RESEARCH ARTICLE

Sensitivity of occipito-temporal cortex, premotor and Broca's areas to visible speech gestures in a familiar language

Vincenzo Maffei[1,2,3☯], **Iole Indovina**[1,4☯*], **Elisabetta Mazzarella**[1], **Maria Assunta Giusti**[2], **Emiliano Macaluso**[5,6], **Francesco Lacquaniti**[1,2], **Paolo Viviani**[1,2]

1 Laboratory of Neuromotor Physiology, IRCCS Santa Lucia Foundation, Rome, Italy, **2** Centre of Space BioMedicine and Department of Systems Medicine, University of Rome Tor Vergata, Rome, Italy, **3** Data Lake & BI, DOT - Technology, Poste Italiane, Rome, Italy, **4** Departmental Faculty of Medicine and Surgery, Saint Camillus International University of Health and Medical Sciences, Rome, Italy, **5** ImpAct Team, Lyon Neuroscience Research Center, Lyon, France, **6** Laboratory of Neuroimaging, IRCCS Santa Lucia Foundation, Rome, Italy

☯ These authors contributed equally to this work.
* i.indovina@hsantalucia.it

OPEN ACCESS

Citation: Maffei V, Indovina I, Mazzarella E, Giusti MA, Macaluso E, Lacquaniti F, et al. (2020) Sensitivity of occipito-temporal cortex, premotor and Broca's areas to visible speech gestures in a familiar language. PLoS ONE 15(6): e0234695. https://doi.org/10.1371/journal.pone.0234695

Editor: José A Hinojosa, Universidad Complutense Madrid, SPAIN

Received: February 27, 2019

Accepted: June 1, 2020

Published: June 19, 2020

Data Availability Statement: The dataset is uploaded to Zenodo and accessible via the following URL: https://doi.org/10.5281/zenodo.3881470.

Funding: This work was supported by the Italian Ministry of Health (Ricerca corrente, IRCCS Fondazione Santa Lucia and PE-2013-02355372), Italian Ministry of University and Research (PRIN grant 2017CBF8NJ_005), Italian Space Agency (I/006/06/0 grant and 2019-11-U.0).

Abstract

When looking at a speaking person, the analysis of facial kinematics contributes to language discrimination and to the decoding of the time flow of visual speech. To disentangle these two factors, we investigated behavioural and fMRI responses to familiar and unfamiliar languages when observing speech gestures with natural or reversed kinematics. Twenty Italian volunteers viewed silent video-clips of speech shown as recorded (Forward, biological motion) or reversed in time (Backward, non-biological motion), in Italian (familiar language) or Arabic (non-familiar language). fMRI revealed that language (Italian/Arabic) and time-rendering (Forward/Backward) modulated distinct areas in the ventral occipito-temporal cortex, suggesting that visual speech analysis begins in this region, earlier than previously thought. Left premotor ventral (superior subdivision) and dorsal areas were preferentially activated with the familiar language independently of time-rendering, challenging the view that the role of these regions in speech processing is purely articulatory. The left premotor ventral region in the frontal operculum, thought to include part of the Broca's area, responded to the natural familiar language, consistent with the hypothesis of motor simulation of speech gestures.

Introduction

Watching the mouth movements of a speaker (so called, visual speech) may help listeners to decode speech in a noisy environment [1–3], and may even alter the auditory perception of speech as in the McGurk effect [4–9].

Observers can discriminate fairly reliably between silent video-clips of a speaker played as recorded (Forward mode) or time-reversed (Backward mode) [1]. It was argued that natural kinematics (recognition of biological motion) rather than linguistic competences had a role in this task [1].

Competing interests: The authors have declared that no competing interests exist.

Normal and time-reversed visual speech differ kinematically in several ways, although the qualitative differences are subtle. With few exceptions (such as long vowels or fricative consonants), phono-articulatory gestures tend to be asymmetric in time. For instance, deceleration phases are longer than acceleration phases [10], and asymmetries are present between the opening and closing movements of the mouth [11]. Moreover, articulatory gestures of speech obey specific constraints imposed by the motor system. Thus, the temporal inversion of these gestures often generates a sequence of unnatural movements hardly repeatable by normal people, although an experienced person can invert the temporal order of the phonemes in a sentence. In fact, the articulatory sequences that generate the phonemes during a speech are extremely complicated to perform in reverse. This is because one should reverse each phono-articulatory manoeuvre required to produce a given phoneme, as well as the specific sequence with which these manoeuvres are chained during the speech [1].

The central nervous system is also sensitive to language familiarity in visual speech. Indeed, a familiar language can be discriminated by the analysis of the speech temporal structure (i.e., rhythm) in auditory as well as in visual modality [12–14]. Temporal duration and variability of vowels and consonants differ between languages [15–20], and the timing of vowels and consonants can be visually assessed since phono-articulatory gestures generating these movements fit into different visual classes [2,12,21–24]. For instance, Spanish monolingual speakers visually distinguished Spanish from Catalan, while this was not possible either for English or for Italian speakers [12,25].

Discrimination ability of a familiar language persists also after a temporal reversal of visual speech stimuli [12]. The rhythmic and global timing structures of speech visible cues (e.g., alternation of consonants and vowels, vowels duration) remain relatively unaltered after a temporal inversion of speech sequences, while semantic, lexical and phonotactic information are lost [12]. Moreover, six-months-old infants are able to discriminate a familiar language from visual speech [26]. These observations suggest that visual spatio-temporal cues play a more important role in identifying familiarity than linguistic competence. The brain networks involved in these processes are unknown. To our knowledge, no study so far has directly investigated the neural correlates of language discrimination of visual speech, while the few papers reporting brain sites activated by inverting the temporal order of natural visual speech measured brain activity using techniques (PET or MEG) other than fMRI [27,28].

In theory, the occipito-temporal cortex (OTC) might have a specific selectivity to the spatio-temporal features of visual speech (i.e., kinematics of biological motion). Indeed, various foci in this region respond to different types of human movements and body forms [29–35]. Studies comparing face movements during a speech with facial movements that cannot be construed as speech reported activations in both lateral and ventral OTC, including the temporal visual speech area (TVSA) [36–38], as well as in auditory association areas of the temporal cortex in the superior temporal gyrus [37,39,40]. However, this comparison might be affected by the presence of confounds in low-level visual features, such as differences in motion speed [38,41]. A contrast immune from low-level visual confounds is the comparison of speech movements rendered normally (Forward) versus time-reversed (Backward).

A recent MEG experiment showed that, during the processing of silently played lip movements, the visual cortex tracks the missing acoustic speech information when played forward as compared to backward, indicating a top-down modulatory control of auditory dorsal stream on visual areas [42]. Also, in a PET study, the contrast Forward versus Backward engaged OTC bilaterally [27]. However, the ability to discriminate plausible speech gestures (i.e., Forward versus Backward video clips) was localised to later stages of processing, such as the parieto-temporal cortex and motor areas in the frontal cortex.

Visual speech stimuli have been shown to engage cortical motor areas involved in speech production, such as the left inferior frontal gyrus (IFG) that includes Brodmann's Area BA 44 and BA 45 (pars opercularis and triangularis of IFG, respectively) thought to overlap with Broca's region [43] and the ventral premotor cortex (PMv), but also more dorsal regions of the premotor cortex (PMd) [27,40,44–49]. Importantly, part of the frontal areas implicated in the control of movements and speech are connected with the visual cortex [50,51]. The motor theory of speech perception [52–54] proposed that the activation of motor speech areas during the observation of speech might represent an implicit motor simulation of the observed gestures conducive to speech understanding [44,55–57].

However, several authors questioned the idea that an automatic engagement of motor areas, such as IFG, during perceptual or cognitive task is evidence of a specific involvement of the motor system in perceptual or cognitive processes [28,58–64]. The dorsal premotor cortex, rather than the Broca's area (BA 44/45), seems to be engaged both in the execution and the observation of speech gestures [62]. Conversely, it was found that the activity in the IFG correlates with hit-rate and response bias during speech perception tasks [27,65]. Since response bias and hit rate are characteristic indexes of the decisional process, these findings might suggest that high level processes related to the generation of the response decision (e.g. whether to respond Yes or No), rather than motor simulations occurred in the IFG during visual speech [66,67]. In summary, the specific role of IFG and Broca's area in the functional architecture of speech perception remains open to debate [43,68].

In the present study, we investigated the neural circuits engaged by language familiarity (Italian vs Arabic) and natural kinematics of biological motion (Forward vs Backward) of visible speech. Italian observers viewed silent video-clips of the mouth movements of Italian and Arabic actors speaking in their native language. Stimuli were rendered either in normal (Forward) mode or after a time reversal (Backward). During an fMRI session, participants were asked to identify the rendering mode (Forward or Backward). The brain regions sensitive to language familiarity and those sensitive to natural mouth movements of speech were identified through the fMRI contrasts Italian vs Arabic (main effect of language) and Forward vs Backward (main effect of rendering mode), respectively. We also computed the interaction between the two main factors by means of the contrast "language x rendering mode" [(Italian Forward vs Italian Backward) vs (Arabic Forward vs Arabic Backward)]. The latter contrast should identify the areas where the effects of Familiarity (Italian vs Arabic) was larger for natural (Forward) than non-natural (Backward) motion (i.e biological motion).

Methods

Participants

Forty healthy right-handed Italian volunteers took part in this study. Twenty participants (14 females, 6 males; mean age: 25 years; age range: 20–42 years) were tested in a preliminary experiment and twenty different participants (13 females, 7 males; mean age: 23 years; age range: 20–35) in the main study (i.e. fMRI and in a follow-up experiment, see later). All participants had normal or corrected to normal vision. None of them had any familiarity with the Arabic language or experience with lip-reading. Written informed consent to procedures approved by the Institutional Review Board of Fondazione Santa Lucia was obtained from each participant. Experimental protocols complied with the Declaration of Helsinki on the use of human subjects in research.

Stimuli

Ten adults (5 females, 5 males) native speakers of Arabic and 10 adults (5 females, 5 males) native speakers of Italian volunteered as actors for generating the stimuli. We chose the Arabic rather than other European languages in order to ensure that the participants would have not been exposed more than occasionally to this language before, and we verified that this was the case. The choice of the Arabic was also motivated because it differed from the Italian more than did most other European languages. Indeed, articulatory movements for the production of words in Arabic and Italian languages are different [18,20].

Two of the authors (P.V. and V.M.) selected Italian speakers so that, after careful visual inspection, the general features of the lower part of the face were roughly similar to those of the Arabic speakers previously selected. None of the actors participated in the main or in the preliminary study. Each actor read four texts excerpted from newspapers in his/her native language (Arabic, A or Italian, I). The preparation of the stimuli involved three steps. First, we recorded the lower part of the face (including the upper/lower lips and the chin) of each actor with a digital camera (25 frames/s, Sony HDR-SR-8E), and stored the results as a sequence of single frames (1024 x 768 pixel, RGB TIFF format). Second, static frames were processed with Photoshop CS6 to equalize for luminance and chromatic spectrum, and cropped to the size of 972 x 694 pixels in order to display only the mouth movements. Finally, by using Virtual Dub, we transformed the sequences of frames into silent AVI video-clips lasting 14 s. The experimental stimuli consisted in the video-clips rendered either as recorded (Forward mode, F) or after reversing the frames order (Backward mode, B). The total number of available stimuli was 2 [rendering mode] x 2 [language] x 10 [actor] x 4 [text] = 160. Four examples of video stimuli, one for each category of interest (I_F, I_B, A_F, A_B), are provided as supplementary material.

Additionally, we checked whether Arabic and Italian video-clips differed in motion energy. For each two consecutive frames of each video-clip, we calculated the mean of the squared differences in the red, green, and blue channels in every pixel [69–72]. The motion energy was estimated as the average of these values across all pixels and frames (350 frames) of each video. We found that the motion energy was not significantly different (unpaired t-test; p = 0.7; t-value (78) = 0.38) between Italian (mean ± SD: 34.0 ± 11.8 pixel^2/frame^2) and Arabic videos (34.9 ± 7.8 pixel^2/frame^2).

