NeuroImage 239 (2021) 118313

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Reading direct speech quotes increases theta phase-locking: Evidence for cortical tracking of inner speech?

Bo Yao[a,*], Jason R. Taylor[a], Briony Banks[b], Sonja A. Kotz[c,d]

[a] Division of Neuroscience and Experimental Psychology, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester M13 9PL, United Kingdom
[b] Department of Psychology, Lancaster University, Lancaster LA1 4YF, United Kingdom
[c] Department of Neuropsychology & Psychopharmacology, Maastricht University, Maastricht 6211 LK, Netherlands
[d] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig 04103, Germany

ARTICLE INFO

Keywords:
Inner speech
Neural oscillations
Phase synchrony
Phase-locking
Reading
Theta activity

ABSTRACT

Growing evidence shows that theta-band (4–7 Hz) activity in the auditory cortex phase-locks to rhythms of overt speech. Does theta activity also encode the rhythmic dynamics of inner speech? Previous research established that silent reading of direct speech quotes (e.g., *Mary said: "This dress is lovely!"*) elicits more vivid inner speech than indirect speech quotes (e.g., *Mary said that the dress was lovely*). As we cannot directly track the phase alignment between theta activity and inner speech over time, we used EEG to measure the brain's phase-locked responses to the onset of speech quote reading. We found that direct (vs. indirect) quote reading was associated with increased theta phase synchrony over trials at 250–500 ms post-reading onset, with sources of the evoked activity estimated in the speech processing network. An eye-tracking control experiment confirmed that increased theta phase synchrony in direct quote reading was not driven by eye movement patterns, and more likely reflects synchronous phase resetting at the onset of inner speech. These findings suggest a functional role of theta phase modulation in reading-induced inner speech.

1. Introduction

Inner speech is the subjective experience of speaking or hearing speech when no-one is talking out loud. It is a pervasive psychological phenomenon in human cognition (Heavey and Hurlburt, 2008). On the one hand, it plays an important role in thinking (Sokolov, 2012), problem solving (Baldo et al., 2005), working memory (Marvel and Desmond, 2012), reading (Yao and Scheepers, 2011, 2018; Yao and Scheepers, 2015; Yao, 2021), and writing (Chenoweth and Hayes, 2003). On the other, dysfunctions of inner speech are often associated with symptoms in mental health disorders such as rumination in depression, auditory verbal hallucinations in schizophrenia, and associated disorders (McCarthy-Jones and Fernyhough, 2011).

The diverse uses and functions of inner speech are supported by its many forms, varying along several phenomenological dimensions (Grandchamp et al., 2019; McCarthy-Jones and Fernyhough, 2011). Regarding its acoustic and structural details and how it is engaged, inner speech can be expanded or condensed, and can be intentional or spontaneous. Expanded inner speech preserves much of the phonological and syntactic qualities of overt speech whereas condensed inner speech keeps only the semantic core without verbal elaboration. Depending on the task at hand and context, inner speech may be engaged deliberately (rehearing phone numbers) or occur spontaneously (sounding words out loud in silent reading). The present study focuses on the neural mechanisms underlying expanded inner speech that is spontaneously induced during silent reading.

Expanded inner speech has been shown to share perceptual features with overt speech, particularly in tempo (Abramson and Goldinger, 1997; Alexander and Nygaard, 2008; Stites et al., 2013; Yao and Scheepers, 2011) and loudness (Tian et al., 2018). For instance, recent research shows that more vivid inner speech (especially in prosodic richness) can be induced by silent reading of direct speech quotations (e.g., *Mary said: "I'm hungry!"*) as compared to linguistically-matched indirect speech quotations (e.g., *Mary said [that] she was hungry*) (Stites et al., 2013; Yao and Scheepers, 2011; Yao et al., 2011). In line with embodied cognition theories (Barsalou, 2008; Zwaan, 2004), such inner speech may be mentally simulated from sensory states related to speech perception. As direct speech quotations are often perceived with more vivid vocal depictions in overt speech (Clark and Gerrig, 1990; Yao, 2011), sensory experiences of such vocal depictions can be mentally re-enacted in silent reading to induce perceptually vivid inner speech. Compared to indirect speech, direct speech quotations were read faster when they were preceded by descriptions indi-

* Corresponding author.
E-mail addresses: Bo.Yao@manchester.co.uk, bo.yao@manchester.ac.uk (B. Yao).

https://doi.org/10.1016/j.neuroimage.2021.118313.
Received 9 February 2021; Received in revised form 28 May 2021; Accepted 24 June 2021
Available online 25 June 2021.

cating a fast (vs. slow) speaking rate (e.g., *He said quickly* vs. *He said slowly*) (Stites et al., 2013; Yao and Scheepers, 2011). This suggests that quotation-induced inner speech must contain speech-like temporal features *in addition to* 'default' phonological processing in silent reading (Yao and Scheepers, 2015). Moreover, direct speech reading times were highly correlated between silent and oral reading, providing further evidence that quotation-induced inner speech shares temporal features with overt speech (Yao and Scheepers, 2011).

At the neural level, expanded inner speech is found to activate areas of the auditory cortex recruited in overt speech (Alderson-Day et al., 2016, 2020; Brück et al., 2014; McGuire et al., 1996; Yao et al., 2011; Yao et al., 2012). Using fMRI and eye tracking, Yao et al. (2011) compared neural responses to silent reading of direct vs. indirect speech quotes. While both kinds of reported speech activated the auditory cortex, direct speech quotes were associated with greater neural activity in areas of the auditory cortex that selectively respond to human voice (Belin et al., 2000). These areas were more active when direct (vs. indirect) speech quotes were read by equally monotonous voices, suggesting more vivid inner speech was induced in a top-down fashion to provide vivid prosodic representations that were expected for direct speech but were absent in the monotonous stimuli (Yao et al., 2012). While identifying the overlapping neural correlates between inner and overt speech represent encouraging progress in understanding inner speech, a detailed understanding of the exact neural mechanisms of inner speech is still lacking, particularly regarding its temporal features.

A recently discovered neural mechanism for encoding overt speech's temporal structure in perception involves the tracking of speech amplitude envelopes by neural oscillations (Giraud and Poeppel, 2012). The phase alignment between neural oscillations and speech envelopes provides an efficient means for encoding acoustic features (Ding and Simon, 2012), parsing syllabic boundaries (Giraud and Poeppel, 2012), and combining smaller linguistic units into larger structures (Ding et al., 2016; Gross et al., 2013). While syllables are tracked predominantly by theta (4–7 Hz) oscillations (Ding et al., 2016), larger linguistic units of words and phrases are tracked by even slower oscillations (<3 Hz, see Keitel et al., 2017; Meyer et al., 2017). Moreover, cortical speech tracking is functionally relevant for comprehension (Peelle and Davis, 2012) as listening to intelligible (vs. unintelligible) speech is often associated with more precise speech tracking (Gross et al., 2013; Luo and Poeppel, 2007).

Given the temporal similarities between speech perception and expanded inner speech, we ask whether neural oscillations play a similar role in inner speech as in overt speech perception. Unlike overt speech, inner speech is not an external signal for the brain to 'track' as such, but emerges directly from neural activity itself. A dominant theory proposes that inner speech is the perceptual consequence of intended articulation (Jack et al., 2019; Scott, 2013; Whitford et al., 2017). Under this framework, phase modulation of neural oscillations in the auditory cortex may be induced and entrained by motor signals (efference copies) from the speech production system (Assaneo and Poeppel, 2018). Since speech rhythms depend on motor constraints inherent in producing speech, the phase coupling between efference copies and neural oscillations would 'transfer' such dynamics to the auditory cortex, giving rise to a quasi-perceptual experience of speech. Other theories suggest that expanded inner speech may be reactivated memories of speech (Tian et al., 2016) or that it is perceptually simulated from a remix of stored speech features (Barsalou, 2008; Yao and Scheepers, 2015). In either case, oscillatory neural firing patterns that encode speech rhythms in perception would be reactivated endogenously, organizing phonological and prosodic representations in speech-like phase structures. Regardless of how expanded inner speech may be generated, its temporal features are likely encoded and modulated by neural oscillations in the auditory cortex.

The present study explored this conjecture. Given that overt speech is predominantly phase-locked to theta oscillations in perception (Assaneo and Poeppel, 2018; Giraud and Poeppel, 2012), we hypothesised a similar relationship for expanded inner speech. This seems plausible given its shared temporal features with overt speech (Stites et al., 2013; Yao and Scheepers, 2011), and that the syllabic rate of expanded inner speech in English (~5.8 per second) falls within the theta range of 4–7 Hz (Netsell et al., 2016). Although we cannot directly track the phase alignment between theta oscillations and inner speech over time, we can nonetheless measure theta phase-locking to the onset of inner speech, whose timing can be determined and aligned across trials in silent reading.

