*). The global larger negative motion for /k/ was clarified by the clustered analyses. In fact, most of the negative shift was accounted by the posterior cluster and to a minor extent by the ventral cluster. Such pattern of spatial segregation of velocities is in agreement with speech production patterns. In fact, velar sounds production requires the raising of the posterior part of the tongue towards the velum and away from the transducer (figure 4*b*), which is exactly what we observed when subjects listened to /k/ as opposed to /t/.

Summing up, in this work we describe a complex regional pattern of movements elicited by TMS during passive listening to speech sounds, which are in detailed agreement with known articulatory descriptions of speech production. This evidence was provided by using TMS in conjunction with

UTDI to extract complex motor synergies from tongue kinematics with an unprecedented level of fine-grained details. Previous research demonstrated that single-pulse TMS can elicit complex patterns of hand movement synergies representing the most frequent grasping configurations [28]. However, this is the first time, to our knowledge, that TMS-induced tongue motor synergies are described and that specific motor patterns can be selected by passive listening speech. Such an improvement in descriptive power shows a clear theoretical advancement with respect to previous studies [3–6,26,27] and opens up to more studies that can be implemented with such a technology. In fact, the selective recruitment of tongue motor synergies during passive listening to speech is the critical prediction of any mirroring hypothesis on brain speech sound classification.

In addition to that, we also show that motor recruitment during speech perception predict subjects' ability to discriminate speech in noise. From a computational perspective, the role of the motor system can be that of compensating for the increased ambiguity of the stimuli [29,30]. Subjects that are more impaired by increasing levels of noise should be in principle relying less on a motor compensation strategy. Here, we show that subjects with a decreased ability to discriminate speech in high levels of noise were those with globally more positive activations in the UTDI–TMS images. Correlation run on clustered data shows that such increase in positive motion was in the direction of reducing differences across motor-evoked patterns (figure 6). This latter results suggest that subjects relying less on a motor simulative strategy when actively discriminating speech, were those that evoked less detailed motor patterns when passively listening to the same material.

As a matter of fact, our previous investigations, with a TMS interference paradigm [31–34], show that the contribution of the motor system to speech discrimination is associated with the signal-to-noise ratio. We suggested that the motor system contributes to speech discrimination, with a process of stimuli reconstruction, when the signal is ambiguous [34]. Such hypothesis was based only on behavioural data (RTs and accuracy). In this study, we find that greater efficiency in discriminating speech in noise is associated with the activation of very detailed motor synergies when passively listening. This is a strong neurophysiological confirmation that the degree of articulatory mirroring is associated with the amount of environmental noise and variability. This result is in agreement with our previous data [31,32] and gives important support to the hypothesis that mirroring in speech is an active process of distal articulatory inference in service of stimuli discrimination.

Motor theories of speech perception started from different assumptions but all converged on the prediction that motor activities should have been observed during passive listening to speech [21–23]. This was long before modern neuroimaging could show that, as it was possible with TMS, such a prediction was true. With the present result, we move one step further by showing that the motor system, during speech listening, replicates the complexity of the motor patterns involved in production. From our results, it emerges that the motor system does not just resonate passively to global effectorwise features [3,4] but that it rather implements a refined simulation of the speaker motor control features. Moreover, the granularity level of the evoked motor representation may vary between individuals and could explain individual differences in discriminating speech in noise. Indeed, this is an additional and radically novel proof that, when dealing with ambiguous speech signals, we implement a compensation strategy that is specifically motor rather than a-modal.

All subjects in the study gave informed consent to the experimental procedures according to the Declaration of Helsinki and the local ethics committee.

Funding statement. This work was supported by European Community grants SIEMPRE (ICT-FET project number 250026) and POETICON++ (STREP project ICT-288382).

References

1. Sebanz N, Bekkering H, Knoblich G. 2006 Joint action: bodies and minds moving together. *Trends Cogn. Sci.* **10**, 70–76. (doi:10.1016/j.tics.2005.12.009)
2. D'Ausilio A, Badino L, Li Y, Tokay S, Craighero L, Canto R, Aloimonos Y, Fadiga L. 2012 Leadership in orchestra emerges from the causal relationships of movement kinematics. *PLoS ONE* **7**, e3575. (doi:10.1371/journal.pone.0035757)
3. 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. (doi:10.1046/j.0953-816x.2001.01874.x)
4. Watkins KE, Strafella AP, 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)
5. Roy AC, Craighero L, Fabbri-Destro M, Fadiga L. 2008 Phonological and lexical motor facilitation during speech listening: a transcranial magnetic stimulation study. *J. Physiol.* **102**, 101–105. (doi:10.1016/j.jphysparis.2008.03.006)
6. D'Ausilio A, Jarmolowska J, Busan P, Bufalari I, Craighero L. 2011 Tongue corticospinal modulation during attended verbal stimuli: priming and coarticulation effects. *Neuropsychologia* **49**, 3670–3676. (doi:10.1016/j.neuropsychologia.2011.09.022)
7. di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G. 1992 Understanding motor events: a neurophysiological study. *Exp. Brain Res.* **91**, 176–180. (doi:10.1007/BF00230027)
8. Beautemps D, Badin P, Bailly G. 2001 Linear degrees of freedom in speech production: analysis of cineradio- and labio-film data and articulatory-acoustic modeling. *J. Acoust. Soc. Am.* **109**, 2165–2180. (doi:10.1121/1.1361090)
9. Sanders I, Mu L. 2013 A three-dimensional atlas of human tongue muscles. *Anat. Rec.* **296**, 1102–1114.
10. Paradiso GO, Cunic DI, Gunraj CA, Chen R. 2005 Representation of facial muscles in human motor cortex. *J. Physiol.* **567**, 323–336. (doi:10.1113/jphysiol.2005.088542)
11. Cruccu G, Inghilleri M, Berardelli A, Romaniello A, Manfredi M. 1997 Cortical mechanisms mediating the inhibitory period after magnetic stimulation of the facial motor area. *Muscle Nerve* **20**, 418–424. (doi:10.1002/(SICI)1097-4598(199704)20:4<418:: AID-MUS3>3.0.CO;2-D)
12. Svensson P, Romaniello A, Wang K, Arendt-Nielsen L, Sessle BJ. 2006 One hour of tongue-task training is associated with plasticity in corticomotor control of the human tongue musculature. *Exp. Brain Res.* **173**, 165–173. (doi:10.1007/s00221-006-0380-3)
13. Svensson P, Romaniello A, Arendt-Nielsen L, Sessle BJ. 2003 Plasticity in corticomotor control of the human tongue musculature induced by tongue-task training. *Exp. Brain Res.* **152**, 42–51. (doi:10.1007/s00221-003-1517-2)
14. Saigusa H, Saigusa M, Aino I, Iwasaki C, Li L, Niimi S. 2006 M-mode color Doppler ultrasonic imaging of

