

# SCIENTIFIC REPORTS



OPEN

## Functional and spatial segregation within the inferior frontal and superior temporal cortices during listening, articulation imagery, and production of vowels

Alessandra Cecilia Rampinini<sup>1</sup> , Giacomo Handjaras<sup>1</sup> , Andrea Leo<sup>1</sup>, Luca Cecchetti<sup>1</sup> , Emiliano Ricciardi<sup>1</sup> , Giovanna Marotta<sup>2</sup> & Pietro Pietrini<sup>1</sup>

Classical models of language localize speech perception in the left superior temporal and production in the inferior frontal cortex. Nonetheless, neuropsychological, structural and functional studies have questioned such subdivision, suggesting an interwoven organization of the speech function within these cortices. We tested whether sub-regions within frontal and temporal speech-related areas retain specific phonological representations during *both* perception and production. Using functional magnetic resonance imaging and multivoxel pattern analysis, we showed functional and spatial segregation across the left fronto-temporal cortex during listening, imagery and production of vowels. In accordance with classical models of language and evidence from functional studies, the inferior frontal and superior temporal cortices discriminated among perceived and produced vowels respectively, also engaging in the non-classical, alternative function – i.e. perception in the inferior frontal and production in the superior temporal cortex. Crucially, though, contiguous and non-overlapping sub-regions within these hubs performed either the classical or non-classical function, the latter also representing non-linguistic sounds (i.e., pure tones). Extending previous results and in line with integration theories, our findings not only demonstrate that sensitivity to speech listening exists in production-related regions and *vice versa*, but they also suggest that the nature of such interwoven organisation is built upon low-level perception.

According to classical models of speech processing, superior temporal and inferior frontal brain regions are consistently involved in perception and production, respectively<sup>1</sup>. However, theories dealing with the relationship between perceived and produced speech have long debated whether and to what extent perceptual and articulatory information are integrated in language acquisition and use, either assuming that perception shapes production, or that production influences perception<sup>2,3</sup>. Other proposals have instead suggested that articulatory coherence and perceptual value *both* contribute to a synergic processing of speech in the brain<sup>4</sup>.

The phoneme-specific specialization of the superior temporal cortex in perception, as well as that of a wide, prefrontal territory around Broca's area in production, are well-known, since quite a few seminal studies have explored the neural encoding of phonological competence<sup>5–8</sup>. Interestingly, while the phoneme itself was a theoretical model debated mostly in Linguistics in the last century, many recent studies revealed that brain activity specific to phonological stimuli could be indeed isolated in the classical foci pertaining to perception and production, with both functional neuroimaging and electrophysiology methods<sup>9</sup>: particularly, the superior temporal cortex has been shown to represent the overall acoustic form of syllables<sup>10</sup>, syllable-embedded perceived consonants or vowel categories<sup>11</sup>, and even tones when phonologically marked<sup>12</sup>, while a precise account of motor involvement during production or imagery of phonemes has received less attention in the existing literature<sup>13</sup>.

<sup>1</sup>IMT School for Advanced Studies, Lucca, 55100, Italy. <sup>2</sup>Department of Philology, Literature and Linguistics, University of Pisa, Pisa, 56100, Italy. Alessandra Cecilia Rampinini and Giacomo Handjaras contributed equally to this work. Correspondence and requests for materials should be addressed to E.R. (email: [emiliano.ricciardi@imtlucca.it](mailto:emiliano.ricciardi@imtlucca.it))

Such rich and mixed picture sparked other questions: do distinct brain regions whatsoever support different aspects of speech processing (such as perception, imagery and production of phonemes)? Do they share specific phonological representations? In the context of theories debating an interwoven organization of speech perception and production, the Motor Theory of Speech Perception (MTSP)<sup>3</sup> has argued in favour of a covert articulatory rehearsal mechanism, which would take place implicitly and automatically whenever a speaker is exposed to language, thus connecting the two ends of the perception-production continuum. Such mechanism was substantiated by findings generalized to other processes, crucially including motor control<sup>14</sup>.

In this respect, functional neuroimaging and electrophysiological studies have recently sought to determine the relationship between the perceptual and articulatory stages of speech, seeking perception-related information in frontal areas engaged by production tasks, and production-related information in temporal areas engaged by perception tasks<sup>15–20</sup>. In these studies, multivariate analyses were exploited to reveal similarities in informational content between regions previously inferred to perform different functions (through classical activation experiments), thus revealing a mixed picture of shared information and cortical space as well, and tangentially supporting integration models such as those described.

Similarly, virtual<sup>21</sup> and real lesion studies failed to validate an exact correspondence between language impairments and information represented in the frontotemporal speech network: damage in one area may, or may not, entail loss of function in the other, as even sub-regions within such well-known perimeters appear to support different functions<sup>22–25</sup>. The idea of an interwoven cortical organization of speech function is also favoured by structural studies that reveal a fine-grained cytoarchitectonic, connectivity- and receptor-mapping-based parcellation of fronto-temporal language areas<sup>26–31</sup>. Therefore, disentangling the nature of the perception-production interface appears far from straightforward.

According to these indications, we tested whether sub-regions within the frontal and temporal speech areas retain specific, functionally segregated phonological representations during both perception and production, and whether a possible covert rehearsal mechanism could be elicited, through articulation imagery, to simulate the production-perception interface postulated by the MTSP (in contrast with hearing imagery<sup>32,33</sup>).

To this aim, using functional Magnetic Resonance Imaging (fMRI) and multivoxel pattern analysis (MVPA), we measured the spatial overlap of brain regions involved in stimulus-specific representations during vowel perception (listening), and production (imagined and overt articulation). Within a set of phonemes, the basic units of words, we selected vowels since they retain acoustic features (i.e., formants) that can combine together, thus to distinguish them in a discrete manner. Moreover, formant combinations emerge from unique articulatory gestures, so that their processing depends upon the same perceptuo-motor model<sup>34</sup>, differently from consonants<sup>5,6,20</sup>. Particularly, while consonants need to be embedded in syllables to be fully heard and articulated, vowels are self-standing phonemes with high salience. Vowels act as syllabic nuclei, prosodic aggregating centres and ultimately, they can carry stress (whereas consonants cannot), around which the phonic profile of words organizes<sup>34</sup>. Therefore, vowels offer an interesting perspective to investigate the workings of the perceptual and motor stages of speech.

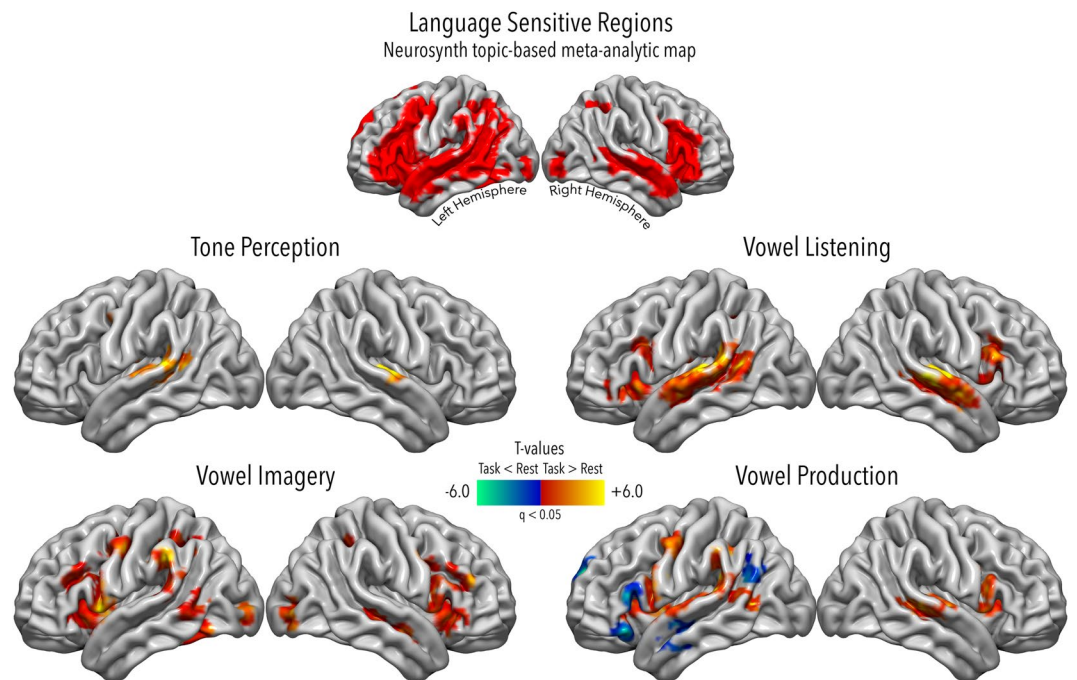
Thus, building on previous knowledge on phoneme representation in the brain, we tried to provide a finer characterization of the fronto-temporal language cortex: in fact, we compared modalities of perception, production and articulation imagery within the same pipeline and testing them with a complex vowel model, where all items carry equal complexity. Crucially, we also assessed whether sub-regions within the frontal and temporal hubs of the speech network support high-level, fully phonological representations of vowels exclusively, rather than sharing sensitivity to lower-level acoustic stimuli (pure tones), not pertaining to categorical perception of the salient, linguistic kind.

