



Full Length Article

Brain ‘talks over’ boring quotes: Top-down activation of voice-selective areas while listening to monotonous direct speech quotations

Bo Yao ^{a,*}, Pascal Belin ^{a,b}, Christoph Scheepers ^a^a Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, G12 8QB, UK^b Centre for Cognitive Neuroimaging (CCNi), University of Glasgow, Glasgow, G12 8QB, UK

ARTICLE INFO

Article history:

Received 26 September 2011

Revised 31 December 2011

Accepted 22 January 2012

Available online 28 January 2012

Keywords:

Direct speech

Indirect speech

fMRI

Speech perception

Mental simulation

Language comprehension

Emotional prosody

ABSTRACT

In human communication, direct speech (e.g., *Mary said, “I’m hungry”*) is perceived as more vivid than indirect speech (e.g., *Mary said that she was hungry*). This vividness distinction has previously been found to underlie silent reading of quotations: Using functional magnetic resonance imaging (fMRI), we found that direct speech elicited higher brain activity in the temporal voice areas (TVA) of the auditory cortex than indirect speech, consistent with an “inner voice” experience in reading direct speech. Here we show that listening to monotonously spoken direct versus indirect speech quotations also engenders differential TVA activity. This suggests that individuals engage in top-down simulations or imagery of enriched supra-segmental acoustic representations while listening to monotonous direct speech. The findings shed new light on the acoustic nature of the “inner voice” in understanding direct speech.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Direct speech (e.g., *Mary said, “I’m hungry”*) and indirect speech (e.g., *Mary said that she was hungry*) are two important reporting styles in everyday communication. The use of direct speech usually coincides with vivid demonstrations of reported speech acts whereas indirect speech provides mere descriptions of what was said (Clark and Gerrig, 1990). For instance, the former usually contains vivid depictions of the reported speaker’s voice while the latter does not. Recently, this “vividness” distinction has been shown to underlie language comprehension of the two reporting styles in written text. Combining event-related fMRI and eye-tracking, Yao et al. (2011) found that silent reading of direct speech elicited higher brain activity in the voice-selective areas or temporal voice areas (TVA; Belin et al., 2000) of the right auditory cortex than silent reading of indirect speech. Moreover, Yao and Scheepers (2011) observed that such an “inner voice” experience in reading direct rather than indirect speech was also reflected in behavioural articulation (oral reading) and eye-movement patterns (silent reading). The findings are in line with embodied cognition in language processing (Barsalou, 1999; Zwaan, 2004), suggesting that individuals are more likely to mentally simulate or imagine the reported speaker’s voice in understanding direct speech as opposed to indirect speech.

However, with no acoustic stimulation as a reference, it is still unclear what constitutes such “inner voice” experiences during *silent reading* of direct as opposed to indirect speech. In Yao et al. (2011), we speculated that the mentally simulated voice representations entail supra-segmental acoustic information of the quoted speaker’s voice (e.g., speech melodies, intonation and emotional prosody), given that a right-lateralised activation pattern was observed. Indeed, “bottom-up” auditory stimulation studies have shown the same lateralisation by contrasting speech or music with acoustically matched noise bursts (Zatorre et al., 1992, 1994), by contrasting speech signals (irrespective of intelligibility) with noise-vocoded signals (Scott et al., 2000), and by contrasting nonverbal sounds comprising extended frequency transitions (supra-segmental) with those comprising rapid frequency transitions (sub-segmental) (Johnsrude et al., 2000). Hence, it seems likely that the right superior temporal gyrus/sulcus (STG/STS) areas are involved in processing dynamic pitch variations which are an important property of supra-segmental vocal information. One type of such information, namely emotional prosody and intonation, is also found to activate similar, right-lateralised activation patterns in various forms including sentences (Mitchell et al., 2003; Wildgruber et al., 2005), words (Wiethoff et al., 2008) and word-like vocalisations (Grandjean et al., 2005). Most importantly, mental simulations of supra-segmental acoustic information in language comprehension would fit well with the notion of direct speech as vivid demonstration – in which vivid depictions of the quoted speaker’s voice are characterised in terms of enriched supra-segmental acoustic information.

* Corresponding author.

E-mail address: b.yao@psy.gla.ac.uk (B. Yao).

In this paper, we attempted to address whether the mentally simulated voices predominantly consist of supra-segmental acoustic information during comprehension of direct as opposed to indirect speech. In order to create acoustic references to verify such supra-segmental information, we prepared audio recordings of short stories in which direct and indirect speech utterances were spoken monotonously. This manipulation preserved (sub)-segmental acoustic information such as the phonological information of the uttered words in the recordings, but minimized supra-segmental acoustic information such as the global intonation patterns over the utterances. Thus, if direct speech is represented in enriched supra-segmental acoustic representations of voices during language comprehension, individuals would have to mentally simulate such representations, since they are minimized in the stimuli, to supplement what they hear while listening to monotonously spoken direct speech utterances. By contrast, indirect speech is not represented in vivid voice representations and hence individuals need not simulate supra-segmental acoustic information when listening to monotonous indirect speech utterances. Thus, we predict that listening to monotonous direct speech quotations would elicit higher “top-down” brain activity in temporal voice areas of the right auditory cortex (i.e., similar to the brain areas identified in Yao et al., 2011) than listening to equally monotonous, meaning-equivalent indirect speech quotations.