We tested these predictions with EEG (Experiment 1) via spontaneous induction of more vivid inner speech in silent reading of direct vs. indirect speech quotes (Alderson-Day et al., 2020; Stites et al., 2013; Yao and Scheepers, 2011; Yao et al., 2011; Yao, 2021). We chose this paradigm because it addresses two methodological limitations in inner speech research. First, many previous studies have elicited inner speech via subvocalisation or phonological judgements of simple words or sentences, which lacks the temporal complexity and spontaneity of everyday inner speech (Jones and Fernyhough, 2007). In contrast, the current paradigm induces naturalistic inner speech in silent reading without explicit instructions to imagine it, which is more ecologically valid than task-elicited inner speech (Hurlburt et al., 2016). Second, many inner speech elicitation tasks such as phonological judgements and speech imagery are often confounded by aspects of language processing (e.g., orthographic, semantic, and syntactic processes). The current paradigm controls such language confounds by manipulating inner speech between visually and linguistically matched reading conditions. As such, differential effects between conditions can only be attributed to inner speech manipulations rather than other language processes involved in reading. In this paradigm, if quotation-induced inner speech is temporally aligned with theta activity, we should observe increased theta phase-locking to the onset of direct compared to indirect speech quotes, with sources of the phase-locked activity estimated in bilateral auditory cortices (Binder et al., 2000; Hickok and Poeppel, 2007; Price et al., 1996; Scott and Johnsrude, 2003). To ensure that any such increased phase-locking reflects differences in phase modulation rather than power, we additionally manipulated the loudness of inner speech which was expected to affect signal amplitude rather than phase (Tian et al., 2018). Finally, we verified that theta phase modulation could not be explained by eye movement patterns for direct and indirect speech quotes in a separate control experiment (Experiment 2).

2. Experiment 1

2.1. Participants

Thirty-two native speakers of British English participated in the EEG study (10 male, 22 female, M_{age}=22.7, SD_{age}=6.4). All were right-handed, had normal or corrected-to-normal vision, no language or learning disorders, and no history of neurological or psychiatric disorders. They were paid £12 for 2 hours of their time. All participants gave written informed consent and the experimental procedure was approved by the University of Manchester Research Ethics Committee (ref: 16248).

2.2. Materials and experimental design

A 2 (Quotation Style: direct vs. indirect speech) × 2 (Loudness: loud- vs. quiet-speaking) within-subject design was used. One-hundred-and-twenty quartets of short stories were written as reading materials. Each story (see Table 1 for an example) described a scenario containing either a direct speech quote (1a, 2a) or an indirect speech quote (1b, 2b). To provide a variety of scenarios, the contexts preceding the quotations described either a loud-speaking (1) or a quiet-speaking (2) scenario. Crucially, critical speech quotations were identical across contexts and were matched word for word between the direct and indirect speech conditions except for unavoidable tense and pronoun changes. This ensured that speech quotations were matched across conditions of each

Table 1
An example quadruple of stories. The critical sentences are highlighted in bold.

1	*The roaring noise of the fire was incredible as flames engulfed the house. Gareth and the other firefighters burst into the house, smashing down the front door to gain entry. As his colleagues found the downstairs to be empty, Gareth battled his way upstairs through the thick smoke*
a	*After checking the upstairs rooms, Gareth bellowed: "**It looks like there is nobody here!**"*
b	*After checking the upstairs rooms, Gareth bellowed that **it looked like there was nobody there**.*
2	*It was a dark night when agents Gareth and James broke into the defence Secretary's residence, looking for evidence of conspiracy. The house seemed really quiet. Carefully forcing the lock, Gareth and his colleague tip-toed inside.*
a	*Gareth turned to James and whispered: "**It looks like there is nobody here!**"*
b	*Gareth turned to James and whispered that **it looked like there was nobody there**.*

item for length (number of words and syllables) and other linguistic characteristics such as grammatical complexity, so as to isolate inner speech from potential linguistic confounds.

In addition to the 120 critical test items, 60 filler stories (without experimental manipulations) were prepared to conceal the intended experimental manipulations. Of the 60 stories, 24 contained direct speech quotes, 12 contained indirect speech quotes, and another 24 did not contain any quoted speech.

The 480 critical stories were allocated to four stimulus lists using a Latin square design. Each list contained 120 stories with 30 stories per condition, plus the 60 filler stories. The order of the stories per list was randomised for each participant. Each list was randomly assigned to 8 participants.

2.3. Task procedure

Participants were seated in a sound-attenuated and electrically-shielded room to silently read a series of written stories. The experiment was run in OpenSesame (Mathôt et al., 2012). The visual stimuli were presented on a gray background in a 30-pixel Sans font on a 24-inch monitor (120 Hz, 1024 × 768 resolution) approximately 100 cm from the participant.

The experiments started with 5 filler trials to familiarise participants with the procedure, after which the remaining 120 critical trials and 55 filler trials were presented in a random order. Each trial began with the trial number for 1000 ms, followed by a fixation dot on the left side of the screen (where the text would start) for 500 ms. The story was then presented in five consecutive segments at the center of the screen. Participants silently read each segment in their own time, and pressed the DOWN key on a keyboard to continue to the next segment. Of the five segments, the first three segments of each story described the story background. The 4th displayed the text preceding the speech quotation (e.g., *After checking the upstairs rooms, Gareth bellowed:*) and the 5th segment displayed the speech quotation (e.g., *"It looks like there is nobody here!"*). In about a third of the trials, a simple question (e.g., *Was the house empty?*) was presented to measure participants' comprehension, which participants answered by pressing the LEFT ('yes') or RIGHT ('no') keys. Answering the question triggered the presentation of the next trial.

Participants were given a short break every 20 trials, and there were 8 breaks in total. The experiment lasted approximately 45–60 min.

2.4. EEG acquisition and preprocessing

EEG and EOG activity was recorded with an analog passband of 0.16–100 Hz and digitised at a sampling rate of 512 Hz using a 64-channel Biosemi Active-Two system. The 64 scalp electrodes were mounted in an elastic electrode cap according to the international 10/20 system. Six external electrodes were used: two were placed on bilateral mastoids, two were placed above and below the right eye to measure vertical ocular activity (VEOG), and another two were placed next to the outer canthi of the eyes to record horizontal ocular activity (HEOG). Electrode-offset values were kept between −25 mV and 25 mV.

The recorded EEG data were preprocessed in EEGLAB v14.1.2b (Delorme and Makeig, 2004). All electrodes were referenced to a mastoid average. EOG activity was calculated by subtracting the EOG signals within each pair of EOG electrodes. The data of 66 (2 EOG) channels were high-pass filtered at 0.3 Hz to remove slow drifts and were down-sampled to 200 Hz because the raw data were recorded at a sampling rate higher than actually needed for the analysis. For each participant, 120 critical trials were segmented from −1200 to 2000 ms relative to the presentation onsets of speech quotations (segment 5). Artifact trials were automatically marked based on (1) whether the EEG amplitudes exceed +/−100 μV in the [−200 1000]ms time window across 64 EEG channels, and (2) whether the probability of observing the trial's data is 5 standard deviations from the mean EEG values within each EEG channel and across all 64 EEG channels (Delorme et al., 2001). The marked trials were then visually reviewed to check for validity. Trials with common, ICA-removable artifacts (e.g., blink-related peaks that exceeded +/−100μV thresholds) and trials with artifacts outside the critical [−200 1000]ms window were kept. This resulted in an overall mean trial loss of 2.2%, with 2.5%, 2.0%, 2.4% and 2.1% in each of the four conditions. The remaining trials were filtered at 2–25 Hz and were submitted to an ICA to isolate eye movement and other artifacts. Using the 'runica' algorithm with the default options, the ICA included both EEG and EOG channels and produced 66 components in total. Artifact components were identified using a semi-automated procedure: Components for eye blinks and muscular artifacts were classified automatically using the EEGLAB extension 'MARA' (Winkler et al., 2011). Components for HEOG were additionally identified if their activation time-courses were strongly correlated with HEOG activity ($|r|>0.7$). All components were visually reviewed before being declared artifact signals and removed (number of components removed per participant ranged 3–23, M=11.3, SD=5.0). The resulting ICA weights (minus artifact components) were projected back to the pre-filtered data before ICA (0.3–100 Hz). The EOG channels were dropped from further analyses.

2.5. Statistical analysis

2.5.1. Reading time analysis

Reading times for direct and indirect speech quotes (sentence segment 5) were determined from the sentence presentation onsets to participants' key presses. They were divided by the number of syllables to account for variation in sentence length. For this analysis only, we excluded trials with extreme reading times that were longer than 500 ms per syllable (1.8% data loss). This cut-off was selected based on the distribution of reading times per syllable and on the fact that fixation durations in silent reading of English rarely go beyond 500 ms (Rayner, 1998). To check any systematic differences in reading times between conditions, a Gamma generalised linear mixed model of *RTs per syllable* was fitted using the *glmer* function in the lme4 package (Bates et al., 2015) in R. We included a full factorial fixed effect structure with deviation-coded *Quotation Style* and *Loudness*, and a maximal random effect structure with *Subject* and *Item* as crossed random factors.

