

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/290625515>

Mental imagery of speech implicates two mechanisms of perceptual reactivation

Article *in* Cortex · January 2016

DOI: 10.1016/j.cortex.2016.01.002

CITATIONS

2

READS

279

3 authors, including:



Xing Tian

New York University Shanghai

20 PUBLICATIONS 322 CITATIONS

SEE PROFILE



David Poeppel

New York University

258 PUBLICATIONS 13,712 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Seeing and Hearing a Word: integration within and across senses [View project](#)

All content following this page was uploaded by [Xing Tian](#) on 26 January 2016.

The user has requested enhancement of the downloaded file. All in-text references [underlined in blue](#) are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.

Accepted Manuscript

Mental imagery of speech implicates two mechanisms of perceptual reactivation

Xing Tian, Ph.D, Jean Mary Zarate, David Poeppel

PII: S0010-9452(16)00003-4

DOI: [10.1016/j.cortex.2016.01.002](https://doi.org/10.1016/j.cortex.2016.01.002)

Reference: CORTEX 1655

To appear in: *Cortex*

Received Date: 3 February 2015

Revised Date: 26 June 2015

Accepted Date: 5 January 2016

Please cite this article as: Tian X, Zarate JM, Poeppel D, Mental imagery of speech implicates two mechanisms of perceptual reactivation, *CORTEX* (2016), doi: [10.1016/j.cortex.2016.01.002](https://doi.org/10.1016/j.cortex.2016.01.002).

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Mental imagery of speech implicates two mechanisms of
perceptual reactivation

Xing Tian^{1,2,3*}, Jean Mary Zarate³, David Poeppel^{3,4}

1. New York University Shanghai
2. NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai
3. Department of Psychology, New York University
4. Department of Neuroscience, Max Planck Institute for Empirical Aesthetics
(MPIEA)

*Corresponding author: Xing Tian, Ph.D.

New York University Shanghai

1555 Century Avenue, Room 1138

Pudong, Shanghai, China 200122

Email: xing.tian@nyu.edu

Phone: +86-21 20595201

Mental imagery of speech implicates two mechanisms of perceptual reactivation

Xing Tian^{1,2,3*}, [Jean Mary Zarate](#)³, David Poeppel^{3,4}

1. New York University Shanghai
2. NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai
3. Department of Psychology, New York University
4. Department of Neuroscience, Max Planck Institute for Empirical Aesthetics (MPIEA)

*Corresponding author: Xing Tian, Ph.D.

New York University Shanghai

1555 Century Avenue, Room 1138

Pudong, Shanghai, China 200122

Email: xing.tian@nyu.edu

Phone: +86-21 20595201

Abstract

Sensory cortices can be activated without any external stimuli. Yet, it is still unclear how this perceptual reactivation occurs and which neural structures mediate this reconstruction process. In this study, we employed fMRI with mental imagery paradigms to investigate the neural networks involved in perceptual reactivation. Subjects performed two speech imagery tasks: articulation imagery (AI) and hearing imagery (HI). We found that AI induced greater activity in frontal-parietal sensorimotor systems, including sensorimotor cortex, subcentral (BA 43), middle frontal cortex (BA 46) and parietal operculum (PO), whereas HI showed stronger activation in regions that have been implicated in memory retrieval: middle frontal (BA 8), inferior parietal cortex and intraparietal sulcus. Moreover, posterior superior temporal sulcus (pSTS) and anterior superior temporal gyrus (aSTG) was activated more in AI compared with HI, suggesting that covert motor processes induced stronger perceptual reactivation in the auditory cortices. These results suggest that motor-to-perceptual transformation and memory retrieval act as two complementary mechanisms to internally reconstruct corresponding perceptual outcomes. These two mechanisms can serve as a neurocomputational foundation for predicting perceptual changes, either via a previously learned relationship between actions and their perceptual consequences or via stored perceptual experiences of stimulus and episodic or contextual regularity.

Keywords: prediction; mental simulation; sensorimotor integration; internal forward model/efference copy/corollary discharge; memory retrieval

Short title: Two mechanisms for perceptual reactivation

Highlights:

- Sensory cortices can be activated without external stimulation.
- Two types of speech imagery were used to induce perceptual reactivation.
- Motor-sensory areas were more strongly activated in imagined speaking relative to imagined hearing.
- Frontal-parietal memory regions were involved in imagined hearing compared to imagined speaking.
- Motor-sensory transformation and memory are mechanisms for perceptual reactivation.

Introduction

Sensory cortices can be activated without any external stimulation (e.g. [Ji & Wilson, 2006](#); [Wheeler, Petersen, & Buckner, 2000](#)). That is, perceptual neural representations can be reconstructed without perceptual processing (referred to as perceptual reactivation). Mental imagery, defined as an internally generated quasi-perceptual experience, is one such example (e.g. [Kosslyn et al., 1999](#); [Kraemer, Macrae, Green, & Kelley, 2005](#)). The ability to form mental images has been hypothesized as a vehicle for generating and representing thoughts. This argument can be found as early as Plato's *Theaetetus* [427-347 BC] (1987) and Aristotle's *De Anima* [384-322 BC] (1986). In the age of enlightenment, mental imagery was considered analogous to perception by philosophers such as Descartes (1642/1984), Hobbes (1651/1968), Berkeley (1734/1965a, 1734/1965b) and Hume (1969). Early experimental psychologists such as Wundt (1913) and James (1890) proposed that ideas were represented as mental images in both visual and auditory domains. Modern research in mental imagery has yielded insight on how thought is represented in cognitive systems ([Kosslyn, 1994](#); [Kosslyn, Ganis, & Thompson, 2001](#); [Paivio, 1971, 1986](#); [Pylyshyn, 1981, 2003](#)).

Recently, an additional computational role of mental imagery has been proposed: a mechanism to plan possible future contingencies. That is, mental imagery has been modeled as a process in which perceptual consequences can be predicted to gain advantages in various aspects of perception, memory, decision making and motor control ([Albright, 2012](#); [Moulton & Kosslyn, 2009](#); [Schacter et al., 2012](#); [Tian & Poeppel, 2012](#)). The reactivation of perceptual neural representations without any external stimulation is the key mechanism mediating this predictive ability ([Moulton & Kosslyn, 2009](#)). Internally induced neural representations, which are highly similar to the ones established in corresponding perceptual processing, have been observed in

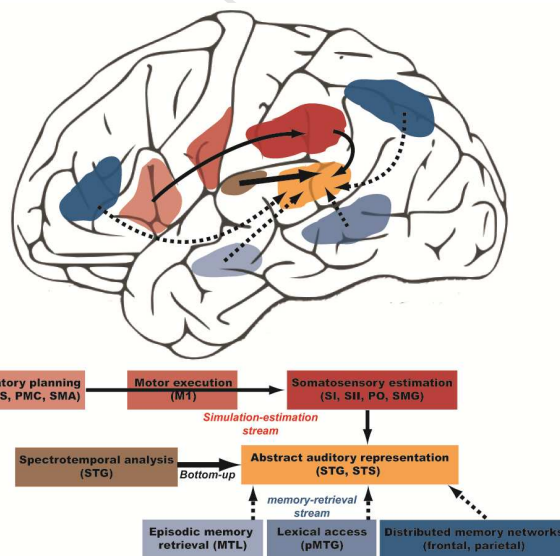
modality-specific areas, such as in visual (e.g. [Kosslyn et al., 1999](#); [O'Craven & Kanwisher, 2000](#)), auditory (e.g. [Kraemer et al., 2005](#); [Shergill et al., 2001](#); [Zatorre, Halpern, Perry, Meyer, & Evans, 1996](#)), somatosensory (e.g. [Yoo, Freeman, McCarthy III, & Jolesz, 2003](#); [Zhang, Weisser, Stilla, Prather, & Sathian, 2004](#)) and olfactory (e.g. [Bensafi et al., 2003](#); [Djordjevic, Zatorre, Petrides, Boyle, & Jones-Gotman, 2005](#)) domains.