## Results

**Univariate results.** To show regions activated by each of the four tasks, brain activity related to tone perception, vowel listening, imagery and production was contrasted with the resting condition ( $p < 0.05$ , corrected for False Discovery Rate<sup>35,36</sup> - FDR), within a topic-based meta-analytic mask of language-sensitive regions selected from the Neurosynth database<sup>37</sup>.

Figure 1 shows the results of this procedure and the extension of the mask. Particularly, the tone perception task activated the bilateral primary auditory cortex (Heschl's gyrus, HG) extending to the superior temporal cortex especially in the left hemisphere, along with the superior part of the precentral sulcus (PrCS) at the border with the precentral gyrus (PrCG). In the vowel listening task, HG and superior temporal cortex were activated bilaterally, with more posterior activations in the left hemisphere only; in the frontal cortex, this task activated the left inferior frontal sulcus (IFS) and the opercular portion of the inferior frontal cortex, the insular cortex (INS), and the horizontal ramus of the sylvian fissure, the right *pars opercularis* of the inferior frontal gyrus (IFGpOp), and a small part of the IFS. In the vowel imagery task the frontal cortex was activated in the bilateral (though mostly left) PrCS, left IFS and PrCG, right middle frontal gyrus (MFG/IFS) and bilateral INS; moreover, this task activated significantly the right superior temporal sulcus (STS), left *planum temporale* and supramarginal gyrus (SMG), bilateral intraparietal sulcus (IPS), left posterior middle temporal gyrus (pMTG) and inferior temporal gyrus (ITG), the bilateral middle/inferior occipital gyrus (MOG/IOG), and finally, the bilateral medial portion of the superior frontal gyrus (SFG) and caudate nuclei. The vowel production task significantly activated the bilateral superior temporal cortex extending to the *planum temporale* in the left hemisphere only, the bilateral INS and PrCS, left PrCG, the medial SFG, and left SMG; in this task, significant deactivations were observed in the left hemisphere, particularly in the left *pars orbitalis*, the vertical ramus of the sylvian fissure, the anterior portion of the medial SFG, anterior and posterior portions of the STS.

**Multivariate results.** A multi-class searchlight-based classifier highlighted three sets of clusters, one for each vowel task, where pattern discrimination was successful. Table 1 summarizes the MNI co-ordinates at each cluster's centre of mass. Figure 2 shows clusters on the cortical volume through axial slices, while Fig. 3 shows the accuracy maps of all experimental tasks projected onto the lateral cortical surfaces.



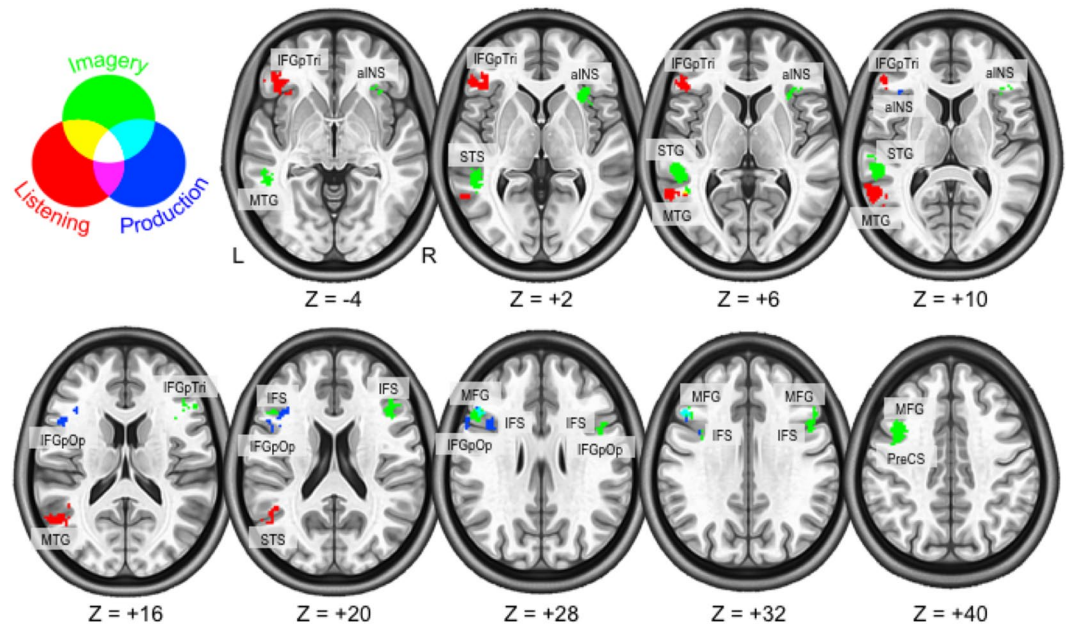
**Figure 1.** Univariate results. Here the results for one-sample, two-tailed t-tests are shown in each of the four tasks against the resting condition ( $p < 0.05$ , FDR corrected). These measures were conducted to assess which regions were activated in each task and restricted to a topic-based meta-analytic mask of language-sensitive regions from the Neurosynth database, whose extension can be appreciated in the top panel of this figure.

TASK	Cluster	Voxels	CMass x	CMass y	CMass z
VOWEL LISTENING	Left pSTS-MTG	263	-56.5	-54.8	+11.3
	Left IFGpTri	257	-46.5	+32.6	-0.3
VOWEL IMAGERY	Left pMTG-STG	352	-53.2	-38.8	+2.5
	Right IFS-MFG	346	+42.9	+21.0	+20.0
	Left IFS-MFG	230	-44.1	+14.5	+33.7
VOWEL PRODUCTION	Left IFS-IFGpOp	211	-45.5	+17.4	+23.9

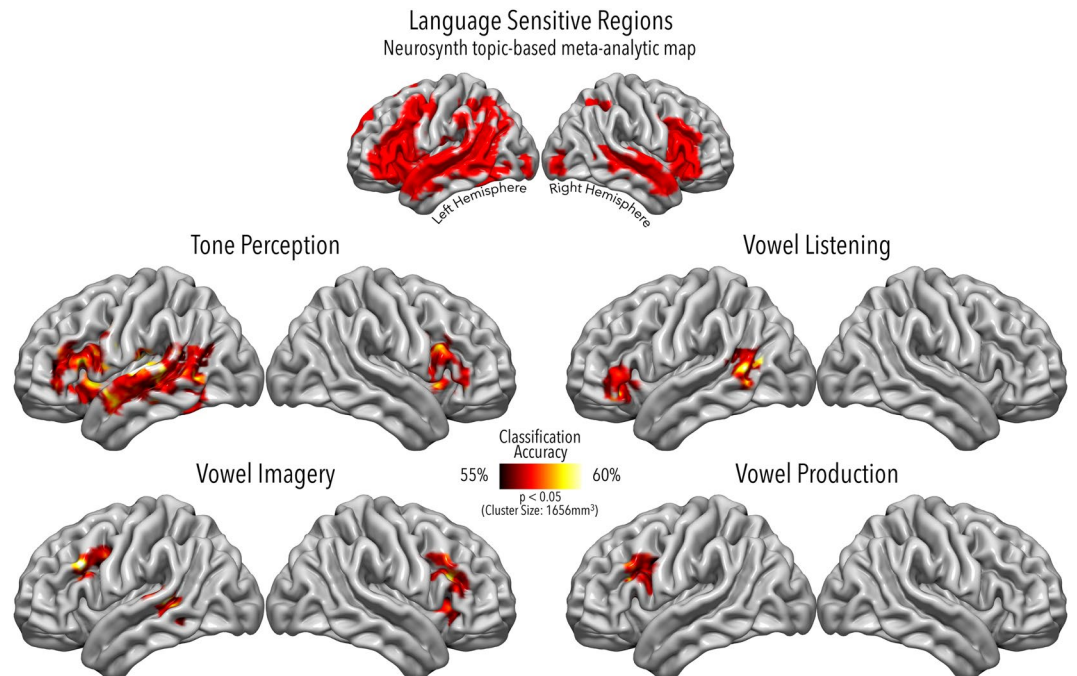
**Table 1.** MNI co-ordinates and centres of mass for the searchlight-based classifier results. Here we show the location and extension of clusters emerging from the classifier run within a mask based on language-related studies in the Neurosynth database. For each task, only significant regions are shown that survived cluster correction. Labels are spelled as follows: pSTS - posterior superior temporal sulcus; pMTG - posterior middle temporal gyrus; IFGpTri - inferior frontal gyrus, pars triangularis; STG - superior temporal gyrus; IFS - inferior temporal sulcus; MFG - middle frontal gyrus; IFGpOp - inferior frontal gyrus, pars opercularis.