2.5.2. Sensor space time-frequency analysis

At the subject level, time-frequency analyses of single-trial EEG data (to calculate intertrial phase-locking values (PLVs) and total power) were conducted using Morlet wavelet decomposition with seven cycles per wavelet, at frequencies from 1 to 30 Hz in SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). PLV is also referred to as intertrial phase coherence (ITC). It measures the variability in the relative

phases over trials. It takes values from 0 to 1 with 0 reflecting no phase synchrony across trials and 1 reflecting identical phase in all trials (see equation 6 in Aydore et al., 2013). Total power averages power over multiple trials and takes positive values (0=no signal). Both PLVs and total power were calculated at each time-frequency point and channel, with the latter being log-scaled and baseline corrected to [−200 0] time window. They were averaged by condition using a robust averaging procedure where statistical outliers in narrow time and frequency ranges were down-weighted without rejecting whole trials (Litvak et al., 2011). The averaged PLV and total power were then converted into Nifti images by condition for each participant, including topography × time images for theta-band analysis and time × frequency images for broad-band analysis. These images are 3D (x, y, time) or 2D (time, frequency) matrices of averaged PLV and power. For topography × time images, the 2D representation of the topography is created by projecting the sensor locations onto a plane, before interpolating the data linearly between them onto a 32 × 32 pixel grid. The converted Nifti images were submitted to a general linear model for statistical analysis using statistical parametric mapping (SPM) to compare conditions over time, frequency, and topographical space. Gaussian smoothing was applied to the scalp × time volumes to accommodate spatial/temporal variability over subjects and ensure the images conform to the assumptions of the topological inference approach (Litvak et al., 2011).

2.5.2.1. Planned theta-band analysis. Topography × time maps were generated, averaging PLVs and total power within the theta (4–7 Hz) frequency band. Data were interpolated to create a 32 × 32 pixel (4.3 mm × 5.4 mm) scalp map for each time point from −200 to 1000 ms relative to the quotation onset (i.e. when the critical speech quotation is presented). Topographic images were stacked to create a 3D space-time image volume. The volume was smoothed with a Gaussian kernel at FWHM=[16 *mm* 16 *mm* 16 *ms*], about 3 times of voxel size, which is in accordance with the assumptions of Random Field Theory (Kiebel and Friston, 2004; Worsley et al., 1996).

2.5.2.2. Broad-band analysis. To verify that the observed results reflect theta-specific rather than broad-band phase-locked responses, a full time-frequency analysis was conducted. Time × frequency maps were generated, averaging PLVs and total power across all channels for each integer frequency from 1 to 30 Hz, and for each time point from −200 to 1000 ms relative to the quotation onset. No smoothing was applied to ensure precise estimation of PLV and total power in the frequency dimension.

2.5.2.3. Group analysis. Group-level analyses used *F*-tests to assess the effects of Quotation Style and Loudness on PLVs and total power across the scalp and time and across time and frequency. The resulting mass-univariate SPMs entail a statistical test at each of tens of thousands of voxels and therefore require correction for multiple comparisons. Familywise error (FWE) correction was applied at the cluster level at $p<.05$, with a cluster defining threshold of $p<.001$ (i.e., clusters consisted of voxels 'surviving' this uncorrected threshold) using Random Field Theory (Flandin and Friston, 2017) which takes image smoothness into account.

2.5.3. Source space analysis

Source estimation was performed on single-trial time-domain data using a template cortical mesh (Mattout et al., 2007). The mesh consists of 8196 nodes, tessellating the gray/white matter boundary of a single individual, with a mean inter-node distance of 4 mm. The neural generators were constrained to a lattice of dipoles on the cortical mesh, oriented perpendicular to its surface. A forward model was defined using a Boundary Element Method (BEM), which was inverted under the minimum norm (MN) hyperprior model. The MN model was chosen over the multiple sparse priors (MSP) model because it deploys reconstructed activity in a non-focal fashion, and hence is most resilient against inter-subject variability in group analysis (Litvak and Friston, 2008).

Estimated source activity was summarised in 3D NIfTI images by condition by subject, which were smoothed with a Gaussian kernel (FWHM=[8 *mm* 8 *mm* 8 *mm*]). The time/frequency window was selected based on the sensor-level results. They were then submitted to group-level *F* tests to assess the effects of Quotation Style and Loudness. Similar to the sensor space analysis, the resulting SPMs were familywise error (FWE) corrected for multiple comparisons at the cluster level ($p<.05$; with a cluster defining threshold of $p<.001$ using Random Field Theory (Flandin and Friston, 2017).

To verify theta phase synchrony in the source space, we extracted source time-courses using a 5-mm spherical ROI at the peak voxel in each cluster. The time series underwent the same time-frequency transformation and averaging as in the sensor data analysis to estimate the PLVs at the same time/frequency windows for inversion.

Table 2
Reading times per syllable and comprehension question accuracies across conditions in Experiment 1.

	Loud-Speaking		Quiet-Speaking	
	Direct Speech	Indirect Speech	Direct Speech	Indirect Speech
	Quotation Reading Times Per Syllable (ms)			
N	946	943	947	935
Mean	166	165	164	164
S.D.	78	80	76	75
	Comprehension Question Accuracy			
N	480	480	480	480
Mean	0.921	0.927	0.888	0.867
S.D.	0.270	0.260	0.316	0.340

Note: *N* = number of trials, S.D.= Standard Deviation.

2.6. Results and discussion

All participants were debriefed after the experiment. They found the experiment 'interesting' and 'enjoyable' but none consciously noticed the experimental manipulations on quotation styles or loudness.

2.6.1. Reading time analysis

Mean reading times per syllable and comprehension question accuracies are summarised by condition in Table 2. The GLMMs of RTs/syllable and comprehension question accuracies showed no significant effects of Quotation Style, Loudness or their interaction ($ps>0.15$), suggesting that reading times and comprehension were statistically indistinguishable between conditions.

2.6.2. Sensor space analysis

We observed significant main effects of Quotation Style in both theta-band and broad-band analyses.

In theta frequencies (4–7 Hz), we found significantly higher PLV over trials in silent reading of direct speech than indirect speech quotes from approximately 250–500 ms following sentence presentation onsets. The increased phase-locked effects were clustered in the left temporal and parietal channels (Fig. 1 top left). No significant Loudness main effects or Quotation Style × Loudness interaction were observed. In terms of total power, we did not find significant power differences between direct and indirect speech in the topography × time analysis (Fig. 1 bottom left), nor did we find any significant Loudness main effects or Quotation Style × Loudness interaction.

In the broad-band (time-frequency) analysis across all channels, we observed increased PLV at 5 Hz but not in other frequencies between 1 and 30 Hz (Fig. 1 top right). No significant Loudness main effects or Quotation Style × Loudness interaction were observed. In comparison, we did not observe significant power differences between direct and indirect speech (Fig. 1 bottom right), nor did we observe any signifi-