It is not clear how these neural representations are reconstructed. Preliminary evidence from an MEG study ([Tian & Poeppel, 2013](#)) suggests that imagining speaking (articulation imagery, AI) and imagining hearing (hearing imagery, HI) differentially modulated neural responses to subsequent auditory stimuli. These distinct modulation effects by different types of imagery suggest that similar auditory neural representations may be internally formed via different neural pathways. A dual stream prediction model (DSPM, Fig. 1) was proposed in which two distinct processes in parallel neural pathways can internally induce the corresponding perceptual neural representation ([Tian & Poeppel, 2012, 2013](#)).

In the *simulation-estimation prediction stream* (Fig. 1), the perceptual consequences of actions are predicted by simulating the movement trajectory, followed by estimating the perceptual changes that would be associated with this movement. AI has been hypothesized to implement the motor-to-sensory transformation for simulation-estimation mechanism ([Tian & Poeppel, 2013](#)). Specifically, during AI, a motor simulation process similar to speech motor preparation is carried out, but without execution and output ([Palmer et al., 2001](#); [Tian & Poeppel, 2010, 2012](#)). Therefore, neural networks that mediate motor simulation should be similar to the ones implicated in motor preparation, including supplementary motor area (SMA), inferior frontal gyrus (IFG), premotor and insula ([Bohland & Guenther, 2006](#); [Palmer et al., 2001](#); [Shuster & Lemieux, 2005](#)). After motor simulation, a copy of the planned motor commands –

known as the efference copy (Von Holst & Mittelstaedt, 1950/1973; for a review see [Wolpert & Ghahramani, 2000](#)) – is sent to the somatosensory areas and is used in a forward model to estimate the somatosensory consequences ([Blakemore & Decety, 2001](#)). This somatosensory estimation is hypothesized to be governed by the networks underlying somatosensory perception ([Blakemore, Wolpert, & Frith, 1998](#); [Tian & Poeppel, 2010, 2012](#)), including primary and secondary somatosensory regions, parietal operculum and the supramarginal gyrus (SMG). Moreover in the context of speech, we hypothesize that auditory consequences are predicted on the basis of somatosensory estimation, and this auditory estimation will recruit neural structures in temporal auditory cortices ([Tian & Poeppel, 2010, 2012, 2013, 2015](#)).

*****insert Figure 1 here *****



In the *memory-retrieval prediction stream* (Fig. 1), the internally induced neural representations are the result of memory retrieval processes -- reconstructing stored perceptual information in modality-specific cortices (Kosslyn, 1994, 2005; [Wheeler et al., 2000](#)). In particular, the retrieved object properties from long-term memory reactivate the sensory cortices

that originally processed the object features (Kosslyn, 1994). In this experiment, we employed HI to probe this memory-retrieval stream. Auditory representations can be retrieved from various memory sources such as episodic memory, which presumably relies on hippocampal structures (Carr, Jadhav, & Frank, 2011; Eichenbaum, Sauvage, Fortin, Komorowski, & Lipton, 2012) with a possible buffer site in parietal cortex (Vilberg & Rugg, 2008; Wagner, Shannon, Kahn, & Buckner, 2005). Auditory representations can also be transformed from lexical and semantic information stored in semantic networks, including frontal (e.g. dorsomedial prefrontal cortex, inferior frontal gyrus, ventromedial prefrontal cortex), parietal (e.g. posterior inferior parietal lobe) and temporal (e.g. middle temporal gyrus) regions (Binder, Desai, Graves, & Conant, 2009; Lau, Phillips, & Poeppel, 2008; Price, 2012). Regardless of the divergent functional roles (episodic or semantic networks), frontal and parietal regions are reliably activated during memory retrieval processes. Therefore, neural activation in a frontal-parietal distributed network – the proposed memory-retrieval prediction stream – should be observed during HI.

This study uses fMRI to investigate three neuroanatomical/functional hypotheses that are generated from the DSPM. First, if the perfect simulation-estimation and memory-retrieval tasks were carried out, two distinct processing streams would be revealed separately. However, because speech imagery could involve both production and perception, we predict that both types of imagery will activate the simulation-estimation stream for simulating speech motor action (Tian & Poeppel, 2013). More importantly, we hypothesize that each type of imagery will recruit each prediction stream to a different extent. Specifically, we predict that AI will induce stronger activation in the simulation-estimation prediction stream, including SMA, IFG, premotor, insula for motor simulation, as well as primary and/or secondary somatosensory areas parietal operculum and SMG for subsequent estimation of somatosensory consequences. On the other

hand, we predict that HI will have more activation in the memory-retrieval prediction stream, which is comprised of frontal, superior and inferior parietal cortices that are associated with memory retrieval ([Binder et al., 2009](#); [Lau et al., 2008](#); [Price, 2012](#); [Vilberg & Rugg, 2008](#); [Wagner et al., 2005](#)).

Second, we suggest that a more precise, detailed auditory prediction can be induced through simulation-estimation mechanisms, comparing to that obtained via memory-retrieval route ([Hickok, 2012](#); [Oppenheim & Dell, 2010](#); [Tian & Poeppel, 2012, 2013](#)). We propose that there is a *one-to-one mapping* between motor simulation and perceptual estimation via a bridge of somatosensory estimation in the simulation-estimation stream. Such a deterministic prediction mechanism, contrasted with the memory-retrieval prediction stream's probabilistic prediction mechanism (narrowing down the target features in distributions of stored memory), presumably suffers less interference and lateral inhibition from similar features and yields a stronger and robust representation ([Tian & Poeppel, 2012, 2013](#)). Based on this hypothesis of enriched auditory representations via simulation and estimation processes, we predict that auditory cortices will be more strongly activated in AI than in HI.