**Vowel listening, imagery and production dissociate in the left inferior frontal cortex.** The left inferior frontal cortex (IFG, IFS) was engaged across all experimental conditions, with the addition of the right homologue in the imagery task only. Particularly, though, clusters of voxels within these macro-regions responded specifically to each task (regions were labelled and their overlap with the result masks was interpreted in accordance with the Harvard-Oxford Cortical Atlas). In details, during vowel listening, the *pars triangularis* of the left IFG (IFGpTri) represented vowels, crossing over anteriorly into the *pars orbitalis*. During vowel imagery, the left IFS and its right homologue intersected superiorly the MFG, with a relative overlap with the INS as well. During production, a slightly more posterior region within the left IFS was engaged, running inferiorly into the *pars opercularis* of the IFG, and superiorly into the MFG.

**Vowel listening and imagery dissociate in the superior temporal cortex.** Temporal regions representing vowels revealed that the left STG and STS running posteriorly and inferiorly towards MTG, were engaged in listening, as well as performing imagery of vowels through covert articulation. Particularly, temporal regions representing vowels during listening were the left pSTS, extending into the pMTG. Vowel imagery engaged a close-by portion of the left pMTG extending superiorly into the STG and STS. No temporal regions represented vowels significantly during overt production.



**Figure 2.** Results mapped on the cortical volume. Here, significant searchlight classifier clusters are shown for the vowel tasks, represented on the cortical volume through axial slices. Colours were assigned by task, and any of their possible combinations were indicated as well in the circle legend. The almost complete contiguity of regions can be appreciated, as marginal overlap emerged only between imagery/production and imagery/listening. No voxels revealed to be shared by all three tasks. Labels are spelled as follows: STS - *superior temporal sulcus*; MTG - *middle temporal gyrus*; IFGpTri - *inferior frontal gyrus, pars triangularis*; STG - *superior temporal gyrus*; IFS - *inferior temporal sulcus*; MFG - *middle frontal gyrus*; IFGpOp - *inferior frontal gyrus, pars opercularis*; aINS - *anterior insular cortex*.



**Figure 3.** Accuracy maps projected onto the lateral surfaces of the brain. Here we show regions where accuracy values were significant across the searchlight area defined by the selected Neurosynth topic-based meta-analytic map (top panel) in each task (bottom panels). The extension and location of these regions was validated through cluster correction in AFNI at a minimum cluster size of 207 voxels ( $p < 0.05$  at voxel level with  $\alpha < 0.05$  for the correction for multiple comparisons).



MASK	Task						
	Cluster	Vowel listening		Vowel imagery		Vowel production	
		Acc. $\pm$ SE	<i>p</i>	Acc. $\pm$ SE	<i>p</i>	Acc. $\pm$ SE	<i>p</i>
VOWEL LISTENING	<i>Left pSTS-MTG</i>	56.7 $\pm$ 1.2	<b>&lt;0.001*</b>	52.8 $\pm$ 0.9	0.167	49.8 $\pm$ 1	0.958
	<i>Left IFGpTri</i>	57.3 $\pm$ 0.8	<b>&lt;0.001*</b>	54.1 $\pm$ 1.4	0.018	50.3 $\pm$ 0.9	0.879
VOWEL IMAGERY	<i>Left pMTG-STG</i>	53.1 $\pm$ 1	0.106	57.3 $\pm$ 0.9	<b>&lt;0.001*</b>	51.4 $\pm$ 0.7	0.645
	<i>Right IFS-MFG</i>	52.2 $\pm$ 1.2	0.315	56.9 $\pm$ 1	<b>&lt;0.001*</b>	52.9 $\pm$ 0.9	0.112
	<i>Left IFS-MFG</i>	51.7 $\pm$ 1.2	0.464	57.1 $\pm$ 1.4	<b>&lt;0.001*</b>	53.2 $\pm$ 1.4	0.087
VOWEL PRODUCTION	<i>Left IFS-IFGpOp</i>	52.5 $\pm$ 1	0.215	53.5 $\pm$ 0.9	0.031	56.5 $\pm$ 1	<b>&lt;0.001*</b>

**Table 2.** Cross-task accuracy measures between vowel tasks. Accuracy measures are shown here for each task in its own significant regions, but also compared to the other tasks by constraining the extraction of accuracy values for one task within the areas that were significant in each of the others. Significant values are reported in bold, and gray shading was used to highlight accuracy values within correspondent masks and tasks. Of note, accuracy values were significant only for a task within its own regions, showing no functional overlap between modalities (accuracies were Bonferroni-corrected at  $p_{bonf} < 0.0028$ ). For label spelling, please refer to Table 1.

MASK	Cluster	Tone Perception	
		Acc. $\pm$ SE	<i>p</i>
		VOWEL LISTENING	<i>Left pSTS-MTG</i>
	<i>Left IFGpTri</i>	54.9 $\pm$ 1.5	<b>0.004*</b>
VOWEL IMAGERY	<i>Left pMTG-STG</i>	55.4 $\pm$ 1	<b>0.002*</b>
	<i>Right IFS-MFG</i>	53.0 $\pm$ 1.1	0.112
	<i>Left IFS-MFG</i>	51.8 $\pm$ 0.8	0.466
VOWEL PRODUCTION	<i>Left IFS-IFGpOp</i>	52.9 $\pm$ 1.2	0.134

**Table 3.** Cross-task accuracy measures of pure tone perception within each vowel mask. Tone perception accuracy results were constrained within the masks defined by the vowel classifier. Significant values are reported in bold. Of note, the Left IFGpTri from the vowel listening task and the Left pMTG-STG from the vowel imagery task were also able to represent tones significantly (accuracies were Bonferroni-corrected at  $p_{bonf} < 0.0083$ ). For region labels, please refer to Table 1.

**Measuring cross-task spatial segregation and tone sensitivity.** No spatial overlap among tasks was revealed, except for a cluster of voxels in the IFS/MFG for vowel imagery and production, and a very small cluster in the MTG for vowel imagery and listening. Moreover, cross-task accuracy measurements revealed that the imagery-sensitive left pMTG-STG region also shared tone representations, as well as IFGpTri during vowel listening. Table 2 summarizes cross-task accuracy results from the calculations performed in each cluster from the vowel tasks, with the associated *p* value and standard errors (SE). Table 3 reports cross-task accuracies for the pure tones within the vowel clusters.

## Discussion

In this study, we combined fMRI and MVPA to study the functional organization of vowel listening, imagery and production. We explored the representation of vowels across these three modalities, as well as determining commonalities and differences with a tone perception control task in a frequency range close to that of our speech stimuli. Specifically, patches of cortex in inferior frontal and superior temporal regions retained information to significantly discriminate the seven vowels of the Italian language in each condition. Within these areas, contiguous, and just minimally overlapping clusters were sensitive to listening, articulation imagery and production of speech sounds. Of note, left IFGpTri and left pMTG-STG shared sensitivity to both tones and vowels.

**Functional segregation and tone sensitivity in brain regions involved in vowel listening, imagery and production.** Several functional studies explored the representation of vowels, consonants and syllables in the fronto-temporal language areas (although more often considering one task at a time): some highlighted their sensitivity to very fine-grained aspects of speech, such as formant structure, manner and place of articulation, and even speaker identity<sup>7,8,15,38</sup>, while others have highlighted the importance of a shared neural code for validating popular theories about the acquisition and processing of language<sup>17</sup>. Univariate results comparing each of the four tasks (tone perception, vowel listening, imagery and production) against resting condition highlighted a set of regions in line with previous findings, revealing frontal and temporal involvement in language perception and production<sup>1</sup>. However, while classical univariate approaches sought to infer specific mental function by comparing regional average activations, and thus were amply exploited to investigate the spatial organization of speech, multivariate analyses show representational content similarities over regional engagement: this, together with a comprehensive comparison of speech modalities, can provide a finer characterization of the speech function across the fronto-temporal language cortex.

Theoretical approaches seeking to support integration between perception and production have suggested that production and the socially-rooted need for intelligibility in infants can shape perception<sup>2</sup>, so that we refine our produced speech output ever since the babbling phase, just by hearing others' voices and ours. Alternatively, some have argued that perceived speech would contain articulatory information<sup>3</sup>. In this context, Schwartz and collaborators have tried to reconcile the contrasting ideas that we acquire language by "saying what we should be hearing" or "hearing what we should be saying", fitting perceptual shaping and motor procedural knowledge together in speech processing<sup>4</sup>. Worth mentioning as well is the functional neuroimaging-based argument of Scott and Johnsrude, suggesting the dual nature of speech as both a sound and an action<sup>39</sup>. In this respect, integration theories argue in favour of a covert articulatory rehearsal mechanism bridging the perception-production gap: such mechanism may be of the utmost relevance in linguistic interactions, whose temporally-fast variations have been frequently associated with the complexity and the computational structure of birdsong, thus integrating functions in sensorimotor learning through efference copies<sup>40</sup>. Importantly, an action-perception dual stream originating in the auditory belt and projecting forward to the inferior frontal cortex, and backward to the parietal lobe, has been proposed for language processes by Rauschecker and Scott<sup>41</sup>, possibly supporting functional integration of the perception-production continuum on the basis of structural connections in humans, and functional studies in the monkey (as well as human) model.