General outline of the study

The study involved three successive experiments: a preliminary, purely behavioural session with the first group of 20 participants in which we estimated the ability to discriminate presentation modes (Backward/Forward); a main session with the second group of 20 participants in which this ability was estimated while measuring brain activity with fMRI; a follow-up session with the same participants of the fMRI session in which we estimated the ability to discriminate language familiarity (Italian/Arabic).

Main task: Identification of rendering mode

In both fMRI and preliminary experiments, participants were informed that the video-clips being shown could be either a faithful or a time-reversed rendering of actual speech movements. They were not informed that in half of the videos the actor's language was Italian (I) and in remaining half was Arabic (A). The task (2-Alternative Forced Choice: 2-AFC) was to indicate whether the video was displayed as recorded (Forward) or reversed in time (Backward). Participants had to wait until the end of the stimulus before responding. Responses were entered by pressing with the right index finger one of two buttons marked "F" (Forward) and "B" (Backward), respectively. Between trials, the display was uniformly grey and participants fixated a central point (0.5° visual angle). No constraints were imposed on oculomotor behaviour during the presentation of the stimuli. Before each experimental session, participants were administered eight warm-up trials, which included at least one example for each combination of actor gender, native language, and rendering mode. The results of these trials were not analysed.

Follow-up task: Identification of language

The participants in the fMRI experiment were retested 10 (± 2) days later in a follow-up experiment outside the scanner. They viewed the same 160 silent video-clips described above. Participants were informed that the actor's language could be either Italian or another, unspecified language, but no further information was provided. Participants were asked to wait until the end of the stimulus before responding (2-AFC). Reponses were entered by pressing with the right index finger a button marked either I (Italian) or NI (not Italian). The aim of this experiment was to gauge the accuracy with which viewers could discriminate a familiar language (Italian) from an unfamiliar one (Arabic) using only visual cues, and to test whether the time-arrow (forward vs. backward rendering mode) affects the judgment of language familiarity. We hypothesized that the visual cues used to discriminate languages (e.g., temporal variability of vowels) are different from those used for discrimination of rendering mode (e.g., the acceleration profiles of opening/closing movements of the mouth). If so, the sensitivity index (see below) should be uncorrelated between the two tasks.

General procedure

In each experiment, the total number of stimuli (160) was divided in 5 runs (32 stimuli/run) with the constraint that successive stimuli never involved the same actor. Stimuli were pseudorandomized and presented using the Presentation software (Neurobehavioural system®). Within runs, interstimulus intervals (ISI) followed a uniform distribution (range: 2 s–4 s; mean: 3 s). The five runs were administered in a single session and were separated by brief pauses. Additionally, in the fMRI experiment, to estimate more accurately the shape of BOLD impulse response [73–75], we pseudo-randomly inter-mixed null events (N = 35, duration 8 s). Thus, the duration of each run was 10' 50" during fMRI and 9' 54" in the follow-up and preliminary experiments.

fMRI experiment: Set-up

Participants lay supine in the MR scanner with the head immobilized with foam cushioning and wore earplugs and headphones to suppress ambient noise. A digital projector (NEC LT158, refresh rate: 60-Hz) projected the stimuli through an inverted telephoto lens onto a semi-opaque Plexiglas screen mounted vertically inside the scanner bore, behind the participant's head. The back-projected image was then viewed via a mirror mounted on the head coil positioned at about 4.5 cm from the eyes. The eye-to-screen equivalent distance was 66 cm, and the angular size of the projected image was 9° (width) × 6.4° (height). Responses were acquired with an MR-compatible response box (fORP, Current Designs).

Follow-up experiment and preliminary experiment: Set-up

The follow-up and preliminary experiments were performed in a quiet, dimly illuminated room. Participants sat in front a 19" LCD monitor and viewed the silent video-clips (9° x 6.4° visual angle) in a pseudo-random order at a distance of about 80 cm. Responses were entered via a high-speed button box (Empirisoft®).

Behavioural data analysis

In the fMRI experiment and in the preliminary experiments, where the task was to identify the rendering mode (see above), responses were collated by language, rendering mode and actor.

The number of responses “Forward” to Forward and Backward stimuli are indicated as $N_{F|F}$ and $N_{F|B}$, respectively, while the number of responses “Backward” to Forward and Backward stimuli for each language are indicated as $N_{B|F}$ and $N_{B|B}$, respectively. For each participant, the sample size was $N_T = 80$ for each language, thus a total of 1600 trials for each language was collected (N_T x 20 participants). For each participant, we computed a sensitivity index d’ = Z{Hit} − Z{False Alarm} and a response bias $c = -0.5^{*}$(Z{Hits} + Z{False Alarm}), where Z{Hit} and Z{False Alarm} are the z-scored transformed values of P{Hit} = P{F|F} = $N_{F|F}/N_T$, and P{False Alarm} = P{F|B} = $N_{F|B}/N_T$, respectively [76]. Moreover, we calculated the probability of correct responses as P{C} = ($N_{F|F} + N_{B|B}$) / N_T.

Similarly, in the follow-up experiment where the task was to identify the actor’s language (see above), responses were collated by language, rendering mode and actor. The number of responses “Italian” to Italian and Arabic stimuli are denoted as $N_{I|I}$ and $N_{I|A}$, respectively, and $N_{A|I}$ and $N_{A|A}$ are the number of responses “Not Italian” to Italian and Arabic stimuli, respectively. For each participant, the sample size was $N_T = 80$ for each rendering mode, thus and a total of 1600 trials for each rendering mode was collected (N_T x 20 participants). We estimated sensitivity and response bias through d’ and *c* indexes, respectively, based on the convention that, in this case, Z{Hit} and Z{False Alarm} are the z-scored transformed values of P{Hit} = P{I|I} = $N_{I|I}$ / N_T and P{False Alarm} = P{I|A} = $N_{I|A}$ / N_T, respectively. Moreover, we calculated the probability of correct responses: P{C} = ($N_{I|I} + N_{A|A}$) / N_T.

We considered d’ and response bias in addition to the probability of correct responses, since the latter might be inflated by response bias and lead to misleading interpretations [76].

We expected that responses to the stimuli depended on whether participants had to discriminate between rendering mode (in the first two experiments) or languages (in the follow-up experiment). Thus, within-subject responses to the rendering-mode and language discrimination task should show different patterns, and the sensitivity index (d’) should be uncorrelated between tasks. To verify these points, we calculated the correlation coefficient of participants’ sensitivity index between the main task and the follow-up experiment task.

fMRI data acquisitions

MR images were acquired with a Siemens Magnetom Allegra 3T head-only scanning system (Siemens Medical Systems, Erlangen, Germany), equipped with a quadrature volume RF head coil. Whole brain BOLD echoplanar imaging (EPI) functional data were acquired with a 3T-optimized gradient echo pulse-sequence (TR = 2.47 s, TE = 30 ms; flip angle = 70˚; FOV = 192mm, fat suppression). 38 slices of BOLD images (volumes) were acquired in ascending order (64 x 64 voxels, 3 x 3 x 2.5 mm^3, distance factor: 50%; inter-slice gap = 1.25 mm; slice thickness = 2.5 mm), covering the whole brain. For each participant, a total of 1315 volumes of functional data were acquired in five consecutive runs. At the end of each run, the acquisition was paused briefly. Structural MRI data were acquired using a standard T1-weighted scanning sequence of 1 mm^3 resolution (MPRAGE; TE = 2.74 ms, TR = 2500 ms, inversion time = 900 ms; flip angle = 8˚; FOV = 256 × 208 × 176 mm^3).

fMRI data preprocessing

Data and statistical analyses were performed using the SPM12 software (Wellcome Trust Centre for Neuroimaging, London, UK) implemented in MATLAB R2013 (The MathWorks Inc., Natick, MA) using standard procedures [77,78]. After discarding the first four volumes of each run, images were corrected for head movements, realigned to the mean image, coregistered to the structural image, and normalized to Montreal Neurological Institute (MNI) space using unified segmentation [79], including resampling to 2 × 2 × 2 mm voxels, and spatially smoothed with a 8 mm full-width at half maximum (FWHM) isotropic Gaussian kernel. Voxel time series were processed to remove autocorrelation using a first-order autoregressive model and high-pass filtered (128-s cut-off).

fMRI analysis

Patterns of brain activations were computed using the general linear model and a Finite Impulse Response (FIR) set of base functions. Here, the FIR approach is ideal to fit brain activity, because it can identify changes of activity over time without making any assumptions about the profile of these changes [80]. Accordingly, for each participant, the FIR estimated the level of activation in 12 successive time-bins. Each time-bin consisted of 1 TR (2.47 s), thus fitting 30 s of the fMRI data for each stimulus. We modelled 5 different event-types: Italian Forward rendering (I_F), Italian Backward rendering (I_B), Arabic Forward rendering (A_F), Arabic Backward rendering (A_B), time-locked to stimulus onset, thus obtaining 12 images (one for each time bin) for each correct trial of the 4 conditions, plus an additional event corresponding to errors trials irrespective of condition. Motion correction parameters were also included as effects of no interest. We analysed the activity related only to stimuli correctly identified (correct trials), since error trials (stimuli not correctly identified) may introduce confounding activation (i.e. contamination of the activation related to poorer performance by increased errors [81,82]. However, to evaluate the effect of error trials on the fMRI activity, we did a supplementary analysis (not reported here) with all trials (correct and error trials). We found that the brain sites activated in the main fMRI analysis (see Results and Table 1) were also activated in the supplementary fMRI analysis, although at an uncorrected level (p-uncorr < 0.05), thus indicating that error trials decrease the signal-to-noise ratio.

At single-subject level, we estimated four effects of interest. First, we calculated the contrast representing the overall mean activity of all stimuli by averaging the estimated parameter of all conditions ($[I_F + I_B + A_F + A_B]/4$) in each bin. Subsequently, we estimated the contrasts of the three effects: (1) main effect of actor's language ($[I_F + I_B]$ vs. $[A_F + A_B]$); (2) main effect of rendering mode ($[I_F + A_F]$ vs. $[I_B + A_B]$); and (3) modulatory effect of actor's language on rendering mode (interaction: $[I_F – I_B]$ vs. $[A_F – A_B]$). The resulting parameters for each contrast (corresponding to 12 images, one for each time bin) in each participant were then entered into second-level group analyses [83].

Four separate one-way ANOVAs with 12 levels (each corresponding to one time-bin) were performed at the second (group) level. We used F-contrasts to highlight brain areas showing differential activity over the 12 time-bins, separately for each of the four ANOVAs. In particular F-contrasts subtracted the activity of the first bin (i.e. one TR at stimulus onset) from each of the other bins, thus capturing the changes of activity over-time. All analyses included appropriate corrections for non-sphericity. Statistical thresholds were set at p-FWE < 0.05, family-wise error corrected for multiple comparisons at cluster level (hereafter, p-corr < 0.05), using a voxel-wise threshold set at $p< 0.001$ [84,85]. Furthermore, post-hoc t-tests on each time bin were false-discovery-rate (FDR) corrected for n multiple comparisons at $p < 0.05$ across the number of bins ($n = 12$).

Regions of interest

In addition to the previous whole-brain analysis, we also performed an analysis based on regions of interest (ROIs). In particular, we defined regions as spheres of 8 mm radius centred on premotor areas that respond to visual speech (Premotor Ventral inferior PMvi/Broca's xyz = −48 12 9, xyz = −51 9 9; Premotor ventral superior / premotor dorsal PMvs/PMd xyz = −39 3 54, xyz = −48 3 42; BA6 and BA 44 xyz = 48 18 18) [86]; visual motion area MT+/V5 (xyz =

Table 1. Peaks of cluster activations.