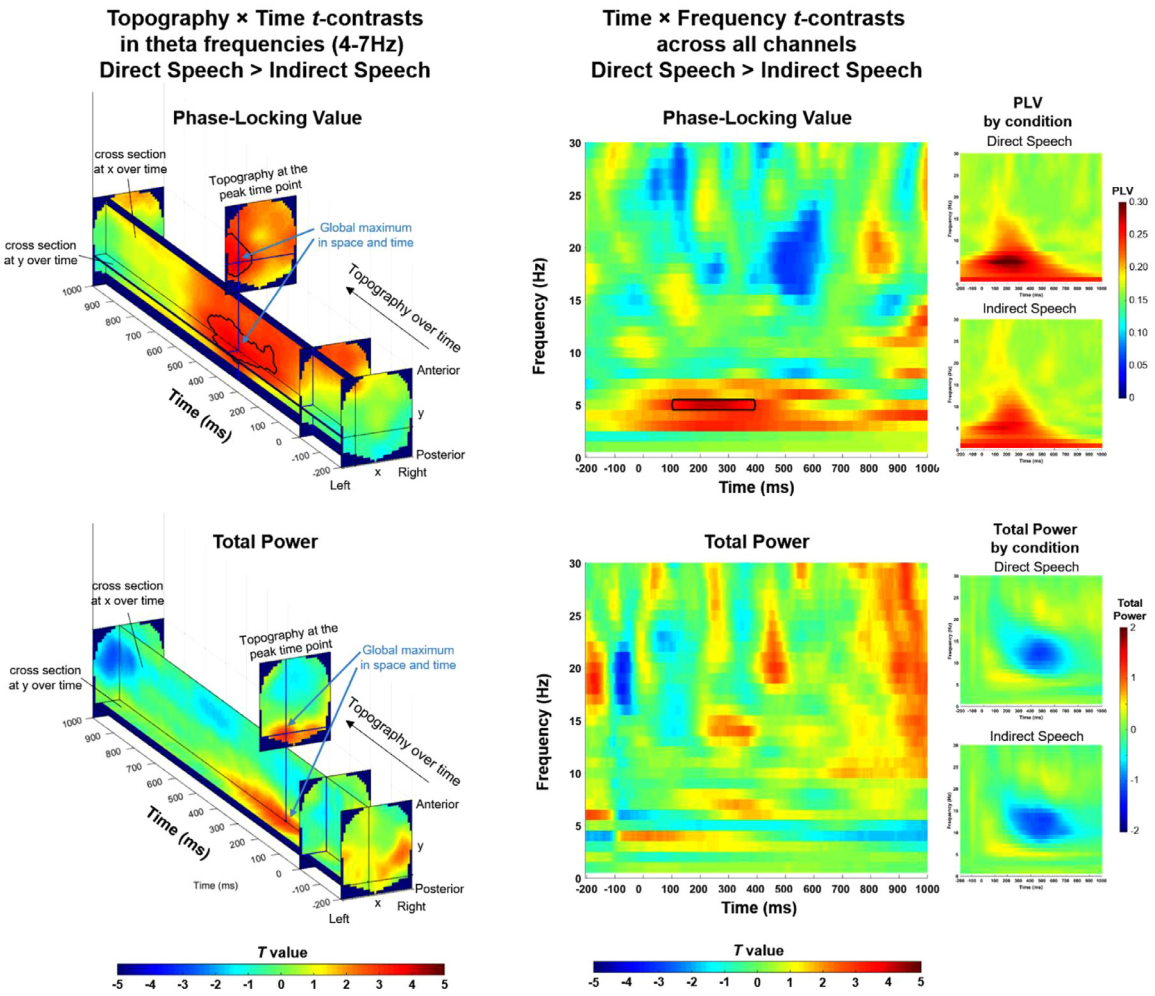

Fig. 1. Direct Speech > Indirect Speech *t*-contrasts for intertrial phase-locking values (top figures) and total power (bottom figures). The left figures show *t*-values over topography and time in theta frequencies (4–7 Hz). The right figures show *t*-values and condition averages across all channels over time and frequency. In all figures, time 0 indicates the onset of speech quotations (segment 5).

Note: Black contours indicate topography-time and time-frequency regions that survived p<.05 FWE-correction at the cluster level. In the topography-time images, blue lines mark the global maximum over topography and time. The image slices (cross sections) pass through this peak coordinate. Also shown is the scalp topography at the peak time point (positioned above the main topography x time image series for clear viewing).

cant Loudness main effects or Quotation Style × Loudness interaction for power.

The results support the hypothesised higher phase synchrony over trials after the reading onset of direct relative to indirect speech quotes. The increased phase synchrony is specific to theta frequencies, particularly at 5 Hz. This phase-locked effect did not coincide with increases in total power which suggests neuronal responses to direct and indirect speech reading differ predominantly in phase modulation of theta activity. The lack of Loudness effects in either phase or power suggest that (1) inner speech in silent reading of direct quotes may not contain detailed loudness information and that (2) the loudness of inner speech may only be detectable indirectly using explicit imagery tasks and neural adaptation paradigms (Tian et al., 2018).

It is worth noting that the PLV topographies in our study are more left-lateralised than in other auditory studies in the literature. For example, the auditory evoked N1 response is typically associated with a fronto-central topography and can be affected by corollary discharge (e.g., Rosburg et al., 2008; Tian et al., 2018). We speculate that internally-generated speech at the sentence level may produce different topographies than externally-perceived auditory stimuli. While biaural auditory stimulation may produce more bilateral responses, inner speech may be internally generated from a more left-lateralised language network. Moreover, the impoverished spectro-temporal details in inner speech means that it may engage a slightly different location in the STG/STS than, e.g., bilateral primary auditory cortex, which may result in a slightly different orientation and different projection to the scalp. Thus, to estimate the possible sources of the theta phase-locked effects, we conducted further analyses in the source space.

2.6.3. Source space analysis

We performed source estimation (see methods) and summarised the phase-locked source energy in a time- (250–500 ms from sentence presentation onset) and frequency- (4–7 Hz) window, based on the sensor space results. We observed different source activity between direct and indirect speech quotes in the left occipito-temporal and fusiform area (BA37), bilateral ventral and middle temporal areas (BA20/21), and the left inferior and middle frontal area (BA45/46). The source activity difference (thresholded at $p<.001$, uncorrected) is illustrated in Fig. 2. The MNI coordinates of the significant peaks and sub-peaks are provided in Table 3 to indicate the cortical regions that are likely to have contributed to the phase-locked effects observed on the scalp. These peaks are several cm apart and sources at these locations should be easily distinguishable via the minimum norm estimation method. It is worth not-

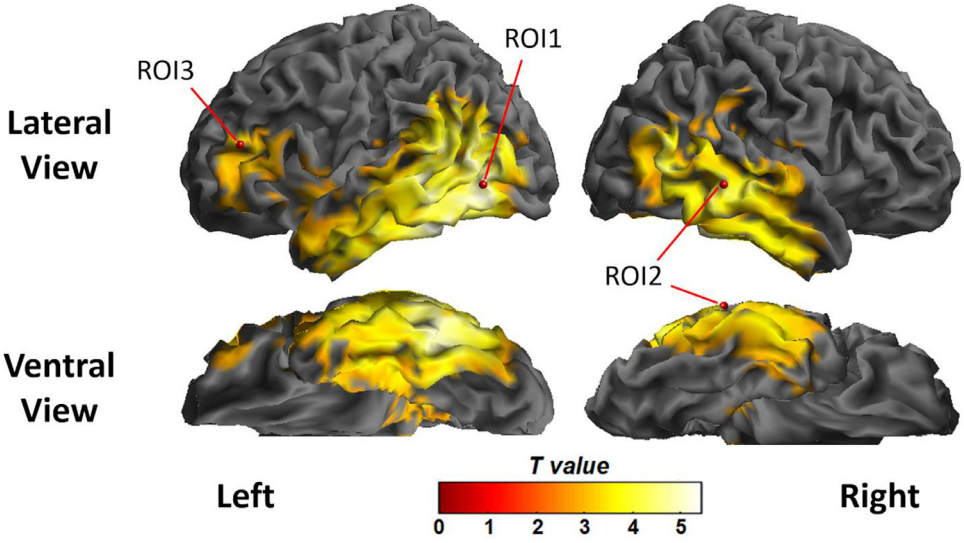

Fig. 2. Direct Speech > Indirect Speech *t*-contrast for source activity at 250–500 ms and 4–7 Hz.
Note: also shown are three ROIs for the source space phase-locking analysis.

Table 3
Whole-brain coordinates for Direct Speech > Indirect Speech *t*-contrast for source activity at 250–500 ms and 4–7 Hz, thresholded at $p<.001$ uncorrected.

Location	X	Y	Z	k	*t*	*z*	$p_{cluster}$
Cluster 1							
L Inferior Occipital / Temporal	−48	−70	−4	6580	5.41	5.03	<0.001
L Fusiform	−42	−52	−12		5.32	4.96	
L Middle Temporal	−62	−46	−10		5.24	4.89	
Cluster 2							
R Middle Temporal	64	−40	−4	3530	4.67	4.41	<0.001
R Inferior / Middle Temporal	60	−48	−10		4.66	4.41	
R Middle Temporal	56	−30	−8		4.64	4.39	
Cluster 3							
L Inferior Frontal (pars triangularis)	−44	40	14	462	4.09	3.91	.069
L Inferior Frontal (pars triangularis)	−36	34	12		3.82	3.67	
L Inferior Frontal (pars triangularis)	−50	30	6		3.76	3.62	

Note: *L* = left, *R* = right, X,Y,Z are coordinates in MNI space, *k* = cluster size, *t* = *t* value, *z* = *z* value, $p_{cluster}$ = *p* value at the cluster level.

ing that source localization errors for 64-channel EEG could amount to ~3 cm in distance (Song et al., 2015). As such, these peaks may not reflect the exact location of the phase-locked source energy, but general estimates of where it may originate from. These general estimates and their corresponding interpretations would not be affected by location errors on the order of ~3 cm.

Intertrial phase-locking analysis of the source ROIs, i.e. the peak voxel in each of the three clusters (Table 3), confirmed significantly higher PLVs in direct than indirect speech quotes at the left inferior occipital-temporal [−48 −70 −4], $F_{(1,31)}$=12.2, p=.001, and the right middle temporal [64 −40 −4] ROIs, $F_{(1,31)}$=12.6, p=.001, with the difference marginally significant at the left inferior frontal [−44 40 14] ROI, $F(1,31)$=3.5, p=.07.

The source space analysis confirmed that the sources of increased theta phase-locked responses in direct speech can be estimated in the bilateral temporal cortices (superior temporal sulcus, medial and inferior parts of temporal lobes) and the left inferior frontal areas. These areas are broadly in line with neural correlates of inner speech identified in previous fMRI studies (Alderson-Day et al., 2016, 2020; Yao et al., 2011). However, they do not quite match a typical motor-to-auditory corollary discharge circuit, which often involves more posterior parts of the inferior frontal cortex (e.g., pars opercularis) for speech planning, and somatosensory cortex and/or secondary auditory cortex for corollary discharge (Tian and Poeppel, 2013). Rather, they agree more with a memory retrieval-based simulation circuit where lexico-semantic information and episodic memories in the prefrontal, medial and inferior temporal, and inferior parietal regions are retrieved and transformed into auditory representations of speech (Price, 2012; Tian et al., 2016).

Notably, the core inner speech circuit is complemented by additional visual processes in the occipital and fusiform areas (cf. similar occipito-fusiform fMRI activations in Yao et al., 2011). One possibility is that direct speech quotations are associated with more vivid *multisensory* simulations that include both auditory and visual aspects of the protagonist speaking. However, because occipito-fusiform activations are not observed when listening to direct vs. indirect speech quotations (Yao et al., 2012), these visual processes may be specific to reading. Increased theta phase synchrony in the fusiform areas may implicate greater phase-locked visual word form processing and orthographic-phonological conversion to inner speech, which may be necessary for inner speech to occur in silent reading.