Despite the variety of models proposed, it appears that any theory considering the sharing of neural information between perceived and produced speech should provide an assessment of their *spatial* organization in the frontal and temporal hubs of the speech network. Indeed, a vast amount of literature reveals mixed comprehension and production deficits associated with cortical lesions in these locations<sup>22–25,42–44</sup>, and particularly within the inferior frontal wide territory pertaining to an extended view of Broca's area<sup>28</sup>, centring around IFGpOp/IFGpTri, touching the lower bank of the PrCG posteriorly and the INS medially. Davis and collaborators, especially, underline that even though a plethora of clinical studies show deficits broadly recollected under the Broca's aphasia label, not all patients diagnosed with Broca's aphasia have lesions in the IFGpOp/IFGpTri and not all patients with these kinds of lesions do, in fact, present with all (or some of) Broca's aphasia-related symptoms<sup>44</sup>; the complexity of lesions and associated disruption of speech along the fronto-temporal network is also reported by Bates and colleagues<sup>45</sup>.

Moreover, recent interest for combining multivariate methods with functional brain data has revealed that phonological information is finely represented in the fronto-temporal language-related cortex: particularly, the superior temporal cortex has been shown to encode perceived phonological features<sup>46</sup>, discrete speech sound categories<sup>7</sup>, and to preserve the representation based on tongue positions together with formant structure<sup>8</sup>. Additional properties have been decoded from perceived phonemes, such as speaker identity, providing an ever-growing account of the complexity of basic speech representations all along the antero-posterior axis of the superior temporal cortex, bilaterally<sup>8,38</sup>.

On the lines of an integrative account, within the prefrontal hub of the speech network, Cheung and colleagues were able to cross-decode *manner* of articulation, a *perceptual* feature of consonants, in *motor* electrodes tested on data previously extracted from the ventral sensorimotor cortex (vSMC) during the production of syllables<sup>17</sup>. Similarly, the involvement of the superior temporal sulcus in processing, at least coarsely, produced syllables was demonstrated, whereas more frontal recordings showed selective firing to specific vowels categories<sup>15</sup>.

Nonetheless, a complete account of the spatial engagement and informational content representation of different speech modalities within the left fronto-temporal cortex is still needed: along these lines, in this study we aimed at extending electrocorticographic findings to the non-invasiveness allowed by fMRI on healthy participants. Notably, while the accuracy and directness of electrocorticography (ECoG) as a measurement of brain function is, indeed, invaluable, fMRI holds the advantage of providing the functional characterization of multiple modalities (perception, production and imagery) across a larger extent of cortex within the same subject, which is not easy to replicate with intracranial recordings, generally tied to clinical needs.

Therefore, to provide a finer spatial and functional account of phonological processing and the production-perception interface, we ran a searchlight classifier of listened, imagined and produced vowels within a mask of neuroimaging studies of the language function. This procedure aimed at measuring the accuracy of vowel discrimination, and, most importantly, the spatial organization and possible overlap between regions controlling the three vowel tasks. Moreover, with the same procedure we attempted tone classification in frequencies close to those of our speech stimuli. Accuracies yielded by each vowel task were also measured in clusters resulting from the classifiers of all the other vowel tasks, as well as tone perception accuracies were tested in the vowel regions.

Globally, our results revealed that speech tasks are indeed processed within two classically linguistic macro-regions in the frontal and temporal cortices. Particularly, though, we did not find production of vowels confined to the inferior frontal cortex, nor perception confined to the superior temporal cortex. Instead, both the inferior frontal and superior temporal cortices represented vowel-specific information in both perception and production (imagined as well as overt). Nonetheless, the three vowel tasks engaged well-defined, *bordering* sub-portions of the inferior frontal and superior temporal hubs, a picture already sustained by lesion studies and pre-operative language function testing<sup>43</sup>. Moreover, the vowel model was well represented in articulation imagery, a task whose aim was to simulate the articulatory rehearsal mechanism assumed by integration theories: even there, segregated regions revealed sensitivity to vowels in contrast with those clusters, adjoining though non-overlapping, which represented perceived and produced stimuli.

Interestingly, though, while no vowel-sensitive regions retained above-chance accuracies for other tasks, two regions represented tones significantly, that is, the IFGpTri involved in listening and the pSTG-MTG involved in imagery of vowels (of note, the region identified in imagery as being tone-sensitive is spatially closer to the primary auditory cortex than the vowel-specific region identified in vowel listening as pSTS-MTG). This result reveals that, while we have regions *within* the frontal and temporal cortices performing both production-related

and perception-related functions in a segregated fashion, these areas also retain low-level non-linguistic information. Specifically, though, high-level information pertains only to the “classical” function associated to that area (production in the inferior frontal and perception in the superior temporal cortex), while the “non-classical” associated function is not language-specific (perception in the inferior frontal and articulation imagery in the superior temporal cortex).

Therefore, regardless of how, in fact, we may approach the issue of perception shaping production or *vice versa*, such mechanisms seem to be in place because globally we do not have regions for production or perception of speech *as a whole*. Instead, our findings seem to suggest that the brain retains a capacity for sub-specialization within the classical language fronto-temporal hubs. Speculatively, one may argue that comprehension deficits resulting from lesions within the inferior frontal cortex, as well as production deficits resulting from lesions within the superior temporal cortex, may arise from disruption of lower-level information processing.

**Vowel listening, imagery and production dissociate in the left inferior frontal cortex.** Our results showed how vowel listening, as well as vowel imagery and production, engage the left inferior frontal cortex, from the IFGpOp crossing over anteriorly into the IFGpTri, superiorly into the IFS and touching the MFG. Within the right hemisphere, vowel imagery engaged the IFS, MFG and aINS. However, vowel tasks engaged the broad “Broca’s territory” in a functionally segregated fashion: left IFGpOp engaged in vowel production, while the IFS engaged in vowel imagery (as well as its right homologue). Finally, a more anterior region in the IFGpTri engaged in vowel listening although it also represented tones, revealing to be non-specific to speech sounds.

A debate exists on the role of the inferior frontal cortex in processing high- rather than low-level language functions in the healthy brain as well as in lesion studies: this region has been broadly implicated in syntactic working memory<sup>47</sup>, perceptuo-motor integration<sup>48</sup> and phonetic/phonological representations<sup>49,49</sup>. Furthermore, along the lines of a functional segregation argument, IFGpOp and IFGpTri within Broca’s area have been associated, respectively, to processes pertaining to syntax and semantics<sup>50</sup>. Still, early evidence from Positron Emission Tomography had already suggested that Broca’s area is primed by *any* phonological differences subtending semantic representations, and not by the processing of meaning *per se*<sup>51</sup>. Moreover, Heim and collaborators do not report additional activations in IFGpTri for semantic *versus* phonological fluency, with only the latter significantly activating IFGpOp<sup>52</sup>.

Along these lines, some have ascribed the disrupted patterns of both complex syntactic *comprehension* and general speech *production* in Broca’s aphasia to a disturbance in the hierarchical chain-processing mechanism at the basis of the phonological loop, which may be controlled by IFGpOp and possibly IFGpTri<sup>44,53</sup>. Recently, it was proposed that Broca’s area in particular mediates the transformation of perceptual information coming first into the superior temporal cortex, thus to be projected back to the PrCG as articulatory instructions for production<sup>54</sup>.

The idea that locations anterior to the PrCG perform sensorimotor transformations and relay information back to the PrCG is in agreement with our findings. Furthermore, we were able to provide a finer characterization of the functional neuroanatomy of the IFG, showing sensitivity to perceived tones and vowels in the *pars triangularis*, and to produced vowels in the *pars opercularis*. Therefore, our results suggest that the language-related inferior frontal cortex, before anything else that may be of a higher level, is concerned *at least* with the representation of perceived speech, as well as non-speech sounds.

The idea that IFGpTri supports simpler, non-linguistic representations, as we found in the cross-task accuracy measurements between vowel listening and tone perception, was previously hinted at by Reiterer and colleagues, who demonstrated IFGpTri involvement in processing tone frequency though not sound pressure, using a pitch *versus* volume discrimination task<sup>55</sup>. On the other hand, Hickok and colleagues reported how IFG-lesioned patients show no *auditory* syllable discrimination deficits whatsoever<sup>23</sup>. Although this result may appear in disagreement with ours, it is reasonable to speculate that the extensions and locations of lesions (as noted by the authors themselves) do not allow for a full comparison with ours and others’ functional results in the healthy brain (as also advised by Ardila and colleagues<sup>25</sup>).