Actor's language (Italian Vs Arabic)						
Anatomical Area	*x*	*y*	*z*	*k*	*F-value*	*FWE corr*
FGa	-34	-66	-14	265	6.44	Whole brain
IOG	-34	-80	-14		3.50	
OTS	38	-72	-4	353	5.33	Whole brain
FGa	38	-74	-12		4.95	
IOG	40	-80	10		3.70	
Precuneus	0	-64	44	169	4.06	Whole brain
	-8	-60	44		3.67	
	-4	-52	42		3.44	
PMvs/PMd	-44	6	50	84	4.54	ROIs
Rendering mode (Forward vs Backward)						
	x	*y*	*z*	*k*	*F-value*	*FWE corr*
IPS	-24	-66	60	376	6.16	Whole brain
	-28	-64	38		3.99	
	-18	-60	46		3.42	
FGb	-30	-60	-4	137	4.98	Whole brain
IPS	30	-68	42	224	4.35	
	32	-70	34		4.08	
	30	-58	44		3.26	
Pmvi	-56	10	6	111	4.09	ROIs
Interaction: Actor's language x Rendering mode						
	x	*y*	*z*	*k*	*F-value*	*FWE corr*
IFG	-40	38	10	193	6.12	Whole brain
	-48	34	4		3.73	
LG	-18	-62	0	464	5.46	Whole brain
	-26	-66	8		3.91	
	-22	-52	-6		3.29	

FG = Fusiform Gyrus, OTS = Occipito—Temporal Sulcus, IPS = Intra-Parietal Sulcus, LG = Lingual gyrus, IFG = Inferior Frontal gyrus; *k = cluster size (in voxels). Family wise error correction (FWE) at p < 0.05 can be whole brain or within ROIs.*

https://doi.org/10.1371/journal.pone.0234695.t001

-42–66 2, xyz = 42–62 6) [87], and sites in the posterior inferior temporal sulcus involved in biological motion processing (pITS xyz = -50–82 0, xyz = 48–78–4) [88]. Finally, we considered ROIs also in the fusiform face area (FFA xyz = -34–62–15, xyz = 34–62–15) [89] and in the temporal visual speech area (TVSA xyz = -57–34 14) [90]. We applied family-wise-error small-volume-correction (FWE-SVC) to each ROI [91,92]. We retained results as significant at p < 0.05 FWE-SVC, further Bonferroni-corrected for the number of regions (n = 12).

Results

Behavioural results

Main task: Identification of the rendering mode (Forward or Backward). During the preliminary and fMRI experiments, observers had to indicate whether the video-clip was played in Forward or Backward mode.

Observers detected the rendering mode (Forward or Backward) of video-clips with an overall probability of correct responses P{C} = 0.590 and P{C} = 0.556 for the preliminary and fMRI experiments respectively (pooled across participants and stimuli), significantly higher than chance level (two-tailed binomial test, $p < 0.001$, Fig 1a). Sensitivity (d') for Italian (d': 0.56 ± 0.14 and d': 0.37 ± 0.09, mean ± s.e.m., for preliminary and fMRI experiments respectively) and Arabic (d': 0.47 ± 0.15 and d': 0.26 ± 0.08, respectively) was not significantly different (paired t-test; p = 0.52; t(19) = 0.66 and p = 0.16; t(19) = 1.44 for preliminary and fMRI experiments respectively, Fig 1b). For both languages, d' was significantly greater than 0 (one sample t-test; all $p < 0.002$; $t(19) > 3.25$ and all $p < 0.004$; $t(19) > 3.32$ for preliminary and fMRI experiments respectively). However, there was a significant response bias (*c* = -0.25 ± 0.07, p = 0.003, t(19) = 3.34 and *c* = -0.36 ± 0.05, p = 0.001, t(19) = 4.05 one-sample t-test, for preliminary and fMRI experiments respectively) in favour of the response "Forward" for the Italian video-clips (Fig 1c), underlying the higher proportion of correct response in this task. By contrast, there was no response bias for the Arabic video-clips (*c* = 0.024 ± 0.06, p = 0.66, t(19) = 0.43 and *c*: -0.01 ± 0.08, p = 0.9, t(19) = 0.12, one-sample t-test, for preliminary and fMRI experiments respectively).

The comparison of sensitivity (d') and response bias (c) indexes between the fMRI and the preliminary experiments, computed for both Italian and Arabic video clips, did not show significant differences (t-test, all $t(19) < 1.28$, $p > 0.21$).

Follow-up experiment: Language identification. All the participants in the fMRI experiment were retested in a follow-up experiment to ascertain their ability to recognize the language by means of visual-only cues. In this experiment, volunteers had to indicate if the actor's language in the silent video clips was Italian or not. Fig 1a (white bars) reports for each condition the probability of correct responses (P{C} = 0.679) pooled across stimuli and participants. In all conditions, P{C} was significantly higher than chance level (two-tailed binomial test, $p < 0.001$). The average d' was not significantly different between video-clips played in forward (d': 1.05 ± 0.12) and backward mode (d': 0.96 ± 0.15) (paired t-test; p = 0.5; t(19) = 0.65) and in both cases d' $>$ 0 (one sample t-test; all $p < 0.001$; $t(19) > 7.02$) (Fig 1d). There was a significant response bias in favour of the response "Italian" in the case of Forward video-clip (c: -0.21 ± 0.05, $p < 0.001$, t(19) = 4.43, one-sample t-test). Conversely, in the case of Backward video-clips, the response bias was in favour of the response "not Italian" (c: 0.12 ± 0.034, $p < 0.01$, t(19) = 2.95, one-sample t-test).

Comparison between tasks. We expected that stimuli were classified differently depending on the task, and that the two tasks had different response patterns across subjects. To verify this hypothesis, we compared the participants' sensitivity index (d') in the fMRI main task (rendering mode discrimination) and in the follow-up experiment (language discrimination). The analysis of d' showed a greater sensitivity to stimuli in the follow-up experiment task compared to stimuli sensitivity in the main task (paired t-test, t(19) = 6.15, $p < 0.001$). An alternative possibility is that the higher d' in the second experiment could be due to a learning process occurring after the first experiment. In this case, we should expect a correlation across participants between tasks. However, the sensitivity indexes of the two tasks were not correlated (Pearson's r = 0.27, p = 0.23), suggesting that the two tasks rely on different processing.

fMRI results

Brain areas engaged by visual speech. We mapped the cortical regions activated by all visual speech stimuli, irrespective of the parameters manipulated experimentally, (i.e., all stimuli vs. rest) by a differential F-test across the 12 time-bins (see Methods). This test highlights regions having different amplitude and/or time-course of the BOLD response between conditions examined in the contrast image (in this case, all stimuli vs. rest condition). As shown in Fig 2, significant effects were observed in occipital and temporal cortices, i.e. regions that are typically involved in audio-visual processing, as well as in parieto-frontal cortices, which are

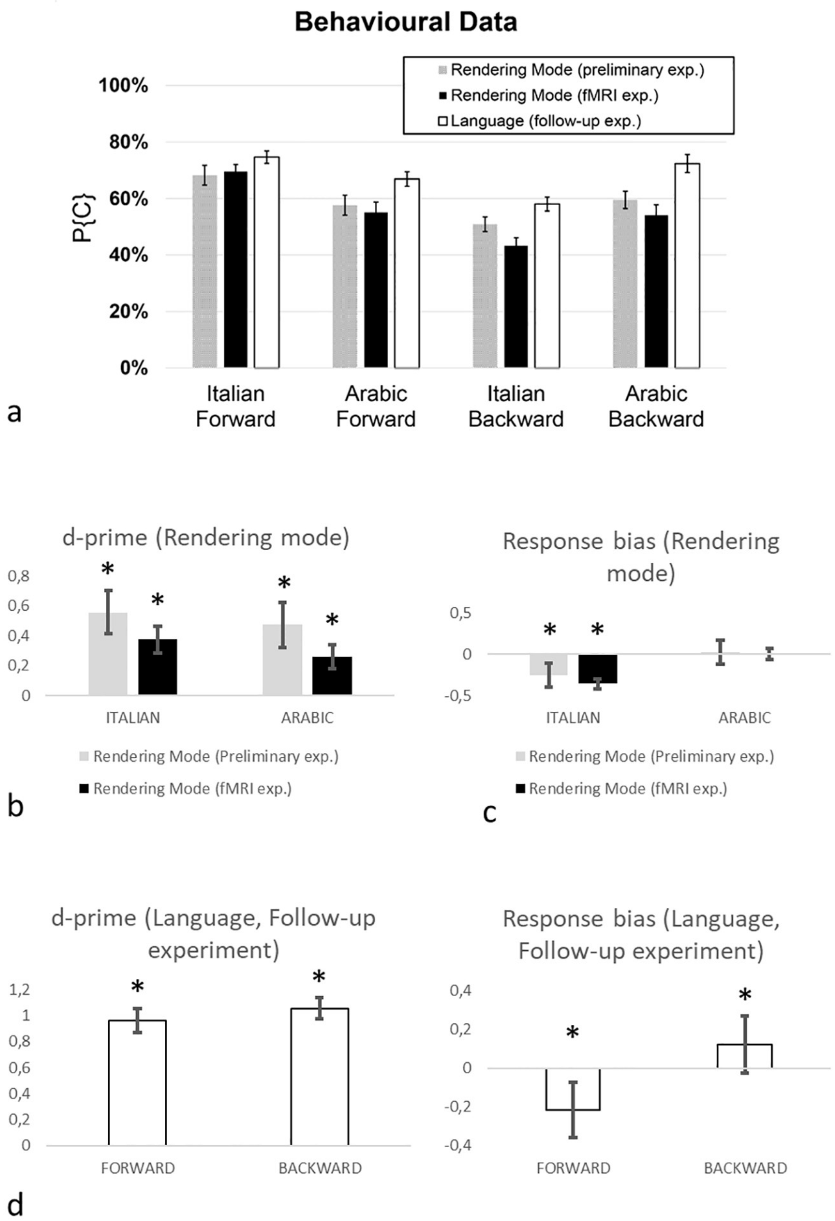

Fig 1. (a) Proportion of correct responses (mean ± s.e.m.) separate for conditions and experiment. (b) D-prime sensitivity to rendering modality in the preliminary and fMRI experiment. (c) Response bias to language during the sensory modality in the preliminary and fMRI experiment. (d) D-prime sensitivity to language and response bias to rendering modality during the follow-up experiment. Asterisks indicate significant values.

https://doi.org/10.1371/journal.pone.0234695.g001

generally engaged by the vision of speech movements (Bernstein et al 2014), in the insula, cingulate cortex, motor and premotor areas. Activity in left motor/premotor areas was presumably related, at least in part, to the right-hand motor responses.