Finally, we hypothesize that the neural networks governing simulation within the simulation-estimation stream overlap with cortical regions underlying motor preparation during speech production ([Tian & Poeppel, 2012](#)). That is, the initial motor processes are the same during articulation (A) and AI until the processes diverge, specifically until the motor signals are not executed during imagery. Therefore, we predict that enhanced activity in SMA, IFG, premotor areas and insula, which has been observed during preparation of overt speech production ([Brendel et al., 2010](#); [Riecker et al., 2005](#)), will be observed in both AI and A. The observation of overlapping neural networks will provide evidence towards potentially shared neural

mechanisms between overt and covert speech production, and furthermore suggests that mental imagery of speech is a valid paradigm to research these shared motor processes.

Methods

Participants. Eighteen volunteers gave informed consent and participated in the experiment (10 males, mean age 28.2 years, range 20 – 44 years). All participants were right-handed, with no history of neurological disorders. The experimental protocol was approved by the New York University Institutional Review Board (IRB).

Materials. Two 600-ms duration consonant-vowel syllables (/ba/, /ki/) were used as auditory stimuli (female voice; sampling rate of 48 kHz). All sounds were normalized to 70 dB SPL and delivered through MR-compatible headphones (MR confon Silenta, MR confon GmbH, Magdeburg, Germany). Four images were used as visual cues to indicate different trial types. Each image was presented foveally, against a black background, and subtended less than 10 degrees visual angle. A label - either '/ba/' or '/ki/' - was superimposed on the center of each picture (< 4 degrees visual angle) to indicate the syllable that participants would produce in the following tasks.

Experimental Procedure. We employed a similar experimental design as Tian and Poeppel (2013) (see Fig. S1). The experiment was comprised of four conditions: articulation (A), hearing (H), articulation imagery (AI), and hearing imagery (HI). In A, participants were asked to overtly generate the cued syllable (gently, to minimize head movement). In AI, participants were required to imagine saying a syllable without any overt movement of the articulators. In H, participants passively listened to one of the syllables. In HI, participants were asked to imagine hearing the cued syllable.

The timing of trials was consistent across conditions (Fig. S1). First, a visual cue appeared in the center of the screen at the beginning of each trial and stayed on for 1000 ms. During the following 2400 ms, participants actively formed a syllable in three of the task conditions (A, AI, and HI) or passively perceived an auditory syllable in H, in which a syllable was presented 1200 ms after the offset of visual cue, followed by a 600 ms interval. Notice that the 2400 ms period was the total duration that participants were allowed to finish the tasks (indicated by the curly bracket, Fig. S1). The actual time of performing task was much shorter, presumably comparable to the syllable duration. Finally, participants were presented with one of the syllables that always followed the task phase. The inter-trial interval was randomly chosen from 4440 to 6660 ms (2 to 3 TRs, see MRI scanning for details), temporally jittered by 46.25-ms increments (length of TR divided by 48, the number of task trials in a run). Twelve trials for each of the four tasks were presented in each run. Six resting trials (length: 9550 ms), which were visually cued with the word 'rest' were also included in each run. In total, the experiment included five runs with 54 trials each, encompassing all four tasks and the rest condition, which were pseudo-randomly presented in each run.

The goal of our earlier MEG study ([Tian & Poeppel, 2013](#)), on which this current study builds, was to assess cross-modal repetition adaptation. In contrast, the goal of this study is to assess the neural networks mediating internal perceptual reactivation by testing the main effects among different tasks, which are independent of the adaptation effects. Because the different syllables were used equally often, we only compared between overt and covert tasks, rather than between different syllables.

Each participant received training for 15-20 minutes before the experiment, and they focused on the timing as well as vividness of imagery. First, only the H trials were presented to introduce

the relative timing among the visual cue, the first auditory stimulus (the same period for active tasks in other conditions), and the subsequent auditory stimuli. After participants were familiar with the timing, they were instructed to use similar timing for the other conditions. This was to prevent any overlap between the internally generated neural responses during tasks and the subsequent responses to the external auditory stimuli. Next, they practiced on A trials while the experimenter observed the overt articulation and provided feedback if needed; participants executed the task with similar timing and without overlaps between their voice and subsequent auditory stimuli, before they moved onto the imagery conditions. For the AI condition, they were told to imagine speaking the syllables “in their mind” without moving any articulators or producing any sounds. They should feel the movement of specific articulators that would associate with actual pronunciation. For the HI condition, they were asked to recreate the female voice from the H condition in their minds, but minimize any feeling of movement in their articulators. If needed, the recorded female voice was presented again to form a better memory. We tried to selectively elicit the motor-induced auditory representation in imagined speaking, while we aimed to target auditory memory retrieval in imagined hearing. Participants were asked to generate a movement intention and kinesthetic feeling of articulation in the AI condition; in the HI condition, such motor-related imagery activity was strongly discouraged. After verbal confirmation of successful distinction between these types of imagery formation, they further practiced on the AI and HI tasks to reinforce the vividness of imagery and the timing requirement in the trials. Lastly, they trained on a practice block in which all four conditions were presented. Timing of the A condition was monitored by the experimenter and verbal confirmation of distinction between imagery tasks was obtained for each participant before proceeding onto the main experiment.

fMRI data acquisition. Scanning was performed with a 3T Siemens Allegra MRI system using a single-channel, whole-head coil. Functional data were acquired using a gradient-echo, echo-planar pulse sequence (TR=2220 ms; TE=30 ms; 38 slices oriented about 30 degrees rotated counter-clockwise from AC-PC line, which was adjusted individually to maximize coverage; (3x3x3) mm³ voxel size, 0.6 mm interslice gap; 244 volume acquisitions in each of five runs). High resolution T1-weighted (MP-RAGE) images were collected from each participant for anatomical visualization.

fMRI pre-processing. Data were analyzed using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/>). The first two volumes of a run (dummy images) were discarded from all analyses. All functional volumes were spatially realigned after motion correction. Structural images were coregistered to the functional images and spatially normalized to a T1-ICBM152 template provided in SPM8. The resulting normalization parameters were applied to the functional images, followed by spatial smoothing with an 8-mm full-width, half maximum isotropic Gaussian kernel.

Statistical analysis. Voxel-wise statistical parametric maps of brain activation were generated by estimating the parameters of a general linear model (GLM). For each of the four conditions (A, AI, H, HI) in each participant, neural activity was modeled as boxcar events spanning the entire 4 s trial period (from onset of visual cue to the offset of auditory stimuli), convolved with a canonical hemodynamic response function, and entered as regressors into a fixed effect GLM. The time series were high-pass filtered with a cut-off at 128 s.

For each comparison of interest in each participant, a contrast of parameter estimates (β weights) was calculated in a voxel-wise manner to produce a contrast image. Two groups of contrasts were defined. The first group of contrasts was the main effects of AI, HI and A: (1) A > H; (2) AI > H; (3) HI > H. Because this study was designed to assess the neural networks that

mediate AI, HI, and A, H was used as a baseline to account for neural responses to the auditory stimuli that cannot be temporally separated from the responses of interest (the tasks) in the experimental design. The second group of contrasts contained direct comparisons between imagery tasks: (4) AI > HI; (5) HI > AI. These direct contrasts revealed the possible differential involvement of neural pathways in different types of speech imagery tasks.