Regarding the *pars opercularis* as the most posterior cluster showing vowel sensitivity, we found produced vowels represented discretely in IFGpOp. In its proximity, the PrCG has been associated to apraxia of speech<sup>42</sup>, a disturbance in the articulatory aspects of production exclusively. Consistently, we were able to discriminate overtly produced vowels at the posterior border of the IFGpOp extending into the PrCS. Instead, vowel imagery involved more anterior regions for the processing of intermediate phonological representations with no sensory output. These arguments appear to sustain the importance of this inferior frontal region at the perceptuo-motor interface for speech.

All in all, our results suggest that both IFGpOp and IFGpTri do perform *phonological* computations, that is, a sub-lexical kind of processing at the basis of any higher-level function (from syntax to semantics, as already mentioned), and their spatial organization is rather driven by the speech task being performed, with perception and production completely detached, and perception being non-specific to speech sounds.

In fact, some of those trying to reconcile the vast literature on inferior frontal cortex involvement in speech processing have argued that, if its engagement is a matter of perceptuo-motor interface, then the IFG *as a whole* should share activations related to different tasks in the speech loop<sup>56</sup>. This argument has been brought forward particularly by those sustaining that region sharing would constitute a neurofunctional correlate of mainframes such as the MTSP<sup>3</sup>. Our results, instead, reveal functional *dissociation* within the inferior frontal cortex for different tasks related to speech sound discrimination, and clarify at least the correlation of *both* IFGpOp and IFGpTri with phonological-level functions.

The processing of produced and imagined speech in close-by regions, as well as more anterior and more rightward activations for imagined speech, were previously reported<sup>57,58</sup>. In our results, we found a cluster of spatial overlap between the regions involved in produced and imagined vowels in the IFS/MFG. This location’s centre of mass was associated to cognitive processes related to working memory in the Neurosynth database (highest

posterior probability: ‘retrieved’ 0.77, ‘memory retrieval’ 0.76, ‘wm task’ 0.76). Of note, our subjects were asked to maintain and then retrieve a heard vowel thus to perform imagery or production, and the searchlight analysis was then conducted on the retrieval phase of the trials. In this sense, the small cluster of spatial overlap that we found between production and imagery could be explained as a common *focus* for the mnemonic-attentive component of the task (vowel retrieval). To reinforce this argument, cross-task accuracy measurements did not reveal shared sensitivity to produced *and* imagined vowels in this region, instead showing complete dissociation: in fact, that cluster of spatial overlap may be shared by the production and imagery-sensitive clusters for task-specific demands, and not information content representation.

Finally, the involvement of the right IFS-MFG homologue, as well as aINS, in the imagery task would be justifiable in that these regions were shown to be involved in mental/imagined speech<sup>59</sup> and aphasia recovery in left IFG/IFS-lesioned patients<sup>57,60</sup>.

**Vowel listening and imagery dissociate in the superior temporal cortex.** In our study, the left superior and middle temporal cortices were largely engaged by vowel listening and vowel imagery. Regarding the engagement of the superior temporal cortex in perceived speech, a large body of evidence suggests that this region retains sensitivity to complex harmonic structures and, generally, spectral features down to a stimulus-specific level, studied with both fMRI<sup>8,38</sup> and ECoG<sup>7,46,61</sup>. The superior temporal cortex has been associated also to imagery of speech, arguing that the pSTG-pSTS-MTG macro-region supports *both* imagery and perception<sup>62,63</sup>. Interestingly, though, our results showed that vowel listening and vowel imagery dissociate spatially, as in the inferior frontal cortex; moreover, pSTG-MTG retains tone-specific representations as well as imagined vowels. This reveals how, in the superior temporal cortex as well as the inferior frontal, the function classically associated to the region is language-specific, while the non-classical function shares sensitivity to lower-level stimuli.

Among those who argued in favour of an integrated model, Murakami and colleagues<sup>64</sup> found that repetitive transcranial magnetic stimulation over the left superior temporal cortex can disrupt phonological fluency, in that it suppresses muscular evoked potential facilitation in the primary motor cortex. This evidence may be of help in characterizing our vowel imagery result in left pSTS-MTG, in that it may validate the idea that mechanisms springing from inferior frontal, speech-generating areas modulate activity in speech-perceiving ones, during covert articulation<sup>65</sup>. It is worth mentioning again that vowels arise from a perceptuo-motor model, with formant structure being determined by unique articulator configurations<sup>34</sup>. Such a model would contain both acoustic and motor information, and thus be represented equally well in superior temporal and inferior frontal areas. These findings are in agreement with previous results obtained with MVPA on functional brain imaging<sup>8</sup> as well as ECoG data<sup>7</sup> showing not only that the auditory cortex can encode vowel-specific information during perception<sup>7</sup>, but also, that it can represent articulated speech sounds<sup>15</sup>. Particularly, though, HG, the primary auditory cortex, did not show sensitivity to single phonemes<sup>8</sup>, as our findings confirm, despite the exquisitely acoustic nature of the task. Nonetheless, in our univariate results HG was significantly activated during vowel listening (see Fig. 1), although it represented pure tones in the multivariate results (see Fig. 3): an extrapolation coming from MVPA is that HG was simply not representing vowels in the listening task, despite being activated, as can be seen from Fig. 1. Of note, as explained in the Methods section, vowels are aggregates of formants above a fundamental frequency, which are perceived as a summation of the fundamental and the overtones, but also as discrete categories<sup>7</sup>. Such kind of complex stimuli with heightened (linguistic) salience might be computed outside the psychophysically low-level HG<sup>66,67</sup>, as our findings seem to suggest in comparison with simpler tones that are, indeed, represented there. Finally, findings from task-dependent decoding of speaker and vowel identity<sup>38</sup> reveal that the primary auditory cortex in the left hemisphere actually represents speaker information over vowel information, which seems reasonable when we consider the higher frequential variability of different speakers (across which is the fundamental frequency that changes), rather than the small changes in different vowels uttered by the same speaker, related to harmonic structure over the same fundamental<sup>34</sup>.

Moreover, in Tankus and colleagues<sup>15</sup>, while STG was further probed to assess its ability to discriminate between a complex system of five vowels, the authors also showed how this classically auditory hub of the cortex actually represents *articulated* speech sounds as well: nevertheless, while neurons in anterior locations such as the medial orbitofrontal cortex (MOF) and the rostral anterior cingulate cortex (rAC) responded to single or coupled vowels, in this study STG did not, in fact, reveal vowel specificity. In agreement with this study, we found STG activated by vowel production (Fig. 1), but crucially it did not classify single vowels (Fig. 3).

Moreover, pSTS-MTG, previously shown to be engaged in articulation imagery over *hearing* imagery<sup>32</sup>, shared sensitivity to mentally articulated vowels, as well as pure tones, in our data: this is supported by a study reporting conflict between vowel imagery and tone perception in the superior temporal cortex<sup>68</sup>. As in our findings, the region showing shared sensitivity to lower- and higher-level stimuli was significantly lateralized in the left, language-dominant hemisphere. Moreover, in our results, the patterns of imagined vowels that were represented in left pSTS-MTG could not be ascribed to any acoustic feedback due to the inner nature of the task itself. In this region, tone sensitivity would therefore sustain higher-level representations pertaining to a non-classical function associated to the location, as well as it did in the inferior frontal cortex.

In conclusion, using fMRI we were able to discriminate the seven vowels of the Italian language in listening, articulation imagery, and production tasks. Globally, these three functions revealed spatial dissociation within language-related brain regions, as well as collateral sensitivity to tone representations. Building on previous evidence, and on suggestions coming from theories postulating the integration of the perceptual and articulatory stages of speech, these findings provide a finer characterisation of the fronto-temporal language-related cortex. Notably, frontal brain regions classically associated to production can also represent acoustic features of both linguistic and non-linguistic stimuli; similarly, temporal regions that process low-level acoustic features (pure tones) retain sensitivity to covertly produced vowels. Importantly, in line with integration theories, not only sensitivity to speech listening exists in production-related regions and *vice versa*, but the nature of such interwoven organization is also built upon low-level perceptual features.



## Methods

**Participants.** Fifteen right-handed (Edinburgh Handedness Inventory<sup>69</sup>, mean laterality index  $0.79 \pm 0.17$ ) healthy, mother-tongue Italian monolingual speakers (9 F; mean age  $28.5 \pm 4.6$  years) participated in this study, after its approval by the Ethics Committee of the University of Pisa. All experimental procedures and methodologies were carried out in accordance with the relevant guidelines and regulations. Informed consent was gathered from all participants.