Although the above interpretations may be plausible, one could also argue that different theta phase-locked activity in occipito-parietal and temporal regions may not be driven by differential vividness of inner speech or grapheme-to-phoneme conversion, but by different eye movement patterns in direct vs. indirect quote reading. Theoretically, the phase of an EEG signal can be affected by oculomotor and visual processes in three ways. First, phase resetting can be caused by oculomotor and muscular activity at saccade onset (Berg and Scherg, 1991). Second, visually evoked responses following fixation onset can change its ongoing phase (Ossandón et al., 2010). Third, phase can be disturbed by subsequent saccades if current fixation duration is not equally distributed between conditions (Nikolaev et al., 2016).

To rule out potential eye movement-related effects on theta phase synchrony, Experiment 2 recorded eye movements in silent reading of the same direct vs. indirect speech quotes. We compared the distribution of saccades preceding and following first fixations to test the effects of oculomotor and muscular activity on phase resetting, and compared the distribution of first fixation onsets and durations to examine the effects of visually evoked responses on phase. Increased theta phase synchrony in silent reading of direct speech quotes could be explained by more concentrated distribution (less dispersion) of first fixations and/or neighbouring saccades, time-locked to sentence presentation onset. If this distribution is not statistically distinguishable between conditions, it would suggest that the increased theta phase synchrony could not be driven by reading differences between conditions, and would be more plausibly explained by increased inner speech when reading direct quotes.

3. Experiment 2

3.1. Participants

Twenty-four native speakers of British English who did not participate in Experiment 1 participated in the eye tracking control experiment (12 male, 12 female, M_{age}=24.3, SD_{age}=6.1). The inclusion criteria were identical to the EEG experiment. They were paid £6 for one hour of their time. All participants gave written informed consent and the experimental procedure was approved by the University of Manchester Research Ethics Committee (ref: 16248).

3.2. Materials and experimental design

The materials and experimental design were identical to the EEG experiment.

3.3. Task procedure

The task procedure was identical to the EEG experiment except that the experiment was conducted in an eye tracking lab. A SR-Research EyeLink 1000 eye tracker was used, running at 500 Hz sampling rate. Viewing was binocular but only the right eye was tracked. A chin rest was applied to keep the viewing distance constant and to prevent strong head movements during reading.

3.4. Eye movement data preprocessing

Raw EDF data files were first converted into ASC files using a file converter provided by SR Research. Fixation and saccade events as well as timestamps for the trial ID and sentence presentation onsets were then extracted.

3.5. Results and discussion

Similar to Experiment 1, all participants were debriefed after the experiment and none of them consciously noticed the experimental manipulations on quotation styles or loudness.

3.5.1. Reading time analysis

As per Experiment 1, we excluded trials with extreme reading times that were longer than 500 ms per syllable (2.3% data loss) for the reading time analysis only. Mean reading times per syllable and comprehension question accuracies are summarised in Table 4. The GLMM of RTs/syllable showed no significant effects (ps>0.21), suggesting that reading performance was statistically indistinguishable between conditions. The GLMM of comprehension question accuracies showed no effects of Quotation Style or Quotation Style × Loudness interaction (ps>0.54). However, it did reveal a significant Loudness main effect, b=0.69, z=2.32, p=.021, with answers to comprehension questions were more accurate following loud-speaking scenarios (97.2%) than quiet-speaking scenarios (94.5%). This 'loudness advantage' in comprehension was mirrored in Experiment 1 which did not reach significance. Since this main effect was not directly relevant to our central contrast of direct vs. indirect speech, it was not explored any further. Overall, reading time and comprehension performance in Experiment 2 are comparable to those in Experiment 1.

Table 4
Reading times per syllable and comprehension question accuracies across conditions in Experiment 2.

	Loud-Speaking		Quiet-Speaking	
	Direct Speech	Indirect Speech	Direct Speech	Indirect Speech
	Quotation Reading Times Per Syllable (ms)			
N	708	704	701	702
Mean	163	162	168	159
S.D.	75	80	82	79
	Comprehension Question Accuracy			
N	360	360	360	360
Mean	0.933	0.922	0.881	0.875
S.D.	0.250	0.268	0.324	0.331

Note: N = number of trials, S.D.= Standard Deviation.

3.5.2. Eye movement analysis

A single ROI around sentence segment 5 (direct and indirect speech quotes) was created. First fixations were defined as the first fixations that landed in the ROI. Pre-first-fixation saccade onsets, first fixation onsets and durations, as well as post-first fixation saccade onsets (first fixation offsets) were calculated from the presentation onset of sentence segment 5. Pre-first-fixation saccades that started before the stimulus presentation were included in the analysis but not shown in distribution plots in Fig. 3.

As we were primarily interested in *statistical dispersion* differences of eye movement distribution, standard deviations of the four eye movement measures were summarised by condition at the subject level. Their means were also calculated to check any fixation/saccade latency differences between conditions. The by-subject standard deviations and means were submitted to paired-sample t-tests at the group level for statistical comparisons. To evaluate the evidence for the alternative (H1) hypothesis, Bayes Factors (BF_{10}) were also calculated at the group level using the Jeffreys-Zellner-Siow (JZS) prior (Bayarri and Garcia-Donato, 2007). Both descriptive and inferential statistics are reported in Table 5.

There were no significant differences in any of the eye movement measures between silent reading of direct and indirect speech quotes $|ts_{(23)}|$<0.895, ps>0.384, BF_{10}s<0.309. All BF_{10}s were below 1/3, providing substantial evidence for H0 over H1 (Jeffreys, 1998). As such, Experiment 2 verified that the increased theta phase synchrony over trials in direct speech quotes could not be attributed to differences in reading patterns between direct and indirect speech quotes and was more likely to reflect differences in inner speech.

4. General discussion

Motivated by findings on the phase alignment between theta activity and overt speech (Assaneo and Poeppel, 2018; Giraud and Poeppel, 2012), the current study tested a similar phase relationship between theta activity and expanded inner speech by focusing on theta phase synchrony over trials at the onset of reading-induced inner speech. We used an established paradigm to induce perceptually vivid inner speech during silent reading of direct (vs. indirect) speech quotes (Alderson-Day et al., 2020; Brück et al., 2014; Stites et al., 2013; Yao and Scheepers, 2011; Yao et al., 2011; Yao, 2021). Using EEG (Experiment 1), we observed increased phase synchrony over trials, but no change in power, in theta frequencies (4–7 Hz) at the onset of direct over indirect speech quotes. Different phase-locked source activity was also observed

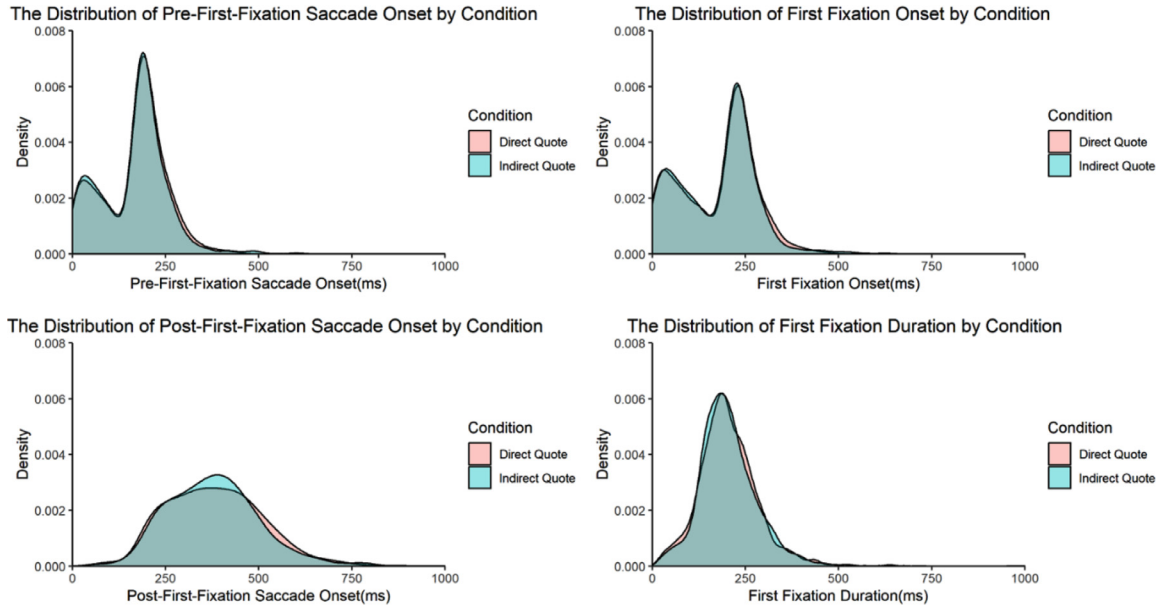

Fig. 3. Density plots of onsets for first fixations and neighbouring saccades (from the sentence presentation onset) and durations for first fixations by condition.