To test this hypothesis, we employed functional magnetic resonance imaging (fMRI) to measure participants' brain activity while they were listening to short stories which contained monotonously spoken direct or indirect speech utterances. Between the two conditions, we compared the evoked BOLD signal changes within participants' temporal voice areas. Moreover, we performed multiple parametric modulation analyses to verify the underlying source of any differential brain activity we observed. To assess consistency of results across studies, we also compared the observed activation patterns with our previous silent reading data (Yao et al., 2011).

Materials and methods

Participants

Twenty-one adult participants were recruited and scanned. They were native English speakers with normal hearing and language abilities, and with no history of neurological or psychiatric disorders. Three participants had to be excluded from analysis due to either (a) no clear response in the voice localiser task and/or excessive head-movements during scanning (2 subjects), or (b) scanning abortion following claustrophobic symptoms (1 subject). Data from the remaining 18 participants (age 18–32 years, 9 males and 9 females) were valid for the final analyses. All of them were right-handed except for one female subject. They signed a consent form and were paid at £6/h for their participation.

Stimuli

Main stimuli

Ninety short stories with different protagonists (indicated by different names) were recorded as stimuli. The stories were exactly the same as in Yao et al. (2011). Transcriptions are available at www.psy.gla.ac.uk/~christop/JOCN_2011/Stimuli.pdf. Each story started with two declarative sentences to set up a scenario (e.g., *Luke and his friends were watching a movie at the cinema. Luke wasn't particularly keen on romantic comedies, and he was complaining a lot after the film.*), followed by either a direct speech or an indirect speech quotation sentence (e.g., *He said: “God, that movie was terrible! I've never been so bored in my life.”* or *He said that the movie was terrible and that he had never been so bored in his life.*). The reported clauses in both conditions (underscored in the above examples) were equivalent in terms of linguistic content. Additional comprehension

questions were also recorded for 23 stories (ca. 25%) to assess participants' overall comprehension accuracy and to ensure that they read the stories attentively.

The stories and questions were spoken by a professional actress. Critically, in one condition, the direct speech utterances were deliberately spoken as monotonously (*Direct-monotonous* condition) as the indirect speech utterances (*Indirect-monotonous* condition), i.e., without providing vivid depictions of the reported speaker's voice. We also recorded “normal” (i.e., vivid) versions of the direct speech utterances which were used as a control condition (*Direct-vivid* condition). Example recordings are available at: www.psy.gla.ac.uk/~boy/fMRI/sampler recordings/.

Three lists of stimuli with counterbalanced item-condition combinations (i.e., 30 Direct-monotonous trials, 30 Indirect-monotonous trials, and 30 Direct-vivid trials per list) were constructed using a Latin square. Each item appeared once per list, but in a different condition across lists. Each list was assigned to one third of our participants. The presentation order of the items per list was randomised for each participant.

Voice localizer stimuli

For the voice localizer session (see Procedure), we presented blocks of vocal sounds and non-vocal sounds provided by the Voice Neurocognition Laboratory (vnl.psy.gla.ac.uk), University of Glasgow. These stimuli were the same as those employed in Belin et al. (2000), and comprised both speech (e.g., spoken vowels) and non-speech (e.g., laughing and coughing) vocal sound clips, as well as non-vocal sound clips (e.g., telephone ringing and dog barking). The contrast in brain activity elicited by vocal versus non-vocal sounds reliably localizes temporal voice areas of the auditory cortex.

Procedure

Participants were positioned in the scanner, wearing MRI-compatible, electrostatic headphones (NordicNeuroLab, Norway) for (1) auditory presentation during both the story listening session and voice localizer session and (2) noise attenuation during fMRI scanning. For the story listening session, participants were instructed to keep their eyes closed, to listen to the stories carefully and to answer comprehension questions which would follow 25% of the short stories they had heard. The stimuli were presented using E-Prime 2.0 (Psychology Software Tools, Inc., USA); each trial started with a 4-second silence period, followed by the presentation of the story and then (in 25% of the trials) a comprehension question regarding the content of the preceding story. Each such question appeared 1 s after termination of the preceding story presentation and prompted a “yes” or “no” response which participants could provide by pressing buttons on a response box with their index or middle fingers, respectively. The 90 listening trials were evenly interspersed with five 30-second “baseline” trials during which no experimental stimulation was present.

After the story listening session, an anatomical scan of the participant's brain was performed, followed by a brief (ca. 10-min) voice localizer scanning session. During the latter, participants were instructed to close their eyes while listening to 20 8-sec blocks of vocal and 20 8-sec blocks of non-vocal auditory stimuli presented in an efficiency optimised, pseudo random order along with 20 8-sec blocks without stimulation, acting as a baseline (cf. Belin et al., 2000).

MRI acquisition

Scanning was performed on a 3-T Siemens Tim Trio MRI scanner using a 12-channel head coil (Erlangen, Germany). Functional scans (for both the story listening session and voice localizer session) were acquired using a T2*-weighted echoplanar imaging (EPI) sequence (32 slices acquired in orientation of the Sylvian fissure;

Download English Version:

<https://daneshyari.com/en/article/6031998>

Download Persian Version:

<https://daneshyari.com/article/6031998>

[Daneshyari.com](https://daneshyari.com)