**Stimuli.** The seven vowels of the Italian language ([i] [e] [ɛ] [a] [ɔ] [o] [u]) were selected as experimental stimuli, along with seven pure tones (450, 840, 1370, 1850, 2150, 2500, 2900 Hz). Pure tones are physically simpler sounds with no harmonic structure, whereas vowels, despite being periodic waves as well, are endowed with acoustic resonances at specific frequency bandwidths, determined by the vocal tract modifying the source signal produced by the laryngeal mechanism. This structure yields a continuous emission of sound with a fundamental frequency (F0) and a number of overtones called *formants* (i.e., F1, F2, F3...), in a combination that is unique for each vowel. The seven vowels from the Italian phonemic inventory can be disambiguated by the two lower formants F1 and F2, with F0 being constant (Fig. 4)<sup>34</sup>.

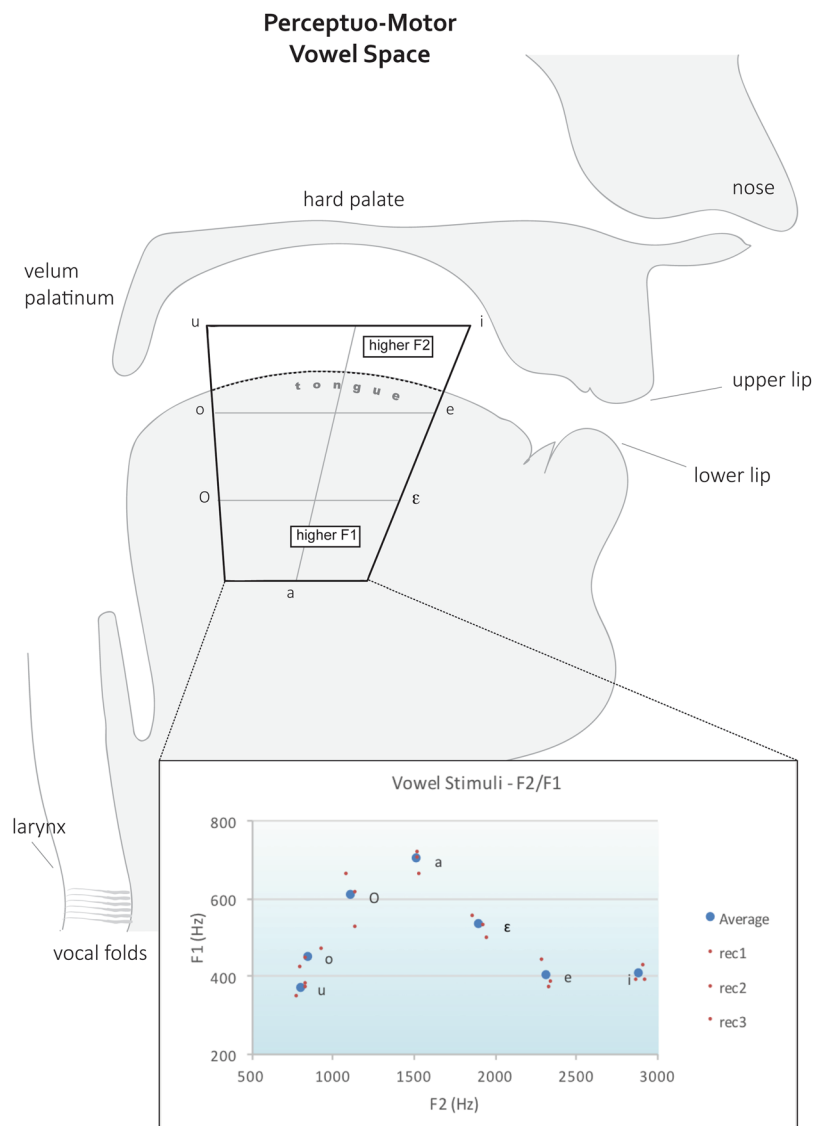
Three separate, 2 s natural voice recordings of each vowel (21 stimuli) were obtained from a female Italian speaker using Praat (©Paul Boersma and David Weenink, <http://www.fon.hum.uva.nl/praat/>) a 44100 Hz frequency sampling rate (F0:  $191 \pm 2.3$  Hz) and spectrograms were visually inspected for abnormalities. Pure tones were selected by dividing the minimum/maximum mean F1 range of the vowel set into seven, equally distanced bins; the resulting values were approximated to the closest Bark scale value and then converted back to Hertz, so that all tones would lie within the sensitive perceptual bands in a psychophysical model<sup>70</sup>. In Audacity (©Audacity Team, <http://audacity.sourceforge.net/>), seven tones were thus generated using the input-frequencies associated to the Bark value obtained through the aforementioned procedure. Table 4 reports mean F1 and F2 across recordings with the associated standard deviations, and the resulting approximated Bark value from which pure tones were generated.

**Experimental procedures.** A slow event-related paradigm was implemented with Presentation (©Neurobehavioral Systems, Inc., <http://www.neurobs.com/>) and comprised two perceptual tasks (tone perception and vowel listening), a vowel imagery task and a vowel production one. To increase the amplitude of individual BOLD responses during scan time, all perceived vowels and tones, as well as the execution of imagery and production, were made to last for 2 whole seconds, with the duration signalled by a green fixation cross that would turn black during resting time. All perceptual stimuli (tones or vowels) were thus administered in trials comprising 2 s stimulus presentation, then followed by 8 s rest. Imagery/production stimuli were administered in trials comprising 2 s stimulus presentation, 8 s maintenance, 2 s task execution and 8 s rest. For the imagery task, participants were instructed to perform mental articulation of a heard vowel with their own voice and simulating speech in their mind without ever moving; for the production task, they were instructed to speak naturally and at a normal volume, with rubber wedges and pillows secured so as to avoid head motion without constraining the chin and jaw. In the perceptual tasks (tone perception and vowel listening) subjects were instructed to lay still and listen attentively to the presented stimuli. Globally, functional scans were 47 m long, divided in 10 runs. Each of the three vowel recordings was presented twice, thus to obtain 42 trials randomized within and across tasks and subjects, with each sound, either vowel or tone, being equally represented.

BOLD activity was measured using GRE-EPI sequences on a GE Signa 3 Tesla scanner (TR/TE = 2500/30 ms; FA = 75°; 2 mm isovoxel; geometry:  $128 \times 128 \times 37$  axial slices). Brain anatomy was provided by a T1-weighted FSPGR sequence (TR/TE = 8.16/3.18 ms; FA = 12°; 1 mm isovoxel; geometry:  $256 \times 256 \times 170$  axial slices). Stimuli were presented using MR-compatible on-ear headphones (30 dB noise-attenuation, 40 Hz to 40 kHz frequency response).

**fMRI pre-processing.** The AFNI software package<sup>71</sup> was used to pre-process functional MRI data. First, all acquired slices were temporally aligned within each volume (*3dTshift*), corrected for head motion (*3dvolreg*), spatially smoothed (*3dmerge*) with a 4 mm FWHM Gaussian filter, and normalized by dividing, within each voxel, every time point by the mean of the time series. A multiple regression analysis was then performed on normalized runs (*3dDeconvolve*), to identify stimulus-related BOLD patterns. Movement parameters and signal trends were included in this procedure as regressors of no interest. Specifically, we used TENT functions for the estimation of BOLD activity (T-values), focusing on the third time point (7.5 seconds) after the acoustic stimulus onset or task execution (imagery or production). By doing this, we aimed at limiting sensory-motor and maintenance-related information, possibly biasing the signal preceding vowel imagery and production<sup>72–74</sup>. BOLD activity related to the acoustic stimulation in the imagery and production tasks was discarded. Afterwards, T1 images were pre-processed in FSL<sup>75</sup> and nonlinearly registered<sup>76</sup> to the Montreal Neurological Institute (MNI) standard space with a 2 mm isovoxel<sup>77</sup>; then, the obtained deformation field was used to warp functional maps for each task type.

**Language-sensitive regions.** Hereon, all analyses were performed within a pre-defined topic-based meta-analytic mask of language-sensitive regions. Specifically, the mask was selected from the Neurosynth database<sup>37</sup>, version 3, topic 21 out of 200, forward inference with a  $p < 0.01$  (FDR corrected)<sup>78</sup>. Keywords included terms related to language and phonological competence, among which “speech, auditory, sounds, processing, perception, voice, pitch, listening, production, vocal, tones, voices, phonetic, syllable, linguistic, speaker, discrimination, spectral, vowel, language”. The extension of the mask was 19093 voxels and comprised the bilateral posterior portion of the IFG/MFG, the left PrCG, the bilateral superior temporal cortex, running more posteriorly in the left hemisphere; the left ITG, SMG and angular gyrus (AG), and the bilateral IPS and MOG/IOG. The mask also included the bilateral caudate nuclei, and the medial portion of the SFG. All analyses, both univariate and multivariate, were performed within this mask.