vertical tongue movement during articulatory movement. *J. Voice* **20**, 38–45. (doi:10.1016/j.jvoice.2005.01.003)

15. Ostry JD, Keller E, Parush A. 1983 Similarities in the control of the speech articulators and the limbs: kinematic of tongue dorsum movement in speech. *J. Exp. Psych. Hum. Perc. Perf.* **9**, 622–636. (doi:10.1037/0096-1523.9.4.622)
16. D'Ausilio A. 2012 Arduino: a low-cost multi purpose lab equipment. *Behav. Res. Methods* **44**, 305–313. (doi:10.3758/s13428-011-0163-z)
17. Rossini PM *et al*. 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. Neuro.* **91**, 79–92. (doi:10.1016/0013-4694(94)90029-9)
18. Norman RW, Komi PV. 1979 Electromechanical delay in skeletal muscle under normal movement conditions. *Acta Physiol. Scand.* **106**, 241–248. (doi:10.1111/j.1748-1716.1979.tb06394.x)
19. Corcos DM, Gottlieb GL, Latash ML, Almeida GL, Agarwal GC. 1992 Electromechanical delay: an experimental artifact. *J. Electromyogr. Kinesiol.* **2**, 59–68. (doi:10.1016/1050-6411(92)90017-D)
20. R Development Core Team. 2008 *R: a language and environment for statistical computing*. Vienna Austria: R Foundation for Statistical Computing.
21. Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. 1967 Perception of speech code. *Psychol. Rev.* **74**, 431–461. (doi:10.1037/h0020279)
22. Stevens KN, Halle M. 1967 Remarks on analysis by synthesis and distinctive features. In *Models for the perception of speech and visual form* (eds W Wathen-Dunn, LE Woods), pp. 88–102. Cambridge, UK: MIT Press.
23. Fowler CA. 1986 An event approach to the study of speech perception from a direct-realist perspective. *J. Phon.* **14**, 3–28.
24. Diehl RL, Lotto AJ, Holt L. 2004 Speech perception. *Annu. Rev. Psych.* **55**, 149–179. (doi:10.1146/annurev.psych.55.090902.142028)
25. Galantucci B, Fowler CA, Turvey MT. 2006 The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* **13**, 361–377. (doi:10.3758/BF03193857)
26. Murakami T, Restle J, Ziemann U. 2011 Effective connectivity hierarchically links temporoparietal and frontal areas of the auditory dorsal stream with the motor cortex lip area during speech perception. *Brain Lang.* **122**, 135–141. (doi:10.1016/j.bandl.2011.09.005)
27. 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)
28. Gentner R, Classen J. 2006 Modular organization of finger movements by the human central nervous system. *Neuron* **52**, 731–742. (doi:10.1016/j.neuron.2006.09.038)
29. Castellini C, Badino L, Metta G, Sandini G, Tavella M, Grimaldi M, Fadiga L. 2011 The use of phonetic motor invariants can improve automatic phoneme discrimination. *PLoS ONE* **6**, e24055. (doi:10.1371/journal.pone.0024055)
30. Canevari C, Badino L, D'Ausilio A, Fadiga L, Metta G. 2013 Modeling speech imitation and ecological learning of auditory-motor maps. *Front. Psych.* **4**, 364. (doi:10.3389/fpsyg.2013.00364)
31. D'Ausilio A, Pulvermúller 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)
32. 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)
33. Bartoli E, D'Ausilio A, Berry J, Badino L, Bever T, Fadiga L. In press. Listener-speaker perceived distance predicts the degree of motor contribution to speech perception. *Cereb. Cortex*.
34. D'Ausilio A, Craighero L, Fadiga L. 2012 The contribution of the frontal lobe to the perception of speech and language. *J. Neurolinguist.* **25**, 328–335. (doi:10.1016/j.jneuroling.2010.02.003)