Main effect of actor's language. *Whole brain analysis*. The differential F-test comparing Italian (familiar) vs. Arabic (unfamiliar) stimuli (irrespective of rendering mode) across the 12 time-bins revealed significant activations (p-corr < 0.05, whole brain) bilaterally in the posterior fusiform gyrus (FG_a, at coordinates almost coinciding with those of FFA), extending to

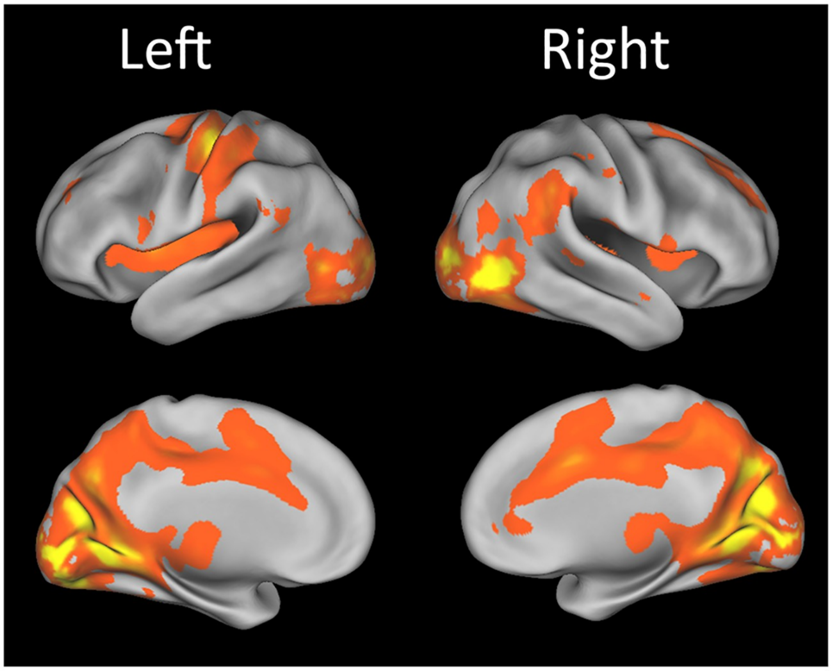

Fig 2. Statistical parametric mapping of GLM analysis. Effects of all stimuli vs. rest (F-test, p-corr < 0.05). Activations maps were overlaid on the standardized (inflated) brain of the PALS-B12 atlas implemented in Caret5 (Van Essen et al. 2005).

https://doi.org/10.1371/journal.pone.0234695.g002

the inferior occipital gyrus (IOG) and right occipito-temporal sulcus (OTS, at coordinates near to those of pITS, a biological motion sensitive area), and in the precuneus (Fig 3, green area, Table 1). The time profiles of the estimated BOLD activity in these regions are plotted in Fig 4a, 4b and 4c. It is important to note that direct inspection of these activity patterns is necessary before any conclusion can be drawn, since significant differences obtained through our statistical analysis (i.e., F-test) could be due to a modulation in amplitude and/or to a time-shift. Time bins presenting a different activity level (post-hoc t-test, p-FDR corrected for multiple comparison < 0.05 across bins) between Italian and Arabic are filled in green. FG_a showed enhanced earlier activity (see bins 1^{th}–3^{th}) for Italian stimuli independently of rendering mode, and later activity for Arabic stimuli (6^{th}–8^{th} bins, compare grey and black lines in Fig 4). IOG and OTS, belonging to the same cluster of FG_a, had a similar temporal profile (e.g., OTS in Fig 4). The precuneus showed the opposite trend, with a decreased activity in the earlier bins for Italian stimuli (see black lines in Fig 4c, 3^{th}–4^{th} bins) and in the later bins for Arabic stimuli (7^{th}–8^{th} bins).

ROI analysis. A significant effect of the familiar language independently of rendering mode (p-corr < 0.05, FWE-SVC Bonferroni) was also found in the left PMvs/PMd (Table 1, Fig 3, green area). These regions showed increased activity for the Italian stimuli independently of rendering mode only in late bins (6^{th}–9^{th} bins) (Fig 5a).

Also left FFA and right pITS showed a main effect of language, already reported in the whole brain analysis. By contrast PMvi, MT+/V5 and TVSA did not show a significant main effect of language.

Main effect of rendering mode. *Whole brain analysis*. Regions with differential responses to normal kinematics (i.e., video-clips played forward) and to implausible kinematics (i.e., video-clips played backward) were identified by the contrast Forward vs. Backward mode

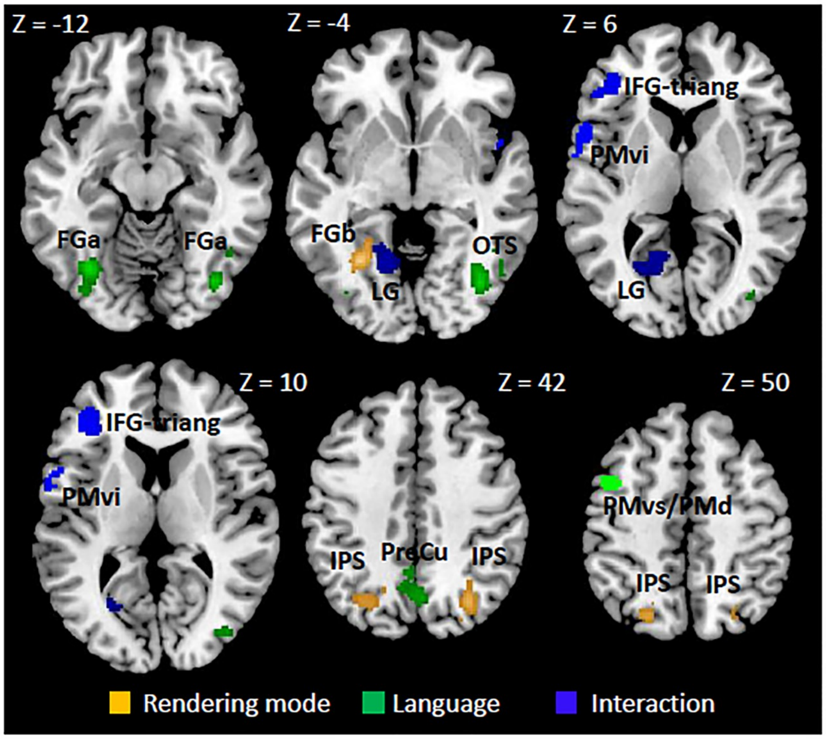

Fig 3. Statistical parametric mapping of GLM analysis. Main effects of actor' language (green), rendering mode (orange) and the interaction actor's language x rendering mode (blue), (F-test, p-corr < 0.05). Maps are projected onto the coronal slices of the 152-MNI template. Z-coordinate of each slice is reported on the top (in mm). $FG_{a,b}$: fusiform gyrus; IFG-triang: inferior frontal gyrus pars triangularis; OTS: Occipito-Temporal Sulcus; PreCu: precuneus; IPS: intraparietal sulcus; LG: lingual gyrus; PMvi: premotor ventral inferior; PMvs/PMd: premotor ventral superior/ Premotor dorsal.

https://doi.org/10.1371/journal.pone.0234695.g003

(main effect of rendering mode), irrespective of language (Italian or Arabic). The regions significantly sensitive to this contrast (orange regions in Fig 3) were found in the intraparietal sulcus (IPS) bilaterally, and in the left posterior-middle fusiform gyrus (FG_b) ($p\text{-corr} < 0.05$, whole brain). Fig 4d, 4e and 4f show the time-course of activity in these regions. Bins in which the activity differed significantly between normal and reversed video-clips are filled in orange (post-hoc t-test, $p<0.05$ FDR corrected for multiple comparisons across bins). In particular, the right IPS (1^{th}–2^{th} bins) and FG_b (1^{th}–2^{th} bins) responded more to Forward than Backward rendering mode (compare continuous and dotted lines in Fig 4e and 4f, respectively). Left IPS showed a similar trend in the early bins (2^{th}–3^{th} bins) at a lower statistical threshold (p-uncorr < 0.05).

Finally, the image resulting from the intersections (logical AND) between the cluster image of the left FG_a (reported above) and the cluster image of left FG_b showed that these two clusters were sharply separated (no voxel in common).

ROI analysis. A significant effect of the rendering mode independently of language (p-corr < 0.05, FWE-SVC Bonferroni) was also found in the PMvi, in the pars opercularis of IFG (p-corr < 0.05, FWE-SVC Bonferroni) (see Table 1, Figs 3 and 5b, blue). PMvi responded more during late bins to the rendering modality (8^{th}–9^{th} bins), but selectively to Italian language, so that also the interaction calculated on the peak was significant (see also below).

By contrast none of the posterior ROIs (MT+/V5, FFA, pITS, TVSA) nor PMvs/PMd showed a main effect of rendering.

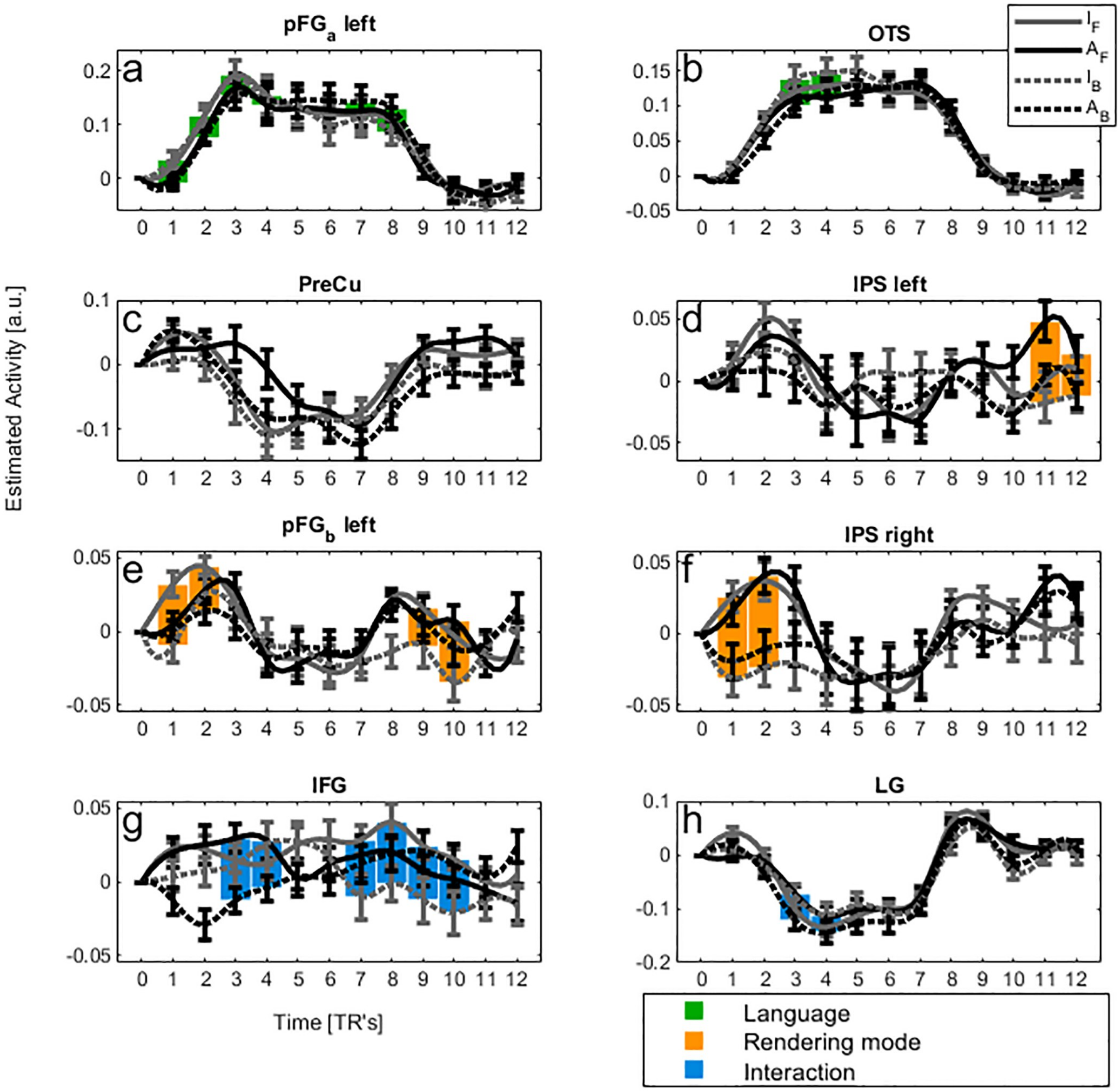

Fig 4. Peri-Stimulus Time Histogram (PSTH) for regions that showed significant effects in the whole-brain analysis. Mean time course (± s.e.m) of estimated BOLD signal at the peak voxel (see Table 1). Abscissa: time in TR's unit (TR = 2.47s), T = 0: trial onset. Continuous and dotted lines indicate forward and backward rendering mode, respectively. Black and grey lines indicate Arabic and Italian video-clips, respectively. Green, orange and blue filled rectangles indicate time bins showing a significant difference in BOLD activity (post-hoc t-test, p-value FDR-corrected for multiple comparison < 0.05) due to Actor's language, rendering mode or interaction, respectively. I_F: Italian forward, A_F: Arabic forward, I_B: Italian Backward, A_B: Arabic Backward.

https://doi.org/10.1371/journal.pone.0234695.g004

Influence of actor's language on the rendering mode discrimination process. *Whole brain analysis*. Through the contrast ($[I_F - I_B]$ vs. $[A_F - A_B]$), we searched for brain sites where the response to rendering mode was affected by language. This analysis revealed significant activations (p-corr < 0.05) in pars triangularis (BA 45) of the left inferior frontal gyrus (IFG-