**Figure 4.** Vowel acoustic and motor spaces. Here, an ideal representation of the perceptuo-motor vowel space can be appreciated through a sagittal view of the head and phonatory apparatus (top). The articulators are labelled and the relationship that lip and tongue positions entertain with the first and second formant (F1 and F2) can be seen from the trapezoid shape representing the Italian vowel system. Below, the real first- and second formant measurements from our experimental stimuli are plotted in the F1/F2 space, reproducing a projection of the pictured perceptuo-motor vowel space. In this chart, averages for each vowel are represented with blue dots, while measures from single recordings are represented with smaller, red dots (see legend: rec - recording).

Vowels	Mean F <sub>1</sub>	STD	Mean F <sub>2</sub>	STD	Bark Value	Tones
i	305	21.1	2170	25.7	5	450
e	303	35.9	1736	30.7	8	840
ε	400	27.1	1428	47.4	11	1370
a	525	28.9	1139	7.1	13	1850
ɔ	455	68.1	836	34.9	14	2150
o	338	23.4	637	71.6	15	2500
u	278	16.2	604	27.0	16	2900

**Table 4.** Mean F1 and F2 across recordings for the vowel stimuli. Here we show the Bark scale -to-Hertz conversion for the pure tones used in the tone perception task. Tone frequencies were obtained starting from mean vowel frequencies (left-to-right in the table). All vowels and tones lasted for 2 seconds.

**Univariate Analysis.** BOLD activity was used to perform one-sample 2-tailed t-test voxel-wise ( $p < 0.05$ , FDR corrected), thus comparing task activity versus rest in each modality.

**Multivariate Analysis.** To assess stimulus discrimination accuracy in each task, the T-value maps were then used in four searchlight-based classifiers<sup>79,80</sup> (rank accuracy; cosine similarity; 6 mm searchlight radius), one for each task (tone perception, vowel listening, imagery and production). A cross-validation *leave-one-stimulus-out* procedure was adopted to measure classification accuracy.

Each classifier was conceived to discriminate among seven classes of stimuli: the seven tones in the tone perception task and the seven vowels in the listening, imagery and production tasks. Accuracies emerging from the tone perception classifier would be used later on, to measure sensitivity to low-level features of acoustic stimuli within clusters defined by the vowel classifiers. Finally, the procedure generated a stimulus discrimination accuracy value for each task, in each voxel and subject. Group accuracies for tone perception, vowel listening, imagery and production were obtained by averaging all single-subject accuracy values, at each voxel.

To assess significance, group accuracies were tested against chance by a permutation test<sup>81–83</sup>, where all stimulus-class labels were shuffled in order to generate 1,000 permuted matrices to be used in a multi-class searchlight-based classifier identical to the one described above. The entire procedure generated a set of 1,000 single-subject null discrimination accuracies for each stimulus class, in each voxel, subject and task. Group null accuracies were obtained by averaging single-subject null accuracies in a distribution of 1,000 null accuracies for each voxel and stimulus class. Group accuracy maps were then corrected for multiple comparisons using AFNI: first, real smoothness in the data (resulting from pre-processing, anatomical and searchlight-related smoothing) was estimated ( $3dFWHMx$ ) from the null distribution defined above; later, cluster correction was performed using Monte Carlo simulations (the latest version of  $3dClustSim$ , 10,000 iterations<sup>84</sup>). This procedure preserved clusters larger than 207 voxels ( $p < 0.05$  at voxel level with  $\alpha < 0.05$  for the correction for multiple comparisons). All the procedures were developed in Matlab (©TheMathWorks, Inc., <http://www.mathworks.com/>), unless otherwise specified, through code developed in-house.

**Cross-task accuracies.** To assess whether vowel-sensitive clusters were specific to each task, we measured the averaged accuracies of each task within the masks defined by each of the others (e.g., accuracy of vowel listening within the vowel production mask;  $3dROIstats$ ). The same procedure was applied to the null distribution used in the aforementioned permutation test, thus to obtain cluster-based accuracies and their associated statistical significance (1,000 permutations, one-tailed rank test,  $p < 0.05$ ). Finally, significance level was adjusted using Bonferroni's correction for multiple comparisons (6 clusters by 3 tasks,  $p < 0.0028$  for  $p_{bonf} < 0.05$ ). The same procedure was employed to assess whether vowel-sensitive clusters represented tone-related information as well, thus to assess their specificity to non-linguistic versus linguistic stimuli; results were Bonferroni-corrected as well (6 clusters by 1 task,  $p < 0.0083$  for  $p_{bonf} < 0.05$ ).

**Data availability.** The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

## References

- Price, C. J. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage* **62**, 816–847 (2012).
- Vihman, M. M. Variable paths to early word production. *Journal of Phonetics* **21**, 61–82 (1993).
- Galantucci, B., Fowler, C. A. & Turvey, M. T. The motor theory of speech perception reviewed. *Psychonomic bulletin & review* **13**, 361–377 (2006).
- Schwartz, J. L., Basirat, A., Ménard, L. & Sato, M. The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics* **25**, 336–354 (2012).
- Bouchard, K. E., Mesgarani, N., Johnson, K. & Chang, E. F. Functional organization of human sensorimotor cortex for speech articulation. *Nature* **495**, 327–332 (2013).
- Obleser, J., Leaver, A. M., VanMeter, J. & Rauschecker, J. P. Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. *Frontiers in psychology* **1**, 232 (2010).
- Chang, E. F. *et al.* Categorical speech representation in human superior temporal gyrus. *Nature neuroscience* **13**, 1428–1432 (2010).
- Formisano, E., De Martino, F., Bonte, M. & Goebel, R. “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science* **322**, 970–973 (2008).
- Rampinini, A. C. & Ricciardi, E. In favor of the phonemic principle: a review of neurophysiological and neuroimaging explorations into the neural correlates of phonological competence. *Studi e Saggi Linguistici* **55**, 95–123 (2017).
- Evans, S. & Davis, M. H. Hierarchical organization of auditory and motor representations in speech perception: evidence from searchlight similarity analysis. *Cerebral cortex* **25**, 4772–4788 (2015).
- Zhang, Q. *et al.* Deciphering phonemes from syllables in blood oxygenation level-dependent signals in human superior temporal gyrus. *European Journal of Neuroscience* **43**, 773–781 (2016).
- Feng, G., Gan, Z., Wang, S., Wong, P. C. M. & Chandrasekaran, B. Task-General and Acoustic-Invariant Neural Representation of Speech Categories in the Human Brain. *Cerebral cortex*, 1–14 (2017).
- Skipper, J. I., Devlin, J. T. & Lametti, D. R. The hearing ear is always found close to the speaking tongue: review of the role of the motor system in speech perception. *Brain and language* **164**, 77–105 (2017).
- Grush, R. The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and brain sciences* **27**, 377–396 (2004).
- Tankus, A., Fried, I. & Shoham, S. Structured neuronal encoding and decoding of human speech features. *Nature communications* **3**, 1015 (2012).
- Correia, J. M., Jansma, B. M. & Bonte, M. Decoding Articulatory Features from fMRI Responses in Dorsal Speech Regions. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **35**, 15015–15025 (2015).
- Cheung, C., Hamiton, L. S., Johnson, K. & Chang, E. F. The auditory representation of speech sounds in human motor cortex. *eLife* **5**, e12577 (2016).
- Arsenault, J. S. & Buchsbaum, B. R. Distributed neural representations of phonological features during speech perception. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **35**, 634–642 (2015).