The parameter estimates from these first-level analyses were then entered into a random (between-subject) effect analysis, and linear contrasts were used to identify responsive regions. Thresholded t-maps were obtained for all contrasts, with a cluster threshold of 25 contiguous voxels whose test statistic exceeded an uncorrected p value of 0.001 ([Lieberman & Cunningham, 2009](#)). Because the effect size of mental imagery in each voxel is weak compared to overt hearing and production, we chose this approach to balance the type I and II errors. The AlphaSim Monte Carlo simulation in the original method paper ([Lieberman & Cunningham, 2009](#)) shows that using the statistical threshold of $P < 0.005$ and cluster size of 10 voxels can achieve a desirable balance between type I and II errors, while using a 20 voxel extent threshold produces an actual false discovery rate (FDR) of 0.05. For this study, an AlphaSim Monte Carlo simulation with our particular scanning and analysis parameters – a smoothing kernel of 8 mm and voxel resolution of 3mm – combined with a more conservative criterion with the magnitude statistical threshold of 0.001 and cluster threshold of 25 voxels yielded an FDR of 0.022. To examine regions that showed significant common neural responses to AI and HI as well as to A and AI, conjunction analyses were performed with the contrasts of interest $[AI > H] \cap [HI > H]$ and $[A > H] \cap [AI > H]$ ([Friston, Penny, & Glaser, 2005](#); [Nichols, Brett, Andersson, Wager, & Poline, 2005](#)).

For visualization purposes, thresholded maps were superimposed on an average, spatially normalized anatomical image obtained from the 18 participants. The locations of neural activity were first classified using the Automated Anatomical Labeling (AAL) map ([Tzourio-Mazoyer et al., 2002](#)), and then were further refined with: 1) neuroanatomical atlases ([Duvernoy, 1991](#); [Schmahmann et al., 1999](#)); 2) probabilistic maps or profiles for primary auditory cortex ([Penhune, Zatorre, MacDonald, & Evans, 1996](#)), planum temporale ([Westbury, Zatorre, & Evans, 1999](#)), inferior frontal gyrus pars opercularis ([Tomaiuolo et al., 1999](#)), and mouth region of primary motor cortex ([Fox et al., 2001](#)); and 3) locations defined by previous reports or reviews on the medial frontal and cingulate areas ([Picard & Strick, 1996, 2001](#)) and subdivisions of the premotor cortex ([Chen, Penhune, & Zatorre, 2008](#)).

Results

Main effects of tasks A, AI and HI. Speech production networks were observed during A, which included bilateral anterior cingulate cortex (ACC), pre-SMA/SMA complex, sensorimotor cortex, middle frontal cortex (BA 46) and right posterior cingulate cortex. Cerebellum and subcortical regions, including thalamus and basal ganglia were also activated (see Table S1 for a complete activation list). Significant activations, as well as in all following analyses, surpassed a threshold of $t > 3.65$ ($P < 0.001$ uncorrected) with an extent of at least 25 voxels which is equivalent to FDR smaller than 0.05 ([Lieberman & Cunningham, 2009](#)).

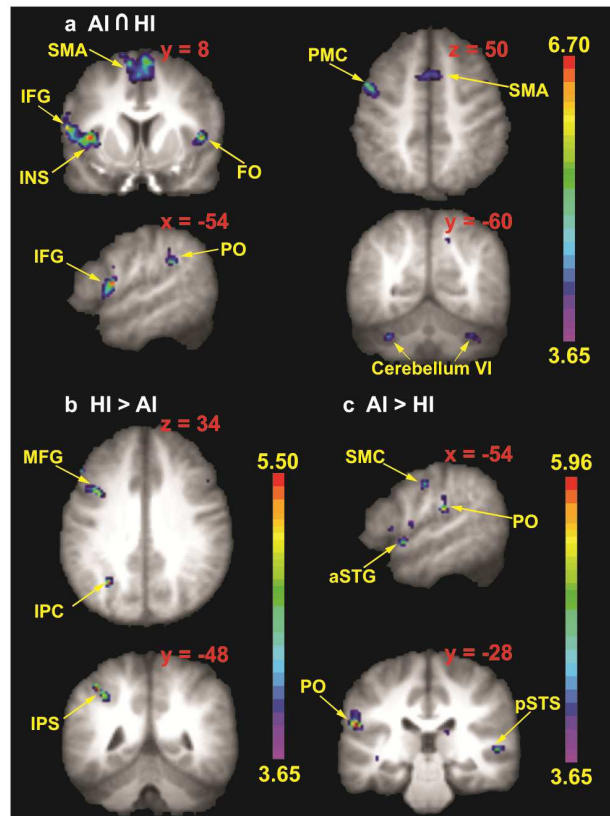
The neural networks that mediated AI were comprised of frontal and motor-related regions, including bilateral pre-SMA, inferior frontal pars opercularis (BA 44), frontal operculum, anterior insula, mid premotor cortex (BA 6), middle frontal cortex (BA 46); this activity extended to left primary motor cortex (near the mouth region). Parietal activation was observed

in the left parietal operculum. Moreover, activity in bilateral cerebellum VI (declive) and globus pallidus was also observed (see Table S2 for a complete list of activation). Activity in auditory cortices was not observed in the contrast of AI-H (Fig. S2).

Similar neural networks were observed during the HI task, including bilateral pre-SMA/SMA, inferior frontal pars opercularis (BA 44), frontal operculum, mid premotor cortex (BA 6) in the frontal lobe, and left parietal operculum in the parietal lobe. Bilateral cerebellum VI (declive) was also engaged (see Table S3 for a complete list of activation). Activity in auditory cortices was not observed in the contrast of HI-H (Fig. S2).

Shared regions for AI and HI. The conjunction analysis between AI and HI revealed the shared neural networks for these imagery tasks (Fig. 2a): bilateral pre-SMA/SMA, inferior frontal pars opercularis (BA 44), and left anterior insula, mid premotor cortex (BA 6) extending to primary motor cortex (near mouth region), and bilateral cerebellum VI (declive). Moreover, activity from both tasks overlapped in left parietal operculum, a somatosensory-related area (see Table 1 for a complete list of activation peaks).

*****insert Figure 2 here*****



*****insert Table 1 here*****

Stronger neural activity during HI. In the direct contrast between HI and AI, stronger activity was observed in HI compared with AI in left middle frontal (BA 8) and in left inferior parietal cortex and intraparietal sulcus (Fig. 2b, also see Table 2 for a complete list). Activity in auditory cortices was not observed in the contrast of HI-AI (Fig. S2).