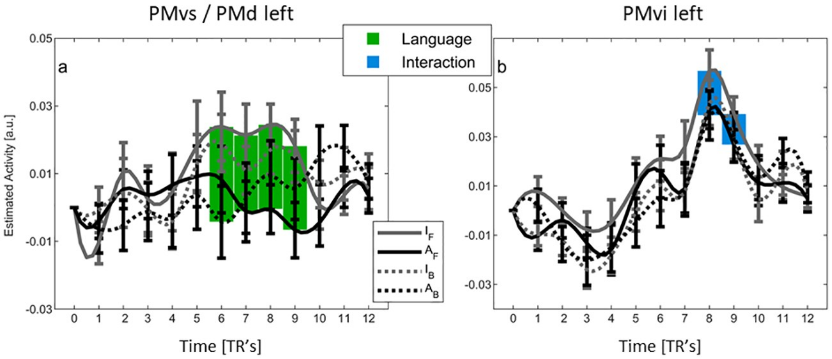

Fig 5. Peri-Stimulus Time Histogram (PSTH) for regions that showed significant effects in the ROI analysis. Mean time course (± s.e.m) of estimated BOLD signal at the peak voxel (see Table 1). Abscissa: time in TR's unit (TR = 2.47s), T = 0: trial onset. Continuous and dotted lines indicate forward and backward rendering mode, respectively. Black and grey lines indicate Arabic and Italian video-clips, respectively. Green and blue filled rectangles indicate time bins showing a significant difference in BOLD activity (post-hoc t-test, p-value FDR-corrected for multiple comparison < 0.05) due to Actor's language or interaction, respectively. I_F: Italian forward, A_F: Arabic forward, I_B: Italian Backward, A_B: Arabic Backward.

https://doi.org/10.1371/journal.pone.0234695.g005

triang) and in lingual gyrus (LG, see blue regions in Fig 3). The temporal profile of BOLD responses (Fig 4g) showed that IFG-triang differentiated the Arabic video-clips played in Backward and Forward mode in the earlier bins, in particular Arabic backward stimuli strongly deactivated IFG-triang (3th–4th bins) (post-hoc t-test, p-FDR corrected for multiple comparison < 0.05 across bins). In a similar way, IFG-triang differentiated Forward from Backward Italian video-clips, but at later times compared to Arabic stimuli discrimination, namely between the 8th and 10th bins (Fig 4g). Moreover, Italian stimuli played backward also showed a marked negative pattern in these bins. Overall, IFG-triang responded similarly to Italian and Arabic stimuli, although the temporal patterns were shifted. Indeed, neither the difference between the maximum peaks for Italian Forward stimuli (bin 8) and Arabic Forward stimuli (bin 3) (t(19) = 0.93; p = 0.36), nor the difference between Forward and Backward condition of Italian stimuli at bin 8 and that of Arabic stimuli at bin 3 (t(19) 0.12; p > 0.9) were significantly different (compare the differences between continuous and dotted grey lines in bin 8 and between continuous and dotted black lines in bin 3, respectively). In sum, the BOLD patterns showed that IFG-triang does not have a clear preferential response to the speech gestures most frequently performed by participants (i.e., Italian Forward stimuli).

LG showed a general de-activation in all four conditions versus rest. In particular, Italian Backward and Arabic Forward stimuli involved a very similar time-course, as did Arabic video-clips played backwards and Italian video-clips played forwards. These two latter conditions were also the two most deactivating (i.e., negative B[93]OLD patterns) conditions in this site (see 3th–4th bins filled in blue in Fig 4h) (post-hoc t-test, p-FDR corrected for multiple comparison < 0.05 across bins).

ROI analysis. This analysis revealed a significant interaction between language and rendering mode in PMvi (IFG pars opercularis), a region that was selective for the Italian language in late bins (8th–9th bins) (Table 1, Figs 3 and 5b, blue). In particular, there was a single peak of higher response to the familiar language with respect to the other three conditions, indicating a clear preferential response to the speech gestures most frequently performed by participants.

By contrast, none of the posterior ROIs (MT+/V5, FFA, pITS, TVSA) nor PMvs/PMd showed an interaction between language and rendering mode.

Finally, we calculated a minimum effect size of 0.15 corresponding to the lowest significant F value reported in Table 1 (F(11,209) = 3.26) [93–95]. A partial eta-squared of 0.15 indicates a large effect [93].

Discussion

We reported differential brain responses to visual speech kinematics, language familiarity and their interaction. Neuroimaging data showed that language familiarity and temporal rendering of silent speech video-clips modulated two distinct areas in the ventral occipito-temporal cortex. Furthermore, language familiarity modulated the left dorsal premotor cortex, while natural familiar language activated the left ventral premotor cortex in the frontal operculum. These results may indicate that phono-articulatory regions resonate in response to the visemes (visual equivalents of phonemes) of a familiar language. Since in our experiments participants generally did not decode the semantic and syntactic content of visual speech, we propose that these results are confined to the visual equivalent of the phonemic axis. Indeed, our results are in agreement with the definition of a phonological pathway more dorsal with respect to the lexical and semantic pathways, which includes IPS, the dorsal premotor region and the pars opercularis of IFG [96,97].

Sensitivity to the time-arrow of visual speech

Participants were able to discriminate above chance level visual speech gestures rendered forwards from those rendered backwards. Behavioural results were consistent across experiments. Noteworthy, the sensitivity index d' estimated in the preliminary and in the main fMRI experiments were not statistically different. These results indicate a sensitivity of the central nervous system for temporal features (i.e., time arrow) of the visible speech, in agreement with the results obtained with a familiar language in a previous study with a similar task [1].

Although lip-reading accuracy of hearing people is generally low and idiosyncratic [98], one cannot rule out a priori that observers were occasionally able to lip-read excerpts of Italian texts in the forward mode, and to use these instances as a cue for discriminating the rendering mode. However, the fact that sensitivity was not significantly different for Italian and Arabic stimuli suggests that lexical competence and speech intelligibility did not play a significant role in the task. The evidence suggests instead that better than chance performance was achieved mainly by a kinematic analysis of movements. If so, the performance reflected the ability to discriminate the motor sequences that are visually perceived as plausible from those that are perceived as implausible from the motoric point of view. This assumption can be sharpened by taking into account the response bias, which describes the position, along the decision axis, of the internal threshold for discriminating the stimuli [76]. In our experiment, there was a response bias in favour of the response "Forward" for the Italian but not for Arabic stimuli, indicating a corresponding shift of the threshold to higher values. This invites the inference that in order to classify a movie reversed in time as 'Backward', it is necessary to detect more motoric incongruences in Italian than in Arabic stimuli. This inference is in keeping with the suggestion [99,100] that a high threshold for detecting speech kinematic anomalies favours the stability of speech perception in environments where such anomalies in a familiar language occur due to inter-individual differences or phonetic peculiarities typical of particular social environments (regional inflexions, slang, etc.). Indeed, the participants to both the preliminary and fMRI experiments had no reason to suspect that in half of the video-clips the language being spoken was not Italian.

In a follow-up experiment, participants were asked to identify the language (Italian or not Italian) spoken by the actors in the silent movies. If participants benefitted from speech intelligibility, then the sensitivity to discriminate languages should be higher for Forward than that for Backward rendered stimuli [101]. The results of the follow-up experiment do not conform with this scenario, because the sensitivity index d' was comparable for Forward and Backward rendered video-clips (see Fig 1), making unlikely that speech intelligibility or lexical processes occurred in our tasks.

The lack of significant correlation across participants between the sensitivity in the fMRI and follow-up experiments suggests that different processes, likely taking into account non-overlapping sets of cues, underlie language and rendering mode discrimination. Conversely, the finding that response bias during rendering mode discrimination depended on the language (and vice-versa) might indicate that, during a late stage of analysis, these signals are merged. This merging might take place in the premotor cortex, where the PMvs/PMd selected the familiar language independently of rendering, while the PMvi responded to the Italian language selectively in the natural kinematics condition.

Ventral occipito-temporal cortex (vOTC)

The main effects of language and rendering mode activated distinct regions in ventral occipito-temporal cortex. In particular, the comparison of Forward with Backward conditions showed a differential pattern of BOLD responses in the left posterior-middle fusiform gyrus (FG_b), while the posterior fusiform bilaterally (FG_a), the inferior occipital gyri and the occipito-temporal sulcus (OTS) were differentially involved with Italian versus Arabic video-clips. The fusiform sites are located posteriorly to the visual areas engaged by semantic and/or lexical processes in the vOTC, which are typically reported at y-coordinates < -50 mm, in the anterior part of the fusiform gyrus [see 102]. Conversely, the fusiform face area (FFA), a region responding selectively to static faces, is located in the posterior region of the fusiform gyrus [35]. The stereotaxic coordinates of FFA centre of mass ([34 –62 –15], [89]) roughly correspond to those of the peak of FG_a reported here, but are posterior to those we found for the main effect of rendering mode (FG_b). Previous studies reported that multiple sites in FG respond to faces [32,89,103]. It is likely that different foci in FG encompass distinct functional modules, as suggested by early PET studies [104,105] showing that gender and face identification activated distinct regions in posterior and middle fusiform gyrus, respectively.

It has been shown that the kinematics of biological movements [30,106–108], as well as the temporal unfolding of faces that express an emotional state [31] engage ventral OTC. In particular, observing facial speech gestures activates FG, although it is unclear whether these activations are specific for speech because some control stimuli, such as gurning faces, activated this region more than talking faces [41]. Indeed, the difference between speech and control stimuli may have been due to differences in low-level features, such as visual motion speed [38,41]. In our study, all four experimental conditions (showing exclusively the lower portion of faces) were comparable in terms of low-level features, such as mean luminance and motion speed. By contrast, time reversal of visual speech stimuli violates motor constraints, and hence produces movements with an implausible kinematics, never occurring during real speech. Previous studies showed that coherent sequences of facial expressions engage the posterior fusiform gyrus more than a scrambled sequence [109]. A possibility is that ventral OTC processed specific kinematic cues embedded in visible speech. Therefore, we speculate that the posterior-middle fusiform site (i.e., FG_b) was sensitive to the kinematic plausibility of speech gestures. Conversely, the more posterior site FG_a was involved mainly in processing kinematic features related to the familiarity of speech, such as the rhythm of speech that is invariant under time reversal but differs across languages (see Introduction).

The previous observations challenge one of the most prominent models about face processing, namely the model proposing that static and dynamic face features are processed separately in ventral OTC and STS, respectively [110]. In fact, our data suggest that ventral OTC has foci sensitive to spatio-temporal (i.e. changeable) characteristics of speech lip movements.

Italian-speech stimuli evoked a higher activity peak than Arabic-speech stimuli in the more posterior site of fusiform gyrus (FG_a), whereas the sustained post-peak activity was greater for Arabic than Italian stimuli. The Arabic-speech stimuli were unfamiliar, and thus they were presumably unexpected. It is thought that a specific class of neurons (the so-called error neurons) responds selectively to unexpected or unusual stimuli [111–113]. These neurons compare the sensory input with an internally generated (prediction) signal coding what is expected in a given context [114]. In case of a mismatch between the predicted and the incoming sensory signal, error units enhance their activity. We surmise that the greater activity for Arabic than Italian stimuli in the post-peak period might be related to the activity of error units. Therefore, depending on the language (familiar or unfamiliar), this class of neurons contributed differently to the overall neural activity in FG_a, so the pattern of neural activity changes according with language. However, this mechanism is not specific to ventral OTC, but is a widespread mechanism governing several brain processes (see Friston 2010). For instance, we recently surmised that this kind of neural processing occurs also in lateral OTC when observing unfamiliar walking movements [107,108], and it might even bias balance control [115].