19. Lee, Y. S., Turkeltaub, P., Granger, R. & Raizada, R. D. S. Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *The Journal of Neuroscience: the official journal of the Society for Neuroscience* **32**, 3942–3948 (2012).
20. Markiewicz, C. J. & Bohland, J. W. Mapping the cortical representation of speech sounds in a syllable repetition task. *NeuroImage* **141**, 174–190 (2016).
21. Schomers, M. R. & Pulvermüller, F. Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. *Frontiers in human neuroscience* **10**, 435 (2016).
22. Josephs, K. A. *et al.* Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. *Brain: a journal of neurology* **129**, 1385–1398 (2006).
23. Hickok, G., Costanzo, M., Capasso, R. & Miceli, G. The role of Broca's area in speech perception: evidence from aphasia revisited. *Brain and Language* **119**, 214–220 (2011).
24. Basilakos, A., Rorden, C., Bonilha, L., Moser, D. & Fridriksson, J. Patterns of poststroke brain damage that predict speech production errors in apraxia of speech and aphasia dissociate. *Stroke* **46**, 1561–1566 (2015).
25. Ardila, A., Bernal, B. & Rosselli, M. Why Broca's Area Damage Does Not Result in Classical Broca's Aphasia. *Frontiers in human neuroscience* **10**, 249 (2016).
26. Amunts, K. *et al.* Broca's region: novel organizational principles and multiple receptor mapping. *PLoS biology* **8**, e1000489 (2010).
27. Anwender, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D. & Knosche, T. R. Connectivity-Based Parcellation of Broca's Area. *Cerebral cortex* **17**, 816–825 (2007).
28. Catani, M., Jones, D. K. & Ffytche, D. H. Perisylvian language networks of the human brain. *Annals of neurology* **57**, 8–16 (2005).
29. Hagmann, P. *et al.* Mapping the structural core of human cerebral cortex. *PLoS biology* **6**, e159 (2008).
30. Fullerton, B. C. & Pandya, D. N. Architectonic analysis of the auditory-related areas of the superior temporal region in human brain. *The Journal of comparative neurology* **504**, 470–498 (2007).
31. Amunts, K. & Zilles, K. Architecture and organizational principles of Broca's region. *Trends in cognitive sciences* **16**, 418–426 (2012).
32. Tian, X., Zarate, J. M. & Poeppel, D. Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex* **77**, 1–12 (2016).
33. Tian, X. & Poeppel, D. The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *Journal of cognitive neuroscience* **25**, 1020–1036 (2013).
34. Hardcastle, W. J., Laver, J. & Gibbon, F. E. *The handbook of phonetic sciences*. (John Wiley & Sons, 2010).
35. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289–300 (1995).
36. Genovese, C. R., Lazar, N. A. & Nichols, T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* **15**, 870–878 (2002).
37. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nature methods* **8**, 665–670 (2011).
38. Bonte, M., Hausfeld, L., Scharke, W., Valente, G. & Formisano, E. Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **34**, 4548–4557 (2014).
39. Scott, S. K. & Johnsrude, I. S. The neuroanatomical and functional organization of speech perception. *Trends in neurosciences* **26**, 100–107 (2003).
40. Troyer, T. W. & Doupe, A. J. An associational model of birdsong sensorimotor learning I. Efference copy and the learning of song syllables. *Journal of Neurophysiology* **84**, 1204–1223 (2000).
41. Rauschecker, J. P. & Scott, S. K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature neuroscience* **12**, 718–724 (2009).
42. Dronkers, N. F. A new brain region for coordinating speech articulation. *Nature* **384**, 159–161 (1996).
43. Long, M. A. *et al.* Functional Segregation of Cortical Regions Underlying Speech Timing and Articulation. *Neuron* **89**, 1187–1193 (2016).
44. Davis, C. *et al.* Speech and language functions that require a functioning Broca's area. *Brain and language* **105**, 50–58 (2008).
45. Bates, E. *et al.* Voxel-based lesion–symptom mapping. *Nature neuroscience* **6**, 448–450 (2003).
46. Mesgarani, N., Cheung, C., Johnson, K. & Chang, E. F. Phonetic feature encoding in human superior temporal gyrus. *Science* **343**, 1006–1010 (2014).
47. Embick, D., Marantz, A., Miyashita, Y., O'Neil, W. & Sakai, K. L. A syntactic specialization for Broca's area. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6150–6154 (2000).
48. Skipper, J. L., Nusbaum, H. C. & Small, S. L. Listening to talking faces: motor cortical activation during speech perception. *NeuroImage* **25**, 76–89 (2005).
49. Papoutsis, M. *et al.* From phonemes to articulatory codes: an fMRI study of the role of Broca's area in speech production. *Cerebral cortex* **19**, 2156–2165 (2009).
50. Goucha, T. & Friederici, A. D. The language skeleton after dissecting meaning: a functional segregation within Broca's Area. *NeuroImage* **114**, 294–302 (2015).
51. Demonet, J. F. *et al.* The anatomy of phonological and semantic processing in normal subjects. *Brain: a journal of neurology* **115**, 1753–1768 (1992).
52. Heim, S., Eickhoff, S. B. & Amunts, K. Specialisation in Broca's region for semantic, phonological, and syntactic fluency? *NeuroImage* **40**, 1362–1368 (2008).
53. Baddeley, A., Lewis, V. & Vallar, G. Exploring the articulatory loop. *The Quarterly journal of experimental psychology* **36**, 233–252 (1984).
54. Flinker, A. *et al.* Redefining the role of Broca's area in speech. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2871–2875 (2015).
55. Reiterer, S., Erb, M., Grodd, W. & Wildgruber, D. Cerebral processing of timbre and loudness: fMRI evidence for a contribution of Broca's area to basic auditory discrimination. *Brain Imaging and Behavior* **2**, 1–10 (2008).
56. Iacoboni, M. The role of premotor cortex in speech perception: evidence from fMRI and rTMS. *Journal of physiology, Paris* **102**, 31–34 (2008).
57. Shuster, L. I. & Lemieux, S. K. An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and language* **93**, 20–31 (2005).
58. Huang, J., Carr, T. H. & Cao, Y. Comparing cortical activations for silent and overt speech using event-related fMRI. *Human brain mapping* **15**, 39–53 (2002).
59. Hinke, R. M. *et al.* Functional magnetic resonance imaging of Broca's area during internal speech. *Neuroreport* **4**, 675–678 (1993).
60. Winhuisen, L. *et al.* Role of the contralateral inferior frontal gyrus in recovery of language function in poststroke aphasia: a combined repetitive transcranial magnetic stimulation and positron emission tomography study. *Stroke* **36**, 1759–1763 (2005).
61. Chakrabarti, S., Sandberg, H. M., Brumberg, J. S. & Krusienski, D. J. Progress in speech decoding from the electrocorticogram. *Biomedical Engineering Letters* **5**, 10–21 (2015).
62. Okada, K. & Hickok, G. Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and language* **98**, 112–117 (2006).



63. Buchsbaum, B. R., Hickok, G. & Humphries, C. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science* **25**, 663–678 (2001).
64. Murakami, T., Kell, C. A., Restle, J., Ugawa, Y. & Ziemann, U. Left dorsal speech stream components and their contribution to phonological processing. *The Journal of Neuroscience: the official journal of the Society for Neuroscience* **35**, 1411–1422 (2015).
65. Shergill, S. S. *et al.* Modulation of activity in temporal cortex during generation of inner speech. *Human brain mapping* **16**, 219–227 (2002).
66. Santoro, R. *et al.* Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS computational biology* **10**, e1003412 (2014).
67. Santoro, R. *et al.* Reconstructing the spectrotemporal modulations of real-life sounds from fMRI response patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 201617622 (2017).
68. Kauramäki, J. *et al.* Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. *The Journal of Neuroscience: the official journal of the Society for Neuroscience* **30**, 1314–1321 (2010).
69. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
70. Zwicker, E. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *The Journal of the Acoustical Society of America* **33**, 248–248 (1961).
71. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research* **29**, 162–173 (1996).
72. Leo, A. *et al.* A synergy-based hand control is encoded in human motor cortical areas. *eLife* **5**, e13420 (2016).
73. Handjaras, G. *et al.* How concepts are encoded in the human brain: a modality independent, category-based cortical organization of semantic knowledge. *NeuroImage* **135**, 232–242 (2016).
74. Connolly, A. C. *et al.* The representation of biological classes in the human brain. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **32**, 2608–2618 (2012).
75. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. Fsl. *NeuroImage* **62**, 782–790 (2012).
76. Andersson, J. L. R., Jenkinson, M. & Smith, S. Non-linear registration, aka Spatial normalisation FMRIB technical report TR07JA2. *FMRIB Analysis Group of the University of Oxford* **2** (2007).
77. Fonov, V. S., Evans, A. C., McKinstry, R. C., Almlí, C. R. & Collins, D. L. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage* **47**, S102 (2009).
78. Poldrack, R. A. *et al.* Discovering relations between mind, brain, and mental disorders using topic mapping. *PLoS computational biology* **8**, e1002707 (2012).
79. Mitchell, T. M. *et al.* Learning to decode cognitive states from brain images. *Machine learning* **57**, 145–175 (2004).
80. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3863–3868 (2006).
81. Winkler, A. M., Ridgway, G. R., Douaud, G., Nichols, T. E. & Smith, S. M. Faster permutation inference in brain imaging. *NeuroImage* **141**, 502–516 (2016).
82. Pereira, F., Mitchell, T. & Botvinick, M. Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage* **45**, S199–S209 (2009).
83. Nichols, T. E. & Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping* **15**, 1–25 (2002).
84. Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C. & Taylor, P. A. FMRI Clustering and False-Positive Rates. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 3370–3371 (2017).

## Author Contributions

A.R. conceived study and wrote manuscript; A.R., G.H., A.L. and L.C. collected, analysed and interpreted data; G.H., A.L. and L.C. constructively reviewed manuscript; E.R., G.M. and P.P. supervised study process, reviewed and approved final version of manuscript.

## Additional Information

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017