Table 5
Group-level standard deviations and means of eye movement measures in ms across conditions.

Dispersion of Onsets / Duration (Standard Deviation)					
	Direct Speech	Indirect Speech	$t_{(23)}$	p	BF_{10}
Pre-1st Fix Sac Onset	113	112	.066	.948	.215
1st Fix Onset	102	103	−.539	.595	.245
1st Fix Duration	83	80	.703	.489	.268
Post-1st Fix Sac Onset	131	126	.894	.381	.308
Means of Onsets / Duration					
	Direct Speech	Indirect Speech	$t_{(23)}$	p	BF_{10}
Pre-1st Fix Sac Onset	134	131	.599	.555	.253
1st Fix Onset	177	175	−.516	.611	.242
1st Fix Duration	205	205	.166	.870	.217
Post-1st Fix Sac Onset	382	379	.507	.617	.241

Note: 1st Fix = First Fixation; Sac = Saccade.

between direct and indirect speech quotes in the left occipito-temporal and fusiform area (BA37), bilateral ventral and middle temporal areas (BA20/21), and the left inferior and middle frontal area (BA45/46). In contrast, no Loudness effect was observed in phase synchrony, which was in line with previous findings that the loudness of inner speech affects EEG amplitude rather than phase. However, the Loudness modulation of power was not observed in the current study, suggesting that the loudness of inner speech may be impoverished in silent reading, and may only be fully activated and reliably detectable during explicit imagery tasks using neural adaptation paradigms (Tian et al., 2018). Using eye tracking (Experiment 2), we found no eye movement differences at the reading onset of direct vs. indirect speech quotes, which ruled out the possibility that the theta phase-locking differences were driven by differential eye movement patterns.

In particular, the differential theta phase synchrony cannot be explained by reading patterns between direct and indirect speech quotes. It is known that oral readers typically insert a pause before a direct speech quote (e.g., *She said:* [pause] *"I am hungry!"*) but not before an indirect speech quote (e.g., *She said that she was hungry*) (Yao, 2011). A similar pause may take place during silent reading of direct speech quotes and cause phase resetting at the onset of direct but not indirect speech quotes. We took precautions by using a self-paced reading paradigm where a key press was required, thereby creating an artificial pause before the presentation of speech quotes in *both* direct and indirect speech conditions. Even if a reading pause did precede direct speech quotes, it would have been equalised by the artificial pauses preceding both conditions. This reading paradigm worked as intended as there was no statistical difference in reading times between conditions in the EEG or the eye tracking experiment. The latter experiment further verified that eye movements (first fixations and the neighbouring saccades) in quotation reading were statistically indistinguishable between direct and indirect speech quotes.

The increased theta phase synchrony over trials at the onset of direct quote reading is therefore more plausibly explained by increased inner speech processing. First, previous research has consistently demonstrated more vivid inner speech during silent reading of direct rather than indirect speech quotes (Stites et al., 2013; Yao and Scheepers, 2011, 2018; Yao et al., 2011). Second, the present study observed increased phase synchrony over trials in theta frequencies only and at 5 Hz in particular, but did not observe theta power differences. Our findings are consistent with the speech tracking literature, which typically reports entrainment and reset of theta phase but not theta power (Gross et al., 2013; Luo and Poeppel, 2007; Peelle et al., 2013) and reports optimal auditory-motor synchrony in the theta range, particularly at ~4.5 Hz (Assaneo and Poeppel, 2018; Giraud and Poeppel, 2012). Third, the sources of the increased phase-locked activity were estimated in bilat-

eral temporal cortices, the left inferior frontal gyrus, and in the occipito-temporal areas. These regions have been respectively associated with auditory speech processing (Binder et al., 2000; Hickok and Poeppel, 2007; Price et al., 1996; Scott and Johnsrude, 2003), covert articulation/verbal working memory (Paulesu et al., 1993; Shergill et al., 2001, 2002), and orthographic-phonological conversion (Blomert, 2011; Hashimoto and Sakai, 2004), all of which are plausible components of inner speech in reading (Alderson-Day and Fernyhough, 2015). Although the exact roles of these brain regions remain to be established in inner speech, they are nonetheless compatible with an inner speech account of the observed theta phase results.

However, it remains inconclusive whether the increased theta phase synchrony reflects greater evoked responses (Obleser et al., 2012) at the onset of inner speech or greater phase modulation of ongoing oscillations (Luo and Poeppel, 2007). One possibility is that evoked responses in direct speech quotes may reflect heightened motor imagery (of vocalization) during inner speech. Given that covert articulation is a key component of inner speech (Alderson-Day and Fernyhough, 2015), readers may engage in stronger, or more effortful subvocalization of direct speech quotes, particular when they are loudly rather than quietly spoken. However, no loudness effects were observed. The observed direct speech effects were neither detected in beta frequencies, which are typically modulated by motor imagery (Kühn et al., 2006), nor observed in articulation-related motor areas (e.g., the premotor cortex) in the source analysis. It was therefore unlikely that the increased theta activity in direct speech quotes was driven by motor imagery. A second possibility relates to potentially greater speech monitoring or attention in inner speech (Perrone-Bertolotti et al., 2014). In direct speech quote reading, higher degrees of self-monitoring or attention may be required to ensure that distinct phonological features are activated to represent the quoted speaker's voice, rather than one's own (Clark and Gerrig, 1990). This may prepare the 'ventral' speech processing network (Hickok and Poeppel, 2007, 2016) in anticipating distinct inner speech to facilitate comprehension of the quoted speech. As the speech processing network is most sensitive at theta frequencies (Giraud and Poeppel, 2012), increased phase-locked responses in this range may reflect a transient burst of anticipatory signals for distinctly vivid inner speech *before* the onset of direct quote reading. This explanation is not backed by the data. If self-monitoring/attention needs to be maintained during distinctly vivid inner speech, power increases should be sustained throughout direct quote reading. Although both direct and indirect quote reading were associated with alpha power decreases, no significant power differences were observed between the two conditions. Moreover, the increased phase synchrony over direct speech quotes was detected at 250–500 ms post reading onset but not before, suggesting that it was not of an anticipatory nature. Moreover, the lack of loudness effects suggests that the observed differences in theta phase synchrony between direct and indirect speech quotes was more likely to reflect distinct phase (rather than power) modulation of theta activity. Thus, the increased phase synchrony is most likely to reflect increased phase resetting of theta oscillations (Luo and Poeppel, 2007), which signals the start of more vivid inner speech in direct speech reading. Just as theta oscillations encode the rhythms of overt speech, they may also encode the rhythms of inner speech in direct speech reading. Although phonological representations are matched almost word for word between direct and indirect speech reading, they may be arranged in different rhythms. While the indirect speech rhythm may follow the default rhythm of reading, the direct speech rhythm may deviate from it in a more speech-like arrangement, giving rise to a more distinct, vivid speech percept than one's default 'reading voice'. Such a rhythmic deviation in direct speech would necessarily cause a reset of ongoing oscillations at the onset of reading, resulting in increased theta phase synchrony over trials. Because different sentences were used, the exact rhythmic structures of inner speech varied largely across trials. While intertrial phase patterns may be more synchronous at the start of direct speech reading (due to increased phase resetting), they became increasingly asynchronous as reading continued (with varying sentence structures) and were eventually indistinguishable from intertrial phase patterns in indirect quote reading.

Several open questions remain to be addressed. One concerns whether inner speech has an 'envelope' similar to overt speech. Indeed, word skipping and regressive eye movements in silent reading means that inner speech could be more fragmented and scrambled, and may not necessarily exhibit the same envelope structure as overt speech in perception (Brumberg et al., 2016). Given the current technology, the 'envelope' of inner speech is not objectively measurable, and we are yet able to provide more direct evidence for cortical tracking of such an 'envelope' like in overt speech. Future research will need to characterize the temporal structures of inner speech in different tasks and develop new methods to directly measure inner speech tracking.

A second open question regards why increased phase synchrony in inner speech is not observed at higher frequencies than those typically observed in overt speech (e.g., theta frequencies). It is commonly recognised that silent reading (and hence inner speech) is faster than overt speech because it does not involve explicit articulation (Alderson-Day and Fernyhough, 2015). It therefore seems counterintuitive that inner and overt speech would share similar timescales in an electrophysiological context. However, although inner speech is evidently faster than overt speech, the rate difference is relatively small. Recent research shows that the rate of expanded inner speech (phonologically detailed inner speech) is only ~11% faster than overt speech (5.8 Hz vs. 5.2 Hz) which is still within the theta range (Netsell et al., 2016). Moreover, word skipping during silent reading (Rayner, 1998) means that only key parts of a sentence are 'sampled' and converted into inner speech. This kind of 'fragmented' inner speech may give the illusion of being faster (as it covers the same amount of text in a shorter time) but may still possess temporal properties of overt speech (Yao and Scheepers, 2011). To test whether the frequencies of phase-locking depend on the rate of inner speech, future research will need to test whether phase-locked response would be observed at higher frequencies for fast vs. slow inner speech.