Lingual gyrus

The interaction between language and rendering mode showed significant activations in the lingual gyrus. Previous reports have already suggested that visually presented speech gestures engage this site [36,116]. The novel finding reported here is that LG responds differently depending on the language. In the follow-up experiment (language discrimination), the d' was similar between Forward and Backward stimuli, while there was a significant bias toward Arabic-speech or Italian-speech response in the Backward or Forward condition, respectively. The temporal profile of LG activity distinguishes the experimental conditions. In particular, the conditions Arabic Backward and Italian Forward were the two most deactivating conditions. Therefore, the response bias in language discrimination task could be related to a decrease of LG activity. Interestingly, in the auditory domain, LG activity has been found to depend on the familiarity of the spoken language [117]. In the latter study, hearing a speech segment in a second language modulated LG activity differently depending on the participant's proficiency in that language. Because LG is involved, together with fronto-parietal regions, in speech control [118], these findings suggest a supra-modal effect of speech language in LG, probably due to feedback from high-order centres.

Premotor prefrontal activity reflects motor simulation of speech gestures

The interaction of rendering mode and language showed a significant effect in the left IFG, comprising the pars opercularis (BA 44), which we have labelled as PMvi, and the pars triangularis (BA45). Most authors agree that the Broca's area includes both BA 44 and BA 45 of the left hemisphere [43,119–121]. Broca's area was initially thought to be involved only in speech production, but current research shows that it has a more complex role possibly involving also speech comprehension [43,64,122,123].

In particular, the activation of IFG during speech perception has been interpreted by some authors as the occurrence of a motor simulation of the observed movements. This idea is in accordance with the motor theory of language holding that a simulation of speech gestures in the motor regions is instrumental for speech perception and understanding [53,124–126].

However, it is still an open issue whether the IFG activation during speech perception is related to a language-specific process, as the putative motor simulation of speech gestures, or represents a general-domain cognitive mechanism [127]. It has also been suggested that distinct IFG foci have different roles during speech perception. Specifically, articulatory rehearsal of speech gestures would occur in pars opercularis, while pars triangularis and orbitalis could be related to cognitive-control mechanisms, such as decisional or working-memory processes [68,128,129]. The rehearsal function of pars opercularis generalizes across different types of movement, as this region was found to respond also to observation of hand movements [130].

The time-course of the BOLD signal that we observed in the pars opercularis of IFG fit with the predictions of the motor simulation hypothesis (Figs 3 and 5b). According to this hypothesis, the familiar stimuli (i.e., Italian Forward) should elicit a higher level of activity than unfamiliar, implausible gestures difficult to reproduce (e.g., Italian Backward or Arabic Forward and Backward stimuli). Conversely, in the pars triangularis of IFG, Italian-speech stimuli and Arabic-speech stimuli, for which latter participants had no motoric expertise, evoked comparable responses although shifted in time (Figs 3 and 4g). Thus, the results do not suggest a specific sensitivity for Italian-speech stimuli in the pars triangularis of IFG, but rather sensitivity for the kinematics of natural mouth movements, a kind of biological motion. Our data, limited to the case of a familiar language (Italian), are also in agreement with those reported by Paulesu et al. [27] in which IFG activity was greater for forward than backward silent movies of a speech in a familiar language.

The issue of the role of intelligibility of silent visual speech should be further investigated, as one could argue that motor simulations occur only or mainly when linguistic competences are required, as with a lexical discrimination task [27,86,131]. However, we believe that the ability of participants to speech-read might be a confound when trying to disentangle motor from higher cognitive functions of Broca's area. In our case, it appears that motor simulation occurs in absence of comprehension of the content of the speech.

Summary and conclusions

Previous studies focused mainly on the role of temporal auditory regions [37,48,132] and frontal regions [86,131,133] in processing visual speech. More recently, it has been shown that a region in the left posterior temporal cortex, the so-called temporal visual speech area (TVSA), is activated in visual phonetic discrimination [38], possibly integrating information coming from high-level visual areas in OTC [3,41,134]. We did not find significant effects in TSVA, as verified through a specific ROI drawn in this region. Our data suggest that the ventral occipito-temporal cortex has a sensitivity to visual speech gestures, contrary to the view that the peculiar analysis of visual speech starts at higher cortical levels [27,135]. Our results support the hypothesis that kinematic cues embedded in visible speech can be extracted through the visual pathways [136], outside the classical areas related to auditory speech and audio-visual integration [36,37,116,137,138]. Finally, the selective responses of PMvs / PMd to the familiar language and of PMvi to the natural familiar language support the hypothesis that motor simulation drives premotor activity during visible speech perception.

Supporting information

S1 Video.
(AVI)

S2 Video.
(AVI)

S3 Video.
(AVI)

S4 Video.
(AVI)

Author Contributions

Conceptualization: Vincenzo Maffei, Paolo Viviani.

Data curation: Vincenzo Maffei, Elisabetta Mazzarella, Maria Assunta Giusti.

Formal analysis: Vincenzo Maffei.

Funding acquisition: Iole Indovina, Francesco Lacquaniti.

Investigation: Vincenzo Maffei.

Methodology: Vincenzo Maffei.

Software: Vincenzo Maffei.

Supervision: Vincenzo Maffei, Emiliano Macaluso.

Validation: Vincenzo Maffei.

Visualization: Vincenzo Maffei.

Writing – original draft: Vincenzo Maffei.

Writing – review & editing: Vincenzo Maffei, Iole Indovina, Emiliano Macaluso, Francesco Lacquaniti, Paolo Viviani.

References

1. Viviani P, Figliozzi F, Lacquaniti F. The perception of visible speech: Estimation of speech rate and detection of time reversals. Exp Brain Res. 2011; 215: 141–161. https://doi.org/10.1007/s00221-011-2883-9 PMID: 21986668
2. Rosenblum LD, Saldaña HM. An audiovisual test of kinematic primitives for visual speech perception. J Exp Psychol Hum Percept Perform. 1996; 22: 318–331. https://doi.org/10.1037//0096-1523.22.2.318 PMID: 8934846
3. Campbell R. The processing of audio-visual speech: empirical and neural bases. Philos Trans R Soc Lond B Biol Sci. 2008; 363: 1001–1010. https://doi.org/10.1098/rstb.2007.2155 PMID: 17827105
4. Reisberg D, McLean J, Goldfield A. Easy to hear but hard to understand: A speechreading advantage with intact auditory stimuli. In: Dodd B, Campbell R, editors. Hearing by eye: The psychology of lip-reading. Lawrence E. London; 1987. pp. 97–113.
5. Dias JW, Rosenblum LD. Visibility of speech articulation enhances auditory phonetic convergence. Attention, Perception, Psychophys. 2016; 78: 317–333. https://doi.org/10.3758/s13414-015-0982-6 PMID: 26358471
6. Erber NP. Auditory–visual perception of speech. J Speech Hear Disord. 1975; 40: 481–492. https://doi.org/10.1044/jshd.4004.481
7. Jeffers J, Barley M. Speechreading (lipreading). Springfield, IL: Thomas; 1971.
8. McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. Nature Publishing Group; 1976; 264: 746–748. https://doi.org/10.1038/264746a0 PMID: 1012311
9. Gonzalez-Franco M, Maselli A, Florencio D, Smolyanskiy N, Zhang Z. Concurrent talking in immersive virtual reality: on the dominance of visual speech cues. Sci Rep. Springer US; 2017; 1–11. https://doi.org/10.1038/s41598-017-04201-x PMID: 28630450
10. Ostry DJ, Flanagan JR. Human jaw movement in mastication and speech. Arch Oral Biol. 1989; 34: 685–693. https://doi.org/10.1016/0003-9969(89)90074-5 PMID: 2624559
11. Gracco VL. Timing factors in the coordination of speech movements. J Neurosci. 1988; 8: 4628–4639.

12. Ronquest RE, Levi SV, Pisoni DB. Language identification from visual only pseech signals. Atten Percept Psychophysiol. 2010; 72: 1601–1613.

13. Ramus F, Pallier C, Dupoux E, Dehaene G. Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. Annu Rev Lang Acquis. 2002; 2: 1–14. https://doi.org/10.1075/arla.2.05ram

14. Ramus F, Mehler J. Language identification with suprasegmental cues: A study based on speech resynthesis. J Acoust Soc Am. 1999; 105: 512–521. https://doi.org/10.1121/1.424522 PMID: 9921674

15. Ramus F, Nespor M, Mehler J. Correlates of linguistic rhythm in the speech signal. Cognition. 1999; 73: 265–292. https://doi.org/10.1016/s0010-0277(99)00058-x PMID: 10585517

16. Tajima K, Zawaydeh A, Kitahara M. a Comparative Study of Speech Rhythm in Arabic, English, and Japanese. Proceedings of the XIV ICPhS. San Francisco, USA; 1999. pp. 3–6.

17. Ghazali S, Hamdi R, Barkat M. Speech rhythm variation in Arabic dialects. Proc Interspeech 2004. 2004; 1613–1616.

18. Abercrombie D. Elements of general phonetics. Aldine; 1971.

19. Dumoulin SO, Bittar RG, Kabani NJ, Baker CL, Le Goualher G, Bruce Pike G, et al. A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. Cereb Cortex. 2000; 10: 454–63. Available: http://www.ncbi.nlm.nih.gov/pubmed/10847595 PMID: 10847595

20. Pike KL. The intonation of American English. Ann Arbor: University of Michigan Publications; 1945.

21. van Son N, Huiskamp TMI, Bosman AJ, Smoorenburg GF. Viseme classifications of Dutch consonants and vowels. J Acoust Soc Am. 1994; 96: 1341–1355. https://doi.org/10.1121/1.411324

22. Campbell CS, Massaro DW. Perception of visible speech: Influence of spatial quantization. Perception. 1997; 26: 627–644. https://doi.org/10.1068/p260627 PMID: 9488886

23. Bernstein L. E., Eberhardt SP, Demorest ME. Judgments of intonation and contrastive stress during lipreading. J Acoust Soc Am. 1986; 80: S78.

24. Green KP. The perception of speaking rate using visual information from a talker's face. Percept Psychophys. 1987; 42: 587–593. https://doi.org/10.3758/bf03207990

25. Soto-Faraco S, Navarra J, Weikum W, Vouloumanos A, Sebastián-Gallés N, Werker JF. Discriminating languages by speech-reading. Percept Psychophys. 2007; 69: 218–231. https://doi.org/10.3758/bf03193744 PMID: 17557592

26. Weikum WM, Vouloumanos A, Navarra J, Soto-Faraco S, Sebastián-Gallés N, Werker JF. Visual language discrimination in infancy. Science. 2007; 316: 1159. https://doi.org/10.1126/science.1137686 PMID: 17525331

27. Paulesu E, Perani D, Blasi V, Silani G, Borghese N a, De Giovanni U, et al. A functional-anatomical model for lipreading. J Neurophysiol. 2003; 90: 2005–2013. https://doi.org/10.1152/jn.00926.2002 PMID: 12750414

28. Pritchett BL, Hoeflin C, Koldewyn K, Dechter E, Fedorenko E. High-level language processing regions are not engaged in action observation or imitation. J Neurophysiol. 2018; jn.00222.2018. https://doi.org/10.1152/jn.00222.2018 PMID: 30156457

29. Downing PE, Peelen M V. The role of occipitotemporal body-selective regions in person perception. Cogn Neurosci. 2011; 2: 186–203. https://doi.org/10.1080/17588928.2011.582945 PMID: 24168534

30. Jastorff J, Orban G a. Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. J Neurosci. 2009; 29: 7315–29. https://doi.org/10.1523/JNEUROSCI.4870-08.2009 PMID: 19494153

31. Reinl M, Bartels A. Face processing regions are sensitive to distinct aspects of temporal sequence in facial dynamics. Neuroimage. The Authors; 2014; 102: 407–415. https://doi.org/10.1016/j.neuroimage.2014.08.011 PMID: 25132020

32. Grill-Spector K, Weiner KS. The functional architecture of the ventral temporal cortex and its role in categorization. Nat Rev Neurosci. Nature Publishing Group; 2014; 15: 536–548. https://doi.org/10.1038/nrn3747 PMID: 24962370

33. Downing PE, Jiang Y, Shuman M, Kanwisher N. A cortical area selective for visual processing of the human body. Science. 2001; 293: 2470–3. https://doi.org/10.1126/science.1063414 PMID: 11577239

34. Epstein R, Harris a, Stanley D, Kanwisher N. The parahippocampal place area: recognition, navigation, or encoding? Neuron. 1999; 23: 115–25. Available: http://www.ncbi.nlm.nih.gov/pubmed/10402198 PMID: 10402198

35. Kanwisher N, McDermott J, Chun MM. The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. J Neurosci. 1997; 17. Available: http://www.jneurosci.org/content/17/11/4302

36. Santi A, Servos P, Vatikiotis-Bateson E, Kuratate T, Munhall K. Perceiving biological motion: dissociating visible speech from walking. J Cogn Neurosci. 2003; 15: 800–9. https://doi.org/10.1162/089892903322370726 PMID: 14511533

37. Calvert GA, Campbell R. Reading Speech from Still and Moving Faces: The Neural Substrates of Visible Speech. 2003; 57–70.