*****insert Table 2 here*****

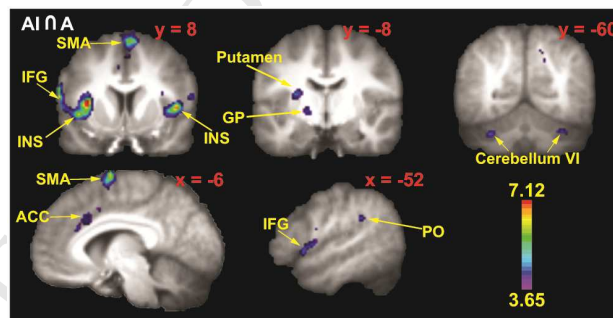
Stronger activation during AI. The direct comparison between AI and HI revealed stronger activation during AI over frontal and parietal areas, including bilateral sensorimotor cortex, left subcentral gyrus (Rolandic operculum, encompassing vocalization areas of primary motor and somatosensory cortex) and middle frontal cortex (BA 46), as well as left parietal operculum (Fig.

2c, also see Table 3). The same direct comparison between AI and HI also revealed stronger activation during AI over temporal cortices, including left anterior superior temporal gyrus (aSTG) and right posterior superior temporal sulcus (pSTS) (Fig. 2c, also see Table 3).

*****insert Table 3 here*****

Common neural networks that mediate A and AI. The conjunction analysis between A and AI revealed the shared networks between overt and covert speech production tasks (Fig. 3). These overlapping areas included bilateral ACC, inferior frontal pars opercularis (BA 44), pre-SMA/SMA, left mid-dorsal insula, right frontal operculum and left parietal operculum, as well as bilateral cerebellum VI (declive), globus pallidus and left putamen (Table 4).

*****insert Figure 3 here*****



*****insert Table 4 here*****

Discussion

We investigated the neural networks that mediate perceptual reactivation using fMRI with speech imagery paradigms. Whereas the neural networks that mediate articulation imagery and hearing imagery largely overlapped in frontal-parietal motor-sensory areas, different subsets of

frontal and parietal regions were involved in each task. This differential involvement of neural networks suggests two possible mechanisms for reactivating perceptual neural representation.

Frontal-parietal neural networks were observed during both AI and HI. The frontal overlapped areas included bilateral pre-SMA/SMA, inferior frontal pars opercularis (BA 44), and left anterior insula, mid premotor cortex (BA 6) (see Fig. 2a). Interestingly, most of the observed overlapped networks between AI and HI (BA 44, pre-SMA/SMA, insula) were also found in the conjunction analysis between AI and actual articulation (A) (see Fig. 3). These frontal/insular regions (SMA, IFG, premotor and insula) have been implicated in motor preparation during overt speech production ([Bohland & Guenther, 2006](#); [Palmer et al., 2001](#); [Shuster & Lemieux, 2005](#)). Therefore, perceptual reactivation processes during AI and HI may recruit these regions to simulate motor preparation without motor execution ([Palmer et al., 2001](#); [Tian & Poeppel, 2010, 2012](#)); motor simulation may then induce activity in sensory cortices. In fact, aside from the observed shared frontal activity between AI and HI, we also observed overlap activation in parietal operculum (see Fig. 2a), an area that relates to somatosensory perception (e.g. [Blakemore et al., 1998](#)). These results are consistent with the internal forward model theory, which hypothesizes that a copy of the planned motor commands – known as the efference copy (Von Holst & Mittelstaedt, 1950/1973; for a review see [Wolpert & Ghahramani, 2000](#)) – is sent to somatosensory areas and is used to estimate the somatosensory consequences of an action ([Blakemore & Decety, 2001](#)). Therefore, the observed activation in frontal-parietal sensorimotor regions during both AI and HI suggests that the motor-sensory interaction via the simulation-estimation process is a potential top-down mechanism to reactivate sensory cortices without external stimuli or output.

Multiple functions such as auditory working memory have been associated with the SMG (e.g. [Paulesu, Frith, & Frackowiak, 1993](#)). In this study, the parietal operculum, an area close to SMG was observed in both articulation and articulation imagery conditions. In the articulation condition, participants were required to say only one syllable after visual cue, so the working memory demand was minimal. As such, the observed parietal opercular activity may not have been elicited by working memory, but rather perception of somatosensory feedback. Together with the conjunction results, the observed parietal opercular activation in articulation imagery may be involved in the estimation of somatosensory consequences in a process similar to that seen during somatosensory perception.

The direct contrast between AI and HI reveals that frontal-parietal sensorimotor regions were activated stronger during AI (see Fig. 2c). The observed greater activation in frontal and motor areas during AI, including bilateral sensorimotor cortex, left subcentral gyrus (Rolandic operculum) and middle frontal cortex (BA 46) is similar to activation patterns indicative of articulation preparation ([Brendel et al., 2010](#); [Price, 2012](#); [Riecker et al., 2005](#)), which suggests that articulation imagery relies more on internally simulating articulatory preparation. Additionally, greater parietal opercular activity during AI may represent stronger somatosensory estimation during motor simulation.

On the other hand, the reverse comparison between HI and AI revealed that activity increased in left middle frontal, inferior parietal cortex and intraparietal sulcus during HI, which may form a subset of proposed distributed memory systems. For example, parietal cortex has been hypothesized as a buffer site for episodic memory ([Vilberg & Rugg, 2008](#); [Wagner et al., 2005](#)). Lexical and semantic information may be stored in semantic networks, including frontal (e.g. dorsomedial prefrontal cortex, inferior frontal gyrus, ventromedial prefrontal cortex) and

parietal (e.g. posterior inferior parietal lobe) regions ([Binder et al., 2009](#); [Price, 2012](#)). The greater activity seen in middle frontal, inferior parietal cortex and intraparietal sulcus during HI may reflect memory retrieval during perceptual reactivation. That is, HI may also rely on two complementary processes: a memory retrieval operation and motor simulation. This combined contribution from motor and memory systems in reactivated perceptual representation may be due to the nature of the HI task: both speech perception and production are related to this particular process of perceptual reactivation, and hence need memory retrieval of information related to speech perception and motor processes to simulate speech production. Therefore, this dissociation of neural pathways between AI and HI tasks implies that (1) two functional pathways exist for perceptual reactivation: one underlies motor-to-perceptual transformation and another mediates memory retrieval; and (2) these two pathways are differentially recruited during perceptual reactivation for different imagery tasks.

Stronger activity in bilateral temporal auditory regions and frontal-parietal sensorimotor systems was observed during AI, compared to activity recruited for both sensorimotor activation and memory retrieval during HI. This supports the hypothesis that detailed auditory representations can be reactivated by the one-to-one ‘deterministic’ mapping between motor and perceptual systems ([Tian & Poeppel, 2012, 2013](#)). This mapping structure provides motor-to-sensory transformation dynamics that enrich the details of the representation, leading perhaps from phonemic to phonetic levels of detail ([Hickok, 2012](#)), which may not be available during memory retrieval. This result is also consistent with the behavioral observation that motor engagement enriches phonetic details, which can then influence speech-error rates at lexical-phonological and phonemic-articulatory levels during a covert tongue twister task ([Oppenheim & Dell, 2010](#)).