A third question concerns how individual differences in the vividness of their inner speech may account for the observation or lack of effects. Vividness of inner speech varies largely between individuals (Alderson-Day et al., 2018) and individuals' sensitivity to the reading of direct speech quotes is also likely to differ. Through our post-experiment conversations with participants, we learned that some consciously imagined vivid voices during speech quote reading while others had no conscious awareness of an internal voice; some imagined specific people's voices (their friends or family) for the quoted speakers while others used their own inner speech during reading. The different levels of awareness and uses of inner speech inevitably introduced noise in our data, which may render weaker effects (e.g., loudness) more difficult to detect. That being said, we did observe significantly increased theta phase-locking during silent reading of direct (vs. indirect) speech quotes, highlighting that phase modulation may be a relatively robust and universal consequence of expanded inner speech in silent reading. To understand the effects of different kinds of inner speech, future research will need to model individual differences in inner speech and capture the specific type of inner speech used on a single-trial basis.

In sum, the present study characterised a neurophysiological correlate of inner speech in silent reading of direct (vs. indirect) speech quotes. The results showed increased theta phase synchrony over trials at the onset of direct quote reading. This phase modulation is most plausibly explained by perceptually vivid inner speech in direct quote reading. Although we cannot directly track the phase alignment between theta activity and inner speech over time, our findings open up an exciting research avenue towards a mechanistic understanding of inner speech. Future investigations will need to develop new methods for measuring the phase structure of inner speech and examine its temporal relations to theta oscillations and to eye movements in reading.

Data and code availability statement

The summarised behavioural data, SPM contrast images and analyses scripts (R scripts and MatLab scripts) that support the findings of this study are available at: https://reshare.ukdataservice.ac.uk/854892/ (DOI: 10.5255/UKDA-SN-854892)

Credit authorship contribution statement

Bo Yao: Conceptualization, Methodology, Software, Investigation, Validation, Formal analysis, Data curation, Visualization, Supervision, Writing – original draft, Writing – review & editing, Funding acquisition. **Jason R. Taylor:** Validation, Formal analysis, Writing – review & editing. **Briony Banks:** Resources, Software, Investigation, Writing – review & editing. **Sonja A. Kotz:** Conceptualization, Supervision, Writing – review & editing.

Acknowledgements

This work was supported by the Economic and Social Research Council (ES/N002784/1) and the Bial Foundation (284/2018) to BY.

References

Abramson, M., Goldinger, S.D., 1997. What the reader's eye tells the mind's ear: silent reading activates inner speech. Percept. Psychophys. 59 (7), 1059–1068.
Alderson-Day, B., Fernyhough, C., 2015. Inner speech: development, cognitive functions, phenomenology, and neurobiology. Psychol. Bull. 141 (5), 931–965. doi:10.1037/bul0000021.
Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., Fernyhough, C., 2018. The varieties of inner speech questionnaire - revised (VISQ-R): replicating and refining links between inner speech and psychopathology. Conscious. Cogn. 65, 48–58. doi:10.1016/j.concog.2018.07.001.
Alderson-Day, B., Moffatt, J., Bernini, M., Mitrenga, K., Yao, B., Fernyhough, C., 2020. Processing speech and thoughts during silent reading: direct reference effects for speech by fictional characters in voice-selective auditory cortex and a theory-of-mind network. J. Cogn. Neurosci. 32 (9), 1637–1653. doi:10.1162/jocn_a_01571.
Alderson-Day, B., Weis, S., McCarthy-Jones, S., Moseley, P., Smailes, D., Fernyhough, C., 2016. The brain's conversation with itself: neural substrates of dialogic inner speech. Soc. Cogn. Affect. Neurosci. 11 (1), 110–120. doi:10.1093/scan/nsv094.
Alexander, J.D., Nygaard, L.C., 2008. Reading voices and hearing text: talker-specific auditory imagery in reading. J. Exp. Psychol. Hum. Percept. Perform. 34 (2), 446–459. doi:10.1037/0096-1523.34.2.446.
Assaneo, M.F., Poeppel, D., 2018. The coupling between auditory and motor cortices is rate-restricted: evidence for an intrinsic speech-motor rhythm. Sci. Adv. 4 (2), eaao3842. doi:10.1126/sciadv.aao3842.
Aydore, S., Pantazis, D., Leahy, R.M., 2013. A note on the phase locking value and its properties. Neuroimage 74, 231–244. doi:10.1016/j.neuroimage.2013.02.008.
Baldo, J.V., Dronkers, N.F., Wilkins, D., Ludy, C., Raskin, P., Kim, J., 2005. Is problem solving dependent on language? Brain Lang. 92 (3), 240–250. doi:10.1016/j.bandl.2004.06.103.
Barsalou, L.W., 2008. Grounded cognition. Annu. Rev. Psychol. 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639.
Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. doi:10.18637/jss.v067.i01.
Bayarri, M.J., Garcia-Donato, G., 2007. Extending conventional priors for testing general hypotheses in linear models. Biometrika 94 (1), 135–152. doi:10.1093/biomet/asm014.
Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403 (6767), 309–312. doi:10.1038/35002078.
Berg, P., Scherg, M., 1991. Dipole models of eye movements and blinks. Electroencephalogr. Clin. Neurophysiol. 79 (1), 36–44. doi:10.1016/0013-4694(91)90154-V.
Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10 (5), 512–528. doi:10.1093/cercor/10.5.512.
Blomert, L., 2011. The neural signature of orthographic-phonological binding in successful and failing reading development. Neuroimage 57 (3), 695–703. doi:10.1016/j.neuroimage.2010.11.003.
Brück, C., Kreifelts, B., Gößling-Arnold, C., Wertheimer, J., Wildgruber, D., 2014. Inner voices ': the cerebral representation of emotional voice cues described in literary texts. Soc. Cogn. Affect. Neurosci. 9 (11), 1819–1827. doi:10.1093/scan/nst180.
Brumberg, J.S., Krusienski, D.J., Chakrabarti, S., Gunduz, A., Brunner, P., Ritaccio, A.L., Schalk, G., 2016. Spatio-temporal progression of cortical activity related to continuous overt and covert speech production in a reading task. PLoS ONE 11 (11), e0166872. doi:10.1371/journal.pone.0166872.
Chenoweth, N.A., Hayes, J.R., 2003. The inner voice in writing. Written Commun. 20 (1), 99–118. doi:10.1177/0741088303253572.
Clark, H.H., Gerrig, R.J., 1990. Quotations as demonstrations. Language (Baltim) 66 (4), 764–805.
Delorme, A., Makeig, S., Sejnowski, T., 2001. Automatic artifact rejection for EEG data using high-order statistics and independent component analysis. In: *Proceedings of the Third International* Independent Component Analysis and Blind Source Decomposition Conference.
Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21. doi:10.1016/j.jneumeth.2003.10.009.
Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19 (1), 158–164. doi:10.1038/nn.4186.
Ding, N., Simon, J.Z., 2012. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J. Neurophysiol. 107 (1), 78–89. doi:10.1152/jn.00297.2011.
Flandin, G., Friston, K.J., 2017. Analysis of family-wise error rates in statistical parametric mapping using random field theory. Hum. Brain Mapp. 40 (7), 2052–2054. doi:10.1002/hbm.23839.
Giraud, A.-L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15 (4), 511–517. doi:10.1038/nn.3063.
Grandchamp, R., Rapin, L., Perrone-Bertolotti, M., Pichat, C., Haldin, C., Cousin, E., … Lœvenbruck, H., 2019. The condialint model: condensation, dialogality, and intentionality dimensions of inner speech within a hierarchical predictive control framework. Front. Psychol. 10, 2019. doi:10.3389/fpsyg.2019.02019.
Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., Garrod, S., 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol. 11 (12), e1001752. doi:10.1371/journal.pbio.1001752.
Hashimoto, R., Sakai, K.L., 2004. Learning letters in adulthood. Neuron 42 (2), 311–322. doi:10.1016/S0896-6273(04)00196-5.
Heavey, C.L., Hurlburt, R.T., 2008. The phenomena of inner experience. Conscious. Cogn. 17 (3), 798–810. doi:10.1016/j.concog.2007.12.006.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8 (5), 393–402. doi:10.1038/nrn2113.
Hickok, G., Poeppel, D., 2016. Neural basis of speech perception. In: Neurobiology of Language. Elsevier, pp. 299–310. doi:10.1016/B978-0-12-407794-2.00025-0.
Hurlburt, R.T., Alderson-Day, B., Kühn, S., Fernyhough, C., 2016. Exploring the ecological validity of thinking on demand: neural correlates of elicited vs. spontaneously occurring inner speech. PLoS ONE 11 (2), e0147932. doi:10.1371/journal.pone.0147932.
Jack, B.N., Le Pelley, M.E., Han, N., Harris, A.W.F., Spencer, K.M., Whitford, T.J., 2019. Inner speech is accompanied by a temporally-precise and content-specific corollary discharge. Neuroimage 198, 170–180. doi:10.1016/j.neuroimage.2019.04.038.
Jeffreys, H., 1998. The Theory of Probability. OUP Oxford.
Jones, S.R., Fernyhough, C., 2007. Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations. Conscious. Cogn. 16 (2), 391–399. doi:10.1016/j.concog.2005.12.003.
Keitel, A., Ince, R.A.A., Gross, J., Kayser, C., 2017. Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. Neuroimage 147, 32–42. doi:10.1016/j.neuroimage.2016.11.062.
Kiebel, S.J., Friston, K.J., 2004. Statistical parametric mapping for event-related potentials: I. Generic considerations. Neuroimage 22 (2), 492–502. doi:10.1016/j.neuroimage.2004.02.012.
Kühn, A.A., Doyle, L., Pogosyan, A., Yarrow, K., Kupsch, A., Schneider, G.-H., … Brown, P., 2006. Modulation of beta oscillations in the subthalamic area during motor imagery in Parkinson's disease. Brain: J. Neurol. 129 (Pt 3), 695–706. doi:10.1093/brain/awh715.
Litvak, V., Friston, K., 2008. Electromagnetic source reconstruction for group studies. Neuroimage 42 (4), 1490–1498. doi:10.1016/j.neuroimage.2008.06.022.
Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Friston, K., 2011. EEG and MEG data analysis in SPM8. Comput. Intell. Neurosci. 2011, 852961. doi:10.1155/2011/852961.
Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54 (6), 1001–1010. doi:10.1016/j.neuron.2007.06.004.
Marvel, C.L., Desmond, J.E., 2012. From storage to manipulation: how the neural correlates of verbal working memory reflect varying demands on inner speech. Brain Lang. 120 (1), 42–51. doi:10.1016/j.bandl.2011.08.005.
Mathôt, S., Schreij, D., Theeuwes, J., 2012. OpenSesame: an open-source, graphical experiment builder for the social sciences. Behav. Res. Methods 44 (2), 314–324. doi:10.3758/s13428-011-0168-7.
Mattout, J., Henson, R.N., Friston, K.J., 2007. Canonical source reconstruction for MEG. Comput. Intell. Neurosci. 67613. doi:10.1155/2007/67613.
McCarthy-Jones, S., Fernyhough, C., 2011. The varieties of inner speech: links between quality of inner speech and psychopathological variables in a sample of young adults. Conscious. Cogn. 20 (4), 1586–1593. doi:10.1016/j.concog.2011.08.005.
McGuire, P.K., Silbersweig, D.A., Wright, I., Murray, R.M., Frackowiak, R.S., Frith, C.D., 1996. The neural correlates of inner speech and auditory verbal imagery in schizophrenia: relationship to auditory verbal hallucinations. Br. J. Psychiatry 169 (2), 148–159. doi:10.1192/bjp.169.2.148.
Meyer, L., Henry, M.J., Gaston, P., Schmuck, N., Friederici, A.D., 2017. Linguistic bias modulates interpretation of speech via neural delta-band oscillations. Cereb. Cortex 27 (9), 4293–4302. doi:10.1093/cercor/bhw228.
Netsell, R., Kleinsasser, S., Daniel, T., 2016. The rate of expanded inner speech during spontaneous sentence productions. Percept. Mot. Skills 123 (2), 383–393. doi:10.1177/0031512516664992.
Nikolaev, A.R., Meghanathan, R.N., van Leeuwen, C., 2016. Combining EEG and eye movement recording in free viewing: pitfalls and possibilities. Brain Cogn. 107, 55–83. doi:10.1016/j.bandc.2016.06.004.