38. Bernstein LE, Jiang J, Pantazis D, Lu ZL, Joshi A. Visual phonetic processing localized using speech and nonspeech face gestures in video and point-light displays. Hum Brain Mapp. 2011; 32: 1660–1676. https://doi.org/10.1002/hbm.21139 PMID: 20853377

39. Zhu LL, Beauchamp MS. Mouth and Voice: A Relationship between Visual and Auditory Preference in the Human Superior Temporal Sulcus. J Neurosci. 2017; 37: 2697–2708. https://doi.org/10.1523/JNEUROSCI.2914-16.2017 PMID: 28179553

40. Campbell R, MacSweeney M, Surguladze S, Calvert G, McGuire P, Suckling J, et al. Cortical substrates for the perception of face actions: An fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). Cogn Brain Res. 2001; 12: 233–243. https://doi.org/10.1016/S0926-6410(01)00054-4

41. Bernstein LE, Liebenthal E. Neural pathways for visual speech perception. Front Neurosci. 2014; 8: 1–18.

42. Hauswald A, Lithari C, Collignon O, Leonardelli E, Weisz N. A Visual Cortical Network for Deriving Phonological Information from Intelligible Lip Movements. Curr Biol. Elsevier Ltd.; 2018; 28: 1453–1459.e3. https://doi.org/10.1016/j.cub.2018.03.044 PMID: 29681475

43. Hagoort P. Nodes and networks in the neural architecture for language: Broca's region and beyond. Curr Opin Neurobiol. Elsevier Ltd; 2014; 28: 136–141. https://doi.org/10.1016/j.conb.2014.07.013

44. Fridriksson J, Moss J, Davis B, Baylis GC, Bonilha L, Rorden C. Motor speech perception modulates the cortical language areas. Neuroimage. 2008; 41: 605–613. https://doi.org/10.1016/j.neuroimage.2008.02.046 PMID: 18396063

45. Nishitani N, Hari R. Viewing Lip Forms: Cortical Dynamics motor cortex, both during execution of hand actions. Neuron. 2002; 36: 1211–1220. Available: http://ac.els-cdn.com/S0896627302010899/1-s2.0-S0896627302010899-main.pdf?_tid=75485b5e-4163-11e7-ab6e-00000aacb35f&acdnat=1495728288_7ab4b401a8f5c1271e68c1fbd55ec075

46. Ruytjens L, Albers F, Van Dijk P, Wit H, Willemsen A. Neural responses to silent lipreading in normal hearing male and female subjects. Eur J Neurosci. 2006; 24: 1835–1844. https://doi.org/10.1111/j.1460-9568.2006.05072.x PMID: 17004947

47. Watkins KE, Strafella AP, Paus T. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia. 2003; 41: 989–94. Available: http://www.ncbi.nlm.nih.gov/pubmed/12667534 PMID: 12667534

48. Calvert G. Activation of auditory cortex during silent lip- reading. Science (80-). 1997; 276: 593–596. https://doi.org/10.1126/science.276.5312.593

49. Turner TH, Fridriksson J, Baker J, Eoute D, Bonilha L, Rorden C. Obligatory Broca's area modulation associated with passive speech perception. Neuroreport. 2009; 20: 492–496. https://doi.org/10.1097/WNR.0b013e32832940a0

50. Mahon BZ, Caramazza A. What drives the organization of object knowledge in the brain? Trends Cogn Sci. Elsevier Ltd; 2011; 15: 97–103. https://doi.org/10.1016/j.tics.2011.01.004 PMID: 21317022

51. Lingnau A, Downing PE. The lateral occipitotemporal cortex in action. Trends Cogn Sci. Elsevier Ltd; 2015; 19: 268–277. https://doi.org/10.1016/j.tics.2015.03.006 PMID: 25843544

52. Galantucci B, Fowler C, Turvey MT. "The motor theory of speech perception reviewed": Erratum. Psychon Bull Rev. 2006; 13: 742. https://doi.org/10.3758/BF03193857 PMID: 17048719

53. Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition. 1985; 21: 1–36. Available: http://www.ncbi.nlm.nih.gov/pubmed/4075760 PMID: 4075760

54. Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychol Rev. 1967; 74: 431–61. Available: http://www.ncbi.nlm.nih.gov/pubmed/4170865 PMID: 4170865

55. Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E. Neural processes underlying perceptual enhancement by visual speech gestures. Neuroreport. 2003; 14: 2213–8. https://doi.org/10.1097/00001756-200312020-00016 PMID: 14625450

56. Skipper JI, Nusbaum HC, Small SL. Listening to talking faces: Motor cortical activation during speech perception. Neuroimage. 2005; 25: 76–89. https://doi.org/10.1016/j.neuroimage.2004.11.006 PMID: 15734345

57. Viviani P. Motor competence in the perception of dynamic events: A tutorial. In: Prinz W, Hommel B, editors. Common mechanisms in perception and action: Attention and performance XIX. Oxford: Oxford University Press; 2002. pp. 406–446.

58. Negri GAL, Rumiati RI, Zadini A, Ukmar M, Mahon BZ, Caramazza A. What is the role of motor simulation in action and object recognition? Evidence from apraxia. Cogn Neuropsychol. 2007; 24: 795–816. https://doi.org/10.1080/02643290701707412 PMID: 18161497

59. Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci. 2009; 12: 718–724. https://doi.org/10.1038/nn.2331 PMID: 19471271

60. Venezia JH, Hickok G. Mirror neurons, the motor system and language: From the motor theory to embodied cognition and beyond. Linguist Lang Compass. 2009; 3: 1403–1416. https://doi.org/10.1111/j.1749-818X.2009.00169.x

61. Lemon R. Is the mirror cracked? Brain. 2015; 138: 2109–2111. https://doi.org/10.1093/brain/awv131

62. Wurm MF, Lingnau A. Decoding Actions at Different Levels of Abstraction. J Neurosci. 2015; 35: 7727–7735. https://doi.org/10.1523/JNEUROSCI.0188-15.2015 PMID: 25995462

63. Cerri G, Cabinio M, Blasi V, Borroni P, Iadanza A, Fava E, et al. The mirror neuron system and the strange case of Broca's area. Hum Brain Mapp. 2015; 36: 1010–1027. https://doi.org/10.1002/hbm.22682

64. Flinker A, Korzeniewska A, Shestyuk AY, Franaszczuk PJ, Dronkers NF, Knight RT, et al. Redefining the role of Broca's area in speech. Proc Natl Acad Sci. 2015; 112: 2871–2875. https://doi.org/10.1073/pnas.1414491112

65. Venezia JH, Saberi K, Chubb C, Hickok G. Response bias modulates the speech motor system during syllable discrimination. Front Psychol. 2012; 3: 0–13. https://doi.org/10.3389/fpsyg.2012.00157 PMID: 22723787

66. Fedorenko E, Varley R. Language and thought are not the same thing: Evidence from neuroimaging and neurological patients. Ann N Y Acad Sci. 2016; 1369: 132–153. https://doi.org/10.1111/nyas.13046 PMID: 27096882

67. Fedorenko E, Duncan J, Kanwisher N. Language-selective and domain-general regions lie side by side within Broca's area. Curr Biol. Elsevier Ltd; 2012; 22: 2059–2062. https://doi.org/10.1016/j.cub.2012.09.011

68. Rogalsky C, Hickok G. The Role of Broca's Area in Sentence Comprehension. J Cogn Neurosci. 2011; 23: 1664–1680. https://doi.org/10.1162/jocn.2010.21530

69. Schippers MB, Roebroeck A, Renken R, Nanetti L, Keysers C. Mapping the information flow from one brain to another during gestural communication. Proc Natl Acad Sci U S A. 2010; 107: 9388–9393. https://doi.org/10.1073/pnas.1001791107 PMID: 20439736

70. Cross ES, Liepelt R, Antonia AF, Parkinson J, Ramsey R, Stadler W, et al. Robotic movement preferentially engages the action observation network. Hum Brain Mapp. 2012; 33: 2238–2254. https://doi.org/10.1002/hbm.21361 PMID: 21898675

71. Bobick AF. Movement, activity and action: the role of knowledge in the perception of motion. Philos Trans R Soc London—Ser B Biol Sci. 1997; 352: 1257–1265. https://doi.org/10.1098/rstb.1997.0108 PMID: 9304692

72. Gardner T, Goulden N, Cross ES. Dynamic modulation of the action observation network by movement familiarity. J Neurosci. 2015; 35: 1561–72. https://doi.org/10.1523/JNEUROSCI.2942-14.2015 PMID: 25632133

73. Henson R. Efficient Experimental Design for fMRI. Text. 2006; 193–210.

74. Friston KJ, Zarahn E, Josephs O, Henson RNA, Dale AM. Stochastic Designs in Event-Related fMRI. Neuroimage. 1999; 10: 607–619. https://doi.org/10.1006/nimg.1999.0498 PMID: 10547338

75. Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RN. A critique of functional localisers. Neuroimage. 2006; 30: 1077–1087. https://doi.org/10.1016/j.neuroimage.2005.08.012 PMID: 16635579

76. Macmillan Na Creelman CD. Detection Theory: A User's Guide. Detection theory: A user's guide (2nd ed). 2005.

77. Mikl M, Mareček R, Hluštík P, Pavlicová M, Drastich A, Chlebus P, et al. Effects of spatial smoothing on fMRI group inferences. Magn Reson Imaging. 2008; 26: 490–503. https://doi.org/10.1016/j.mri.2007.08.006 PMID: 18060720

78. Lindquist MA. The Statistical Analysis of fMRI Data. Stat Sci. 2008; 23: 439–464. https://doi.org/10.1214/09-STS282

79. Ashburner J, Friston KJ. Unified segmentation. Neuroimage. 2005; 26: 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018 PMID: 15955494

80. Dale A, Buckner R. Selective averaging of rapidly presented individual trials using fMRI. Hum Brain Mapp. 1997; Available: http://math.bu.edu/people/horacio/tutorials/DaleBucknerHumBrainMap1997_fmri_evtrelated.pdf

81. Murphy K, Garavan H. Artifactual fMRI group and condition differences driven by performance confounds. Neuroimage. 2004; 21: 219–228. https://doi.org/10.1016/j.neuroimage.2003.09.016 PMID: 14741659

82. Daselaar SM. Neuroanatomical correlates of episodic encoding and retrieval in young and elderly subjects. Brain. 2003; 126: 43–56. https://doi.org/10.1093/brain/awg005 PMID: 12477696

83. Penny W, Holmes A. Random-Effects Analysis. Human Brain Function: Second Edition. 2003. https://doi.org/10.1016/B978-012264841-0/50044-5

84. Woo CW, Krishnan A, Wager TD. Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. Neuroimage. Elsevier Inc.; 2014; 91: 412–419. https://doi.org/10.1016/j.neuroimage.2013.12.058 PMID: 24412399

85. Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC. Assessing the significance of focal activations using their spatial extent. Hum Brain Mapp. 1994; 1: 214–220.

86. Callan DE, Jones JA, Callan A. Multisensory and modality specific processing of visual speech in different regions of the premotor cortex. Front Psychol. 2014; 5: 1–10.