STS recruitment is commonly observed in speech and song production studies that manipulate auditory feedback (e.g. [Niziolek & Guenther, 2013](#); [Tourville, Reilly, & Guenther, 2008](#); [Zarate, Wood, & Zatorre, 2010](#); [Zarate & Zatorre, 2008](#); [Zheng, Munhall, & Johnsrude, 2010](#)). The observation of increased STS activity during AI in our study suggests similar computations between articulation imagery and self-monitoring during speech production. Whereas the auditory feedback manipulation during speech production actually creates the discrepancy between expectation and auditory input, the lack of auditory feedback during articulation imagery can also be considered as a similar violation of a sensory expectation generated after motor preparation and simulation. To support this, the location of our STS activation in the AI task [54, -26, 2] resembles the locations of STS activity reported in feedback perturbation studies: [52.8, -32.1, 4.4] in ([Niziolek & Guenther, 2013](#)), [58,-28,6] in ([Tourville et al., 2008](#)), and [54, -18, -6] in ([Zheng et al., 2010](#)). Therefore, we suggest that similar mechanisms for generating auditory predictions and subsequent comparisons with incoming auditory feedback may be carried out in STS during both speech imagery and speech monitoring.

[Price \(2012\)](#) implicates the aSTG in the early auditory processing of complex sounds. This anterior region of temporal gyrus has been found to be sensitive to rapid frequency transition ([Belin et al., 1998](#)) and spectral variation ([Zatorre & Belin, 2001](#)). Rapid frequency modulations are a key feature in words that might need to be internally reconstructed and parsed to distinguish between particular phonemes or syllables in speech. Thus, the observation of increased aSTG activity during AI might suggest that auditory representations of spectral transitions, similar to those seen during speech perception, can be internally induced without any external stimulation.

Our observed increase in activity within associative auditory cortices aSTG and pSTS (but not within primary auditory cortex) during speech imagery is consistent with findings in earlier

auditory imagery studies (e.g. [Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005](#); [Halpern & Zatorre, 1999](#); [Herholz, Halpern, & Zatorre, 2012](#); [Shergill et al., 2001](#); [Zatorre et al., 1996](#)). It should be noted, however, that some auditory imagery work has reported primary auditory cortex activation (e.g. [Kraemer et al., 2005](#); see [Zatorre & Halpern, 2005](#) for a review), but we speculate that imagining different levels of content complexity may require multiple levels of auditory processing, which could result in the recruitment of different stages along the auditory perceptual hierarchy. In the current study, we use spoken syllables as stimuli. Given the complex nature of these stimuli, the internal reconstruction of syllabic representation may occur beyond the computations and representations that are mediated by primary auditory cortex. Our MEG studies support this view -- the response latencies were modulated by the content of stimuli, for example at 200 ms for syllables ([Tian & Poeppel, 2013](#)) and 100ms for pitch ([Tian & Poeppel, 2015](#)), suggesting that simpler stimuli may only recruit lower (and therefore faster) level areas for auditory processing, whereas more complex stimuli recruit higher-order auditory regions within the auditory perceptual hierarchy and thus require additional processing time. Additional studies will need to be conducted to determine whether an auditory processing hierarchy can be accessed by increasingly complex imagined stimuli, as has been reported in the visual domain ([Kosslyn & Thompson, 2003](#)).

In summary, this study complements and extends beyond our earlier MEG study ([Tian & Poeppel, 2013](#)) by offering neuroanatomical evidence that supports the existence of two complementary neural pathways for perceptual reactivation. Two speech imagery tasks differentially recruit a motor-to-sensory transformation pathway and a memory-retrieval pathway. Moreover, stronger auditory responses in articulation imagery suggest that motor system involvement leads to stronger perceptual reactivation.

Acknowledgments

We thank Keith Sanzenbach for his technical assistance with fMRI recording, Tobias Overath and Thomas Schofield for their discussion and guidance with fMRI analyses, and Jess Rowland for her comments on this manuscript. This study was supported by MURI ARO #54228-LS-MUR, NIH 2R01DC 05660, a grant from the GRAMMY Foundation®, Major Projects Program of the Shanghai Municipal Science and Technology Commission (STCSM) 15JC1400104 and National Natural Science Foundation of China 31500914.

References

- Albright, T. D. (2012). On the perception of probable things: neural substrates of associative memory, imagery, and perception. *Neuron*, *74*(2), 227-245.
- Aristotle. (1986). *De Anima (on the soul)* (H. Lawson-Tancred, Trans.). London: Penguin Books.
- Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.-C., & Samson, Y. (1998). Lateralization of speech and auditory temporal processing. *Journal of Cognitive Neuroscience*, *10*(4), 536-540.
- Bensafi, M., Porter, J., Pouliot, S., Mainland, J., Johnson, B., Zelano, C., . . . Khan, R. (2003). Olfactomotor activity during imagery mimics that during perception. *Nature neuroscience*, *6*(11), 1142-1144.
- Berkeley, G. (1734/1965a). Three Dialogues between Hylas and Philonus In C. M. Turbayne (Ed.), *Principles, Dialogues, and Correspondence*. Indianapolis: Bobbs-Merrill.
- Berkeley, G. (1734/1965b). A Treatise Concerning the Principles of Human Knowledge. In C. M. Turbayne (Ed.), *Principles, Dialogues, and Correspondence*. Indianapolis: Bobbs-Merrill.
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767-2796.
- Blakemore, S.-J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*(8), 561-567.
- Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature neuroscience*, *1*(7), 635-640.
- Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage*, *32*(2), 821-841.
- Brendel, B., Hertrich, I., Erb, M., Lindner, A., Riecker, A., Grodd, W., & Ackermann, H. (2010). The contribution of mesiofrontal cortex to the preparation and execution of repetitive syllable productions: an fMRI study. *Neuroimage*, *50*(3), 1219-1230.
- Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H.-J., & Jancke, L. (2005). Scanning silence: mental imagery of complex sounds. *Neuroimage*, *26*(4), 1119-1127.
- Carr, M. F., Jadhav, S. P., & Frank, L. M. (2011). Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature neuroscience*, *14*(2), 147-153.
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, *18*(12), 2844-2854.
- Descartes, R. (1642/1984). *Meditations on First Philosophy* (J. Cottingham, R. Stoothoff & D. Murdoch, Trans.) *The Philosophical Writings of Descartes* (Vol. 2). Cambridge: Cambridge University Press.
- Djordjevic, J., Zatorre, R. J., Petrides, M., Boyle, J., & Jones-Gotman, M. (2005). Functional neuroimaging of odor imagery. *Neuroimage*, *24*(3), 791-801.
- Duvernoy, H. (1991). *The human brain: Structure, three-dimensional sectional anatomy and MRI*. Wien: Springer-Verlag.
- Eichenbaum, H., Sauvage, M., Fortin, N., Komorowski, R., & Lipton, P. (2012). Towards a functional organization of episodic memory in the medial temporal lobe. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1597-1608.
- Fox, P. T., Huang, A., Parsons, L. M., Xiong, J.-H., Zamarippa, F., Rainey, L., & Lancaster, J. L. (2001). Location-probability profiles for the mouth region of human primary motor-sensory cortex: model and validation. *Neuroimage*, *13*(1), 196-209.
- Friston, K. J., Penny, W. D., & Glaser, D. E. (2005). Conjunction revisited. *Neuroimage*, *25*(3), 661-667.
- Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, *9*(7), 697-704.

- Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience*, *24*(6), 1382-1397.
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135-145.
- Hobbes, T. (1651/1968). *Leviathan*. In C. Macpherson (Ed.). London: Penguin Books.
- Hume, D. (1969). *A Treatise of Human Nature*. In E. C. Mossner (Ed.). London: Penguin Books.
- James, W. (1890). *The principles of psychology* (Vol. 2). London: MacMillan.
- Ji, D., & Wilson, M. A. (2006). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature neuroscience*, *10*(1), 100-107.
- Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
- Kosslyn, S. M. (2005). Mental images and the brain. *Cognitive Neuropsychology*, *22*(3-4), 333-347.
- Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*(9), 635-642.
- Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J., Ganis, G., . . . Alpert, N. (1999). The role of area 17 in visual imagery: convergent evidence from PET and rTMS. *Science*, *284*(5411), 167.
- Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological bulletin*, *129*(5), 723.
- Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: sound of silence activates auditory cortex. *Nature*, *434*(7030), 158-158.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920-933.
- Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience*, *4*(4), 423-428.
- Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1273-1280.
- Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, *25*(3), 653-660.
- Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *The Journal of Neuroscience*, *33*(29), 12090-12098.
- O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013-1023.
- Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & cognition*, *38*(8), 1147-1160.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
- Palmer, E. D., Rosen, H. J., Ojemann, J. G., Buckner, R. L., Kelley, W. M., & Petersen, S. E. (2001). An event-related fMRI study of overt and covert word stem completion. *Neuroimage*, *14*(1), 182-193.
- Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature*, *362*(6418), 342-345.
- Penhune, V., Zatorre, R. J., MacDonald, J., & Evans, A. (1996). Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex*, *6*(5), 661-672.
- Picard, N., & Strick, P. L. (1996). Motor areas of the medial wall: a review of their location and functional activation. *Cerebral Cortex*, *6*(3), 342-353.
- Picard, N., & Strick, P. L. (2001). Imaging the premotor areas. *Current opinion in neurobiology*, *11*(6), 663-672.

- Plato. (1987). *Theaetetus* (R. A. Waterfield, Trans.). London: Penguin Books.
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, *62*(2), 816-847.
- Pylyshyn, Z. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, *88*(1), 16-45.
- Pylyshyn, Z. (2003). Return of the mental image: are there really pictures in the brain? *Trends in cognitive sciences*, *7*(3), 113-118.
- Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., & Ackermann, H. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology*, *64*(4), 700-706.
- Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. *Neuron*, *76*(4), 677-694.
- Schmahmann, J. D., Doyon, J., McDonald, D., Holmes, C., Lavoie, K., Hurwitz, A. S., . . . Petrides, M. (1999). Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. *Neuroimage*, *10*(3), 233-260.
- Shergill, S. S., Bullmore, E., Brammer, M., Williams, S., Murray, R., & McGuire, P. (2001). A functional study of auditory verbal imagery. *Psychological medicine*, *31*(02), 241-253.
- Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and language*, *93*(1), 20-31.
- Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Front. Psychology*, *1*(166). doi:10.3389/fpsyg.2010.00166
- Tian, X., & Poeppel, D. (2012). Mental imagery of speech: linking motor and sensory systems through internal simulation and estimation. *Front. Hum. Neurosci.*, *6*(314). doi: 10.3389/fnhum.2012.00314
- Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience*, *25*(7), 1020-1036.
- Tian, X., & Poeppel, D. (2015). Dynamics of Self-monitoring and Error Detection in Speech Production: Evidence from Mental Imagery and MEG. *Journal of Cognitive Neuroscience*, *27*(2), 352-364.
- Tomaiuolo, F., MacDonald, J., Caramanos, Z., Posner, G., Chiavaras, M., Evans, A. C., & Petrides, M. (1999). Morphology, morphometry and probability mapping of the pars opercularis of the inferior frontal gyrus: an in vivo MRI analysis. *European Journal of Neuroscience*, *11*(9), 3033-3046.
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, *39*(3), 1429-1443.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., . . . Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*(1), 273-289.
- Vilberg, K. L., & Rugg, M. D. (2008). Memory retrieval and the parietal cortex: a review of evidence from a dual-process perspective. *Neuropsychologia*, *46*(7), 1787-1799.
- Von Holst, E., & Mittelstaedt, H. (1950/1973). The reafference principle (R. Martin, Trans.) *The behavioral physiology of animals and man: The collected papers of Erich von Holst* (pp. 139-173): University of Miami Press.
- Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in cognitive sciences*, *9*(9), 445-453.
- Westbury, C., Zatorre, R. J., & Evans, A. (1999). Quantifying variability in the planum temporale: a probability map. *Cerebral Cortex*, *9*(4), 392-405.
- Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, *97*(20), 11125-11129.

- Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature neuroscience*, 3, 1212-1217.
- Wundt, W. M. (1913). *Elemente der Völkerpsychologie: Grundlinien einer psychologischen Entwicklungsgeschichte der Menschheit*. Leipzig: Kröner.
- Yoo, S. S., Freeman, D. K., McCarthy III, J. J., & Jolesz, F. A. (2003). Neural substrates of tactile imagery: a functional MRI study. *Neuroreport*, 14(4), 581.
- Zarate, J. M., Wood, S., & Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*, 48(2), 607-618.
- Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage*, 40(4), 1871-1887.
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 11(10), 946-953.
- Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: musical imagery and auditory cortex. *Neuron*, 47(1), 9-12.
- Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience*, 8(1), 29-46.
- Zhang, M., Weisser, V. D., Stilla, R., Prather, S., & Sathian, K. (2004). Multisensory cortical processing of object shape and its relation to mental imagery. *Cognitive, Affective, & Behavioral Neuroscience*, 4(2), 251-259.
- Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, 22(8), 1770-1781.

Figure captions

Figure 1. Dual stream prediction model (DSPM). Top: approximate cortical regions in the hypothesized dual streams. Bottom: schematic diagram of the DSPM (color scheme corresponds to the anatomical locations above). The abstract auditory representations (orange) can be induced via perception and perceptual reactivation and are formed around STG and STS. The perceptual reactivation process can be carried out in either the memory-retrieval or simulation-estimation prediction pathway. The memory-retrieval stream (blue) includes pMTG, MTL and distributed frontal-parietal networks for retrieval from long-term lexical items, episodic and semantic memory, respectively. The simulation-estimation stream (red) includes the frontal motor system and parietal somatosensory system. The articulatory trajectory is planned in frontal motor regions, including IFG, PMC, INS and SMA. If covert production is the goal, the planned articulation signal bypasses M1 and is simulated internally. The somatosensory consequence of the simulated articulation is estimated over parietal somatosensory regions, including SI, SII, PO and SMG. The auditory consequences -- in the form of an abstract auditory representation -- is derived from the subsequent estimation. A highly specified auditory representation (thick arrow) is generated in the bottom-up perceptual process that goes through spectrotemporal analysis of external stimuli in STG (brown). The stream containing the motor simulation and perceptual estimation processes can enrich the specificity of predicted auditory representations (solid arrows), compared to enrichment from the memory-retrieval stream (dotted arrows).