Obleser, J., Herrmann, B., Henry, M.J., 2012. Neural oscillations in speech: don't be enslaved by the envelope. Front. Hum. Neurosci. 6, 250. doi:10.3389/fnhum.2012.00250.

Ossandón, J.P., Helo, A.V., Montefusco-Siegmund, R., Maldonado, P.E., 2010. Superposition model predicts EEG occipital activity during free viewing of natural scenes. J. Neurosci. 30 (13), 4787–4795. doi:10.1523/JNEUROSCI.5769-09.2010.

Paulesu, E., Frith, C.D., Frackowiak, R.S., 1993. The neural correlates of the verbal component of working memory. Nature 362 (6418), 342–345. doi:10.1038/362342a0.

Peelle, J.E., Davis, M.H., 2012. Neural oscillations carry speech rhythm through to comprehension. Front. Psychol. 3, 320. doi:10.3389/fpsyg.2012.00320.

Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb. Cortex 23 (6), 1378–1387. doi:10.1093/cercor/bhs118.

Perrone-Bertolotti, M., Rapin, L., Lachaux, J.P., Baciu, M., Lœvenbruck, H., 2014. What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. Behav. Brain Res. 261, 220–239. doi:10.1016/j.bbr.2013.12.034.

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. doi:10.1016/j.neuroimage.2012.04.062.

Price, C.J., Wise, R.J., Warburton, E.A., Moore, C.J., Howard, D., Patterson, K., Friston, K.J., 1996. Hearing and saying. The functional neuro-anatomy of auditory word processing. Brain: J. Neurol. 119 (Pt 3), 919–931. doi:10.1093/brain/119.3.919.

Rayner, K., 1998. Eye movements in reading and information processing: 20 years of research. Psychol. Bull. 124 (3), 372–422. doi:10.1037/0033-2909.124.3.372.

Rosburg, T., Boutros, N.N., Ford, J.M., 2008. Reduced auditory evoked potential component N100 in schizophrenia–a critical review. Psychiatry Res. 161 (3), 259–274. doi:10.1016/j.psychres.2008.03.017.

Scott, M., 2013. Corollary discharge provides the sensory content of inner speech. Psychol. Sci. 24 (9), 1824–1830. doi:10.1177/0956797613478614.

Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26 (2), 100–107. doi:10.1016/S0166-2236(02)00037-1.

Shergill, S.S., Bullmore, E.T., Brammer, M.J., Williams, S.C., Murray, R.M., McGuire, P.K., 2001. A functional study of auditory verbal imagery. Psychol. Med. 31 (2), 241–253. doi:10.1017/s003329170100335x.

Shergill, S.S., Brammer, M.J., Fukuda, R., Bullmore, E., Amaro, E., Murray, R.M., McGuire, P.K., 2002. Modulation of activity in temporal cortex during generation of inner speech. Hum. Brain Mapp. 16 (4), 219–227. doi:10.1002/hbm.10046.

Sokolov, A., 2012. Inner Speech and Thought. Springer Science & Business Media.

Song, J., Davey, C., Poulsen, C., Luu, P., Turovets, S., Anderson, E., … Tucker, D., 2015. EEG source localization: sensor density and head surface coverage. J. Neurosci. Methods 256, 9–21. doi:10.1016/j.jneumeth.2015.08.015.

Stites, M.C., Luke, S.G., Christianson, K., 2013. The psychologist said quickly, "dialogue descriptions modulate reading speed!. Mem. Cognit. 41 (1), 137–151. doi:10.3758/s13421-012-0248-7.

Tian, X., Ding, N., Teng, X., Bai, F., Poeppel, D., 2018. Imagined speech influences perceived loudness of sound. Nat. Hum. Behav. 2 (3), 225–234. doi:10.1038/s41562-018-0305-8.

Tian, X., Poeppel, D., 2013. The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. J. Cogn. Neurosci. 25 (7), 1020–1036. doi:10.1162/jocn_a_00381.

Tian, X., Zarate, J.M., Poeppel, D., 2016. Mental imagery of speech implicates two mechanisms of perceptual reactivation. Cortex 77, 1–12. doi:10.1016/j.cortex.2016.01.002.

Whitford, T.J., Jack, B.N., Pearson, D., Griffiths, O., Luque, D., Harris, A.W., … Le Pelley, M.E., 2017. Neurophysiological evidence of efference copies to inner speech. Elife 6. doi:10.7554/eLife.28197.

Winkler, I., Haufe, S., Tangermann, M., 2011. Automatic classification of artifactual ICA-components for artifact removal in EEG signals. Behav. Brain Funct. 7, 30. doi:10.1186/1744-9081-7-30.

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

Yao, B., 2011. *Mental Simulations in Comprehension of Direct* Versus *Indirect Speech Quotations*.

Yao, B., Scheepers, C., 2011. Contextual modulation of reading rate for direct versus indirect speech quotations. Cognition 121 (3), 447–453. doi:10.1016/j.cognition.2011.08.007.

Yao, B., Scheepers, C., 2018. Direct speech quotations promote low relative-clause attachment in silent reading of English. Cognition 176, 248–254. doi:10.1016/j.cognition.2018.03.017.

Yao, B., 2021. Mental simulations of phonological representations are causally linked to silent reading of direct versus indirect speech. J. Cognit. 4 (1), 6. doi:10.5334/joc.141.

Yao, B., Belin, P., Scheepers, C., 2011. Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. J. Cogn. Neurosci. 23 (10), 3146–3152. doi:10.1162/jocn_a_00022.

Yao, B., Belin, P., Scheepers, C., 2012. Brain "talks over" boring quotes: top-down activation of voice-selective areas while listening to monotonous direct speech quotations. Neuroimage 60 (3), 1832–1842. doi:10.1016/j.neuroimage.2012.01.111.

Yao, B., Scheepers, C., 2015. Inner voice experiences during processing of direct and indirect speech. In: Frazier, L., Gibson, E. (Eds.), Explicit and Implicit Prosody in Sentence Processing. Springer International Publishing, Cham, pp. 287–307. doi:10.1007/978-3-319-12961-7_15 Vol. 46.