87. Sunaert S, Van Hecke P, Marchal G, Orban G a. Motion-responsive regions of the human brain. Exp Brain Res. 1999; 127: 355–370. https://doi.org/10.1007/s002210050804 PMID: 10480271

88. Jastorff J, Orban GA. Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. J Neurosci. 2009; 29: 7315–29. https://doi.org/10.1523/JNEUROSCI.4870-08.2009 PMID: 19494153

89. Weiner KS, Grill-Spector K. Neural representations of faces and limbs neighbor in human high-level visual cortex: Evidence for a new organization principle. Psychol Res. 2013; 77: 74–97. https://doi.org/10.1007/s00426-011-0392-x PMID: 22139022

90. Borowiak K, Schelinski S, von Kriegstein K. Recognizing visual speech: Reduced responses in visual-movement regions, but not other speech regions in autism. NeuroImage Clin. Elsevier; 2018; 20: 1078–1091. https://doi.org/10.1016/j.nicl.2018.09.019 PMID: 30368195

91. Worsley KJ, Marrett S, Neelin P, Vandal aC, Friston KJ, Evans aC. A unified statistical approach for determining significant voxels in images of cerebral activation. Hum Brain Mapp. 1996; 4: 58–73. https://doi.org/10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O

92. Friston KJ. Eigenimages and multivariate analyses. Hum Brain Mapp. 1997; 1–17.

93. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd Edition). New Jersey: Lawrence Erlbaum Associates; 1988.

94. Bakeman R. Recommended effect size statistics for repeated measures designs. Behav Res Methods. 2005; 37: 379–384. https://doi.org/10.3758/bf03192707 PMID: 16405133

95. Wilkinson L, Task Force on Statistical Inference. Statistical methods in psychology journals: Guidelines and explanations. Am Psychol. 1999; 54: 594–604. https://doi.org/10.1037/0003-066X.54.8.594

96. Wu T, Long X, Wang L, Hallett M, Zang Y, Li K, et al. Functional connectivity of cortical motor areas in the resting state in Parkinson's disease. Hum Brain Mapp. 2011; 32: 1443–1457. https://doi.org/10.1002/hbm.21118

97. Xiang H-D, Fonteijn HM, Norris DG, Hagoort P. Topographical functional connectivity pattern in the perisylvian language networks. Cereb Cortex. 2010; 20: 549–60. https://doi.org/10.1093/cercor/bhp119 PMID: 19546155

98. Bernstein LE, Demorest ME, Tucker PE. Speech perception without hearing. Percept Psychophys. 2000; 62: 233–252. https://doi.org/10.3758/bf03205546 PMID: 10723205

99. Bhat J, Pitt MA, Shahin AJ. Visual context due to speech-reading suppresses the auditory response to acoustic interruptions in speech. Front Neurosci. 2014; 8: 1–9.

100. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995; 270: 303–4. Available: http://www.ncbi.nlm.nih.gov/pubmed/7569981 PMID: 7569981

101. Viviani P, Figliozzi F, Campione GC, Lacquaniti F. Detecting temporal reversals in human locomotion. Exp Brain Res. 2011; 214: 93–103. https://doi.org/10.1007/s00221-011-2809-6 PMID: 21814834

102. Binder JR, Desai RH, Graves WW, Conant LL. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex. 2009; 19: 2767–2796. https://doi.org/10.1093/cercor/bhp055 PMID: 19329570

103. Pinsk MA, Arcaro M, Weiner KS, Kalkus JF, Inati SJ, Gross CG, et al. Neural Representations of Faces and Body Parts in Macaque and Human Cortex: A Comparative fMRI Study. J Neurophysiol. 2008; 101: 2581–2600. https://doi.org/10.1152/jn.91198.2008 PMID: 19225169

104. Sergent J, Ohta S, Macdonald B. Functional Neuroanatomy of Face and Object Processing. Brain. 1992; 115: 15–36. https://doi.org/10.1093/brain/115.1.15 PMID: 1559150

105. Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL. The functional organization of human extrastriate cortex: a PET- rCBF study of selective attention to faces and locations. J Neurosci. 1994; 14: 6336–6353. https://doi.org/10.1523/JNEUROSCI.14-11-06336.1994 PMID: 7965040

106. Grossman ED, Blake R. Brain Areas Active during Visual Perception of Biological Motion. Neuron. 2002; 35: 1167–75. Available: http://www.ncbi.nlm.nih.gov/pubmed/12354405 PMID: 12354405

107. Maffei V, Indovina I, Macaluso E, Ivanenko YP, Orban GA, Lacquaniti F. Visual gravity cues in the interpretation of biological movements: Neural correlates in humans. Neuroimage. Elsevier Inc.; 2015; 104: 221–230. https://doi.org/10.1016/j.neuroimage.2014.10.006 PMID: 25315789

108. Maffei V, Giusti MA, Macaluso E, Lacquaniti F, Viviani P. Unfamiliar Walking Movements Are Detected Early in the Visual Stream: An fMRI Study. Cereb Cortex. 2015; 25: 2022–2034. https://doi.org/10.1093/cercor/bhu008 PMID: 24532318

109. Furl N, Van Rijsbergen NJ, Kiebel SJ, Friston KJ, Treves A, Dolan RJ. Modulation of perception and brain activity by predictable trajectories of facial expressions. Cereb Cortex. 2010; 20: 694–703. https://doi.org/10.1093/cercor/bhp140 PMID: 19617291

110. Haxby J V., Hoffman EA, Gobbini MI. The distributed human neural system for face perception. Trends Cogn Sci. 2000; 4: 223–233. https://doi.org/10.1016/s1364-6613(00)01482-0 PMID: 10827445

111. Summerfield C, Trittschuh EH, Monti JM, Mesulam M-M, Egner T. Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci. 2008; 11: 1004–1006. https://doi.org/10.1038/nn.2163 PMID: 19160497

112. Friston K. The free-energy principle: a unified brain theory? Nat Rev Neurosci. Nature Publishing Group; 2010; 11: 127–138. https://doi.org/10.1038/nrn2787 PMID: 20068583

113. Egner T, Monti JM, Summerfield C. Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream. J Neurosci. 2010; 30: 16601–16608. https://doi.org/10.1523/JNEUROSCI.2770-10.2010 PMID: 21147999

114. Koster-Hale J, Saxe R. Theory of Mind: A Neural Prediction Problem. Neuron. Elsevier Inc.; 2013; 79: 836–848. https://doi.org/10.1016/j.neuron.2013.08.020 PMID: 24012000

115. Balestrucci P, Daprati E, Lacquaniti F, Maffei V. Effects of visual motion consistent or inconsistent with gravity on postural sway. Exp Brain Res. Springer Berlin Heidelberg; 2017; 0: 1–12. https://doi.org/10.1007/s00221-017-4942-3 PMID: 28326440

116. Capek CM, MacSweeney M, Woll B, Waters D, McGuire PK, David AS, et al. Cortical circuits for silent speechreading in deaf and hearing people. Neuropsychologia. 2008; 46: 1233–1241. https://doi.org/10.1016/j.neuropsychologia.2007.11.026 PMID: 18249420

117. Perani D, Paulesu E, Galles NS, Dupoux E, Dehaene S, Bettinardi V, et al. The bilingual brain. Proficiency and age of acquisition of the second language. Brain. 1998; 121: 1841–1852. https://doi.org/10.1093/brain/121.10.1841 PMID: 9798741

118. Fuertinger S, Horwitz B, Simonyan K. The functional connectome of speech control. PLoS Biol. 2015; 13: 1–31. https://doi.org/10.1371/journal.pbio.1002209 PMID: 26204475

119. Musso M, Moro A, Glauche V, Rijntjes M, Reichenbach J, Büchel C, et al. Broca's area and the language instinct. Nat Neurosci. 2003; 6: 774–781. https://doi.org/10.1038/nn1077

120. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos Trans R Soc B Biol Sci. 2001; 356: 1293–1322. https://doi.org/10.1098/rstb.2001.0915 PMID: 11545704

121. Amunts K, Zilles K. Architecture and organizational principles of Broca's region. Trends in Cognitive Sciences. 2012. pp. 418–426. https://doi.org/10.1016/j.tics.2012.06.005

122. Shapiro K a, Pascual-Leone a, Mottaghy FM, Gangitano M, Caramazza a. Grammatical distinctions in the left frontal cortex. J Cogn Neurosci. 2001; 13: 713–20. https://doi.org/10.1162/08989290152541386 PMID: 11564316

123. Caramazza A, Zurif EB. Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. Brain Lang. 1976; 3: 572–582. https://doi.org/10.1016/0093-934x(76)90048-1 PMID: 974731

124. Rizzolatti G, Arbib MA. Language within our grasp. 1998; 2236: 1667–1669. https://doi.org/10.1016/S0166-2236(98)01260-0

125. Rizzolatti G, Craighero L. The mirror-neuron system. Annu Rev Neurosci. 2004; 27: 169–92. https://doi.org/10.1146/annurev.neuro.27.070203.144230 PMID: 15217330

126. Pulvermüller F, Fadiga L. Active perception: sensorimotor circuits as a cortical basis for language. Nat Rev Neurosci. Nature Publishing Group; 2010; 11: 351–360. https://doi.org/10.1038/nrn2811 PMID: 20383203

127. Fedorenko E, Nieto-Castañón A, Kanwisher N. Syntactic processing in the human brain: What we know, what we don't know, and a suggestion for how to proceed. Brain Lang. 2012; 120: 187–207. https://doi.org/10.1016/j.bandl.2011.01.001

128. Hickok G, Rogalsky C. What Does Broca's Area Activation to Sentences Reflect? J Cogn Neurosci. 2011; 23: 2629–2631. https://doi.org/10.1162/jocn_a_00044

129. Basilakos A, Smith KG, Fillmore P, Fridriksson J, Fedorenko E. Functional Characterization of the Human Speech Articulation Network. Cereb Cortex. 2017; 1–15. https://doi.org/10.1093/cercor/bhw362

130. Iacoboni M. Cortical Mechanisms of Human Imitation. Science (80-). 1999; 286: 2526–2528. https://doi.org/10.1126/science.286.5449.2526 PMID: 10617472

131. Hasson U, Skipper JI, Nusbaum HC, Small SL. Abstract coding of audiovisual speech: beyond sensory representation. Neuron. 2007; 56: 1116–26. https://doi.org/10.1016/j.neuron.2007.09.037 PMID: 18093531

132. Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Tarkiainen A, et al. Primary auditory cortex activation by visual speech: an fMRI study at 3 T. Neuroreport. 2005; 16: 125–128. https://doi.org/10.1097/00001756-200502080-00010 PMID: 15671860

133. Skipper JI, Van Wassenhove V, Nusbaum HC, Small SL. Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. Cereb Cortex. 2007; 17: 2387–2399. https://doi.org/10.1093/cercor/bhl147 PMID: 17218482

134. Frey S, Campbell JSW, Pike GB, Petrides M. Dissociating the Human Language Pathways with High Angular Resolution Diffusion Fiber Tractography. J Neurosci. 2008; 28: 11435–11444. https://doi.org/10.1523/JNEUROSCI.2388-08.2008 PMID: 18987180

135. Campbell R. Speechreading and the Bruce-Young model of face recognition: Early findings and recent developments. Br J Psychol. 2011; 102: 704–710. https://doi.org/10.1111/j.2044-8295.2011.02021.x PMID: 21988379

136. Ponton CW, Bernstein LE, Auer ET. Mismatch negativity with visual-only and audiovisual speech. Brain Topogr. 2009; 21: 207–215. https://doi.org/10.1007/s10548-009-0094-5 PMID: 19404730

137. Venezia JH, Fillmore P, Matchin W, Lisette Isenberg A, Hickok G, Fridriksson J. Perception drives production across sensory modalities: A network for sensorimotor integration of visual speech. Neuroimage. Elsevier B.V.; 2016; 126: 196–207. https://doi.org/10.1016/j.neuroimage.2015.11.038 PMID: 26608242

138. Bernstein LE, Auer ET, Wagner M, Ponton CW. Spatiotemporal dynamics of audiovisual speech processing. Neuroimage. 2008; 39: 423–435. https://doi.org/10.1016/j.neuroimage.2007.08.035 PMID: 17920933