Abbreviations: STG, superior temporal gyrus; STS, superior temporal sulcus; pMTG, posterior middle temporal gyrus; MTL, middle temporal lobe; IFG, inferior frontal gyrus; PMC, premotor cortex; INS, insula; SMA, supplementary motor area; M1, primary motor cortex; PO, parietal

operculum; SI, primary somatosensory cortex; SII, secondary somatosensory cortex; and SMG, supramarginal gyrus.

Figure 2. Functional MRI data for the AI and HI tasks. Statistical parametric t-maps indicate strength of the BOLD signal ($p < 0.001$ uncorrected) with cluster size greater than 25 voxels. (a) Shared cortical regions that mediate both AI and HI. (b) Regions show greater activity in HI compared to that in AI, including MFG, IPC and IPS, consistent with frontal-parietal distributed memory retrieval networks. (c) Regions show greater activity in AI compared to that in HI, including SMC, PO, which demonstrate the greater recruitment of simulation-estimation stream in AI. Greater activity is also observed in aSTG and pSTS, suggesting that more specific and robust auditory representation is obtained during AI. Abbreviations: SMA, supplementary motor area; IFG, inferior frontal gyrus; INS, insula; FO, frontal operculum; PO, parietal operculum; PMC, premotor cortex; MFG, middle frontal gyrus; IPC, inferior parietal cortex; IPS, intraparietal sulcus; SMC, sensorimotor cortex; aSTG, anterior superior temporal gyrus; pSTS, posterior superior temporal sulcus.

Figure 3. Functional MRI data for common activation in AI and A. Statistical parametric t-maps indicate strength of the BOLD signal ($p < 0.001$ uncorrected) with cluster size greater than 25 voxels. Abbreviations: SMA, supplementary motor area; IFG, inferior frontal gyrus; INS, insula; ACC, anterior cingulate cortex; GP, globus pallidus; PO, parietal operculum.

Tables

Table 1. Shared neural network for AI and HI, revealed by conjunction analysis. Multiple peaks in a cluster are presented, with local maxima at least 8 mm apart. The threshold is set at $t > 3.65$, $P < 0.001$ uncorrected, with an extent of at least 25 voxels. The clusters are presented in descending size order (by number of voxels). Shading represents different clusters, if multiple distinct regions are found in the same cluster. Abbreviations: BA, Brodmann area; SMA, supplementary motor area.

Brain region	# voxels	t-stat	z-stat	MNI coordinates		
				x	y	z
Frontal						
L supplementary motor area (SMA)	1013	6.10	4.38	-6	-2	70
R pre-SMA		5.73	4.22	4	12	58
L inferior frontal pars opercularis (BA 44)	831	6.70	4.63	-58	6	12
L anterior insula		6.44	3.89	-36	10	6
R frontal operculum	380	5.99	4.33	48	8	6
R inferior frontal pars opercularis (BA 44)		4.48	3.59	54	14	4
L mid premotor cortex (BA 6)	146	5.92	4.30	-46	2	52
L primary motor cortex (near mouth region)		3.86	3.22	-42	-10	46
Parietal						
L parietal operculum	50	5.38	4.06	-54	-42	28
Cerebellum / Subcortical						
L cerebellum VI (declive)	30	4.81	3.77	-32	-60	-30
R cerebellum VI (declive)	45	4.66	3.69	32	-58	-30

Table 2. Regions of peak neural activity during HI compared with AI. Abbreviations: BA, Brodmann area; SMA, supplementary motor area.

Brain region	# voxels	t-stat	z-stat	MNI coordinates		
				x	y	z
Frontal						
L middle frontal (BA 8)	121	5.26	4.00	-38	10	36
L inferior frontal pars opercularis (BA 44)		4.31	3.49	-44	4	26
R inferior frontal pars triangularis (BA 45)	54	5.50	4.11	46	32	14
L pre-SMA	33	5.34	4.04	-6	18	50
Parietal						
L intraparietal sulcus	87	5.28	4.01	-42	-48	48
L inferior parietal gyrus	29	5.23	3.98	-28	-62	34

Table 3. Regions of peak neural activity during AI compared with HI. Abbreviations: BA, Brodmann area; aSTG, anterior superior temporal gyrus; pSTS, posterior temporal sulcus.

Brain region	# voxels	t-stat	z-stat	MNI coordinates		
				x	y	z
Frontal & Parietal						
R superior frontal (BA 9)	60	4.87	3.8	18	60	26
R primary motor cortex	40	5.88	4.29	24	-20	60
L middle frontal (BA 46)	34	4.45	3.57	-32	48	26
L sensorimotor cortex	31	4.78	3.75	-54	-12	40
L subcentral gyrus (Rolandic operculum)	31	4.44	3.57	-52	-2	8
R sensorimotor cortex	25	4.65	3.68	56	0	18
L parietal operculum	78	5.96	4.32	-58	-28	22
Temporal / Insular						
L aSTG	66	4.88	3.80	-54	6	-6
R pSTS	40	5.04	3.89	54	-26	2
L posterior insula	29	5.17	3.96	-38	-10	-12
Cerebellum / Subcortical						
R caudate	41	5.28	4.01	20	22	4

Table 4. Shared neural network for A and AI, revealed by conjunction analysis. Abbreviations:

BA, Brodmann area; SMA, supplementary motor area; ACC, anterior cingulate cortex.

Brain region	# voxels	t-stat	z-stat	MNI coordinates		
				x	y	z
Frontal / Insular						
L mid-dorsal insula	679	7.12	4.78	-34	8	10
L inferior frontal pars opercularis (BA 44)		5.84	4.27	-60	8	14
L SMA	584	6.03	4.35	-6	-2	72
R pre-SMA		5.71	4.21	6	4	68
R frontal operculum	425	6.36	4.49	46	6	6
R inferior frontal pars opercularis (BA 44)		5.18	3.96	56	14	2
R ACC (BA 32)	222	4.71	3.72	2	22	34
L ACC (BA 32)		4.34	3.51	-6	28	24
Parietal						
L parietal operculum	33	4.52	3.62	-52	-40	28
Cerebellum / Subcortical						
L putamen	70	4.94	3.84	-28	-8	14
L globus pallidus	66	6.55	4.57	-16	-12	0
R globus pallidus	45	5.25	4.00	14	0	-6
R caudate		3.81	3.20	8	6	4
R cerebellum VI (declive)	41	4.44	3.57	34	-60	-28
L cerebellum VI (declive)	32	4.81	3.77	-32	-60	-30

