The emotional consequences of hearing inner speech during silent reading of direct speech quotations

A thesis submitted to The University of Manchester for the degree of Master of Philosophy in the Faculty of Biology, Medicine and Health

2021

School of Biological Sciences

Division of Neuroscience and Experimental Psychology

Hongrui Lin

Table of Contents

ABSTRACT **3**

DECLARATION **4**

COPYRIGHT STATEMENT **5**

GENERAL INTRODUCTION **6**

Inner Speech and Overt Speech *7*
Inner Speech in Silent Reading *10*
Inner Speech when Reading Direct Quotes *12*
Inner Speech and Emotional Prosody *14*

THE CURRENT PROJECT **15**

EXPERIMENT 1 19
Participants *20*
Stimuli *21*
Procedure *23*
Results and Discussion *24*
EXPERIMENT 2 30
Participants *30*
Stimuli *31*
Procedure *32*
Results and Discussion *33*
EXPERIMENT 3A 39
Participants *40*
Stimuli *40*
Procedure *42*
Results and discussion *42*
EXPERIMENT 3B 45
Participants *46*
Stimuli *46*
Procedure *46*
Results and discussion *47*
EXPERIMENT 3C 50
Participants *50*
Stimuli *51*
Procedure *51*
Results and discussions *51*

GENERAL DISCUSSION **54**

CONCLUSION **63**

REFERENCES **65**

Word Count: 14162

Abstract

Recently, research has shown that people produce inner speech – a mental simulation of overt speech – in silent reading of, especially, direct speech quotations. Among those studies, some have reported neural overlaps (activations of the right temporal voice areas) between inner speech and emotional prosody, suggesting that inner speech may have emotional consequences. This project explores this possibility by assessing individuals' performance on emotional judgements during silent reading. Participants were asked to judge categories of emotions (to choose one out of six facial expressions or six emotion words) for the reported speakers in the first experiments, while the latter three experiments indicated readers to judge the intensity of emotions by giving emotional arousal ratings. The results demonstrated that readers were faster to make emotional judgements when reading quotations in direct speech than in indirect speech. This 'quantitative' effect supports the hypothesis that hearing inner speech in direct speech conditions helps readers to access the reported speakers' emotional states. However, the findings of the 'qualitative' effects captured by making more appropriate emotional judgements were mixed, suggesting the effects may be more complex than increasing the intensity of emotional activations. Several limitations, such as the experimental setup, were discussed. Future studies are suggested to confirm that the advantage of response time in direct speech conditions is due to inner speech and explore the 'qualitative' effects on emotional judgements of hearing it.

Keywords: inner speech, inner voice, emotional prosody, direct speech, silent reading

Declaration

The MPhil candidate hereby declares that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

General Introduction

In communication, people decode other's feelings, attitudes, and intentions via non-verbal cues such as speech prosody (i.e., the rhythms, stresses, and intonation of speech). Communicating in written language, conversely, may lead to misunderstandings due to its lack of speech prosody. However, recent evidence has shown that people actually produce "inner speech" – a mental simulation of overt speech – even in silent reading, particularly with regards to direct speech quotations (Stites et al., 2013; Yao et al., 2011, 2012). This interesting form of speech has attracted considerable attention in various fields since Vygotsky (1934/1987) first tried to understand it systematically (see Alderson-Day & Fernyhough, 2015 for a recent review). However, the reasons why inner speech appears remain unclear. One way to explore them is by comparing overt speech, which is similar to inner speech in many ways (e.g., Alderson-Day et al., 2020; Alexander & Nygaard, 2008; Oppenheim & Dell, 2008; Watson, 1913; Yao & Scheepers, 2011). Evidence from spoken language, though, has proved that overt speech plays an important role in communicating and understanding emotions (e.g., emotional prosody indicates the speaker's emotions by changing its vocal features; Banse & Scherer, 1996). Therefore, it is feasible that inner speech leads to emotional consequences that help readers to understand the emotional aspects of the texts. The present project will explore this possibility.

In this report, I will first review the relationship between inner and overt speech before focusing on inner speech in silent reading, especially of direct speech quotations. I will then explore the evidence on the possible emotional quality of inner speech. Finally, I will attempt to address whether inner speech conveys emotions in several experiments,

followed by my plan for future work. Please note that terms like covert speech, inner voice, verbal thinking, or auditory verbal imagery have been used interchangeably in the literature. Therefore, this report will use ‘inner speech’ as a general term for the silent production of words or the auditory experience in one’s mind.

Inner Speech and Overt Speech

As one of the early attempts to understand inner speech, Vygotsky’s (1934/1987) theory has assumed that inner speech is the “internalised form” of spoken language (from social speech to private speech to inner speech). Vygotsky viewed this developmental process as crucial to cognitive development. In 1913, Watson proposed a “motor simulation” view, arguing that inner speech is a “weakened form of overt speech”. This view suggested that inner speech is essentially the same as overt speech, only with less articulation involved.

Both theories assume that inner speech is grounded in overt speech and should share many features with overt speech. Indeed, recent evidence shows that inner speech can vary in tempo (Alexander & Nygaard, 2008; Kurby et al., 2009; Yao & Scheepers, 2011) see the following section). Inner speech has also been proved to affect the perceived loudness of sound, which is another acoustic feature of overt speech (Tian et al., 2018). In addition to acoustic features, researchers believed that inner speech and overt speech shared linguistic and structural features (e.g., Brocklehurst & Corley, 2009; Corley et al., 2011; Filik & Barber, 2011; Oppenheim & Dell, 2008). Research studying errors and delays in producing inner speech (in tongue-twister recitation tasks) has shown similar patterns of lexical bias (the tendency to create real words rather than nonwords during tongue-twister recitation)

and phonemic similarity effects (the tendency to substitute similar phonemes, e.g., *patch* and *batch*) with overt speech (e.g., Corley et al., 2011; Oppenheim & Dell, 2008; Yao, 2021).

The shared features between inner speech and overt speech suggest that they share a common processing mechanism. A recent case study (Vercueil & Perronne-Bertolotti, 2013) reported a patient with jargon aphasia (a type of aphasia in which an individual produces incomprehensible speech). Apart from producing overt jargon, this patient also reported her experience of jargon-like inner speech. Her jargon involving both inner and overt speech led the authors to suggest that inner and overt speech may engage the same speech production system.

Another source of evidence for the similarities between inner and overt speech comes from neuroimaging studies. When comparing silent and oral reading tasks, inner speech and overt speech both activate Broca's and Wernicke's areas in the left hemisphere - areas that are linked to speech production and perception (see Price, 2012 for a review). Yao and colleagues (2011) also observed that inner speech in silent reading of direct speech quotations elicited increased neural activity in parts of the auditory cortex. Specifically, they reported right temporal voice areas (TVAs) that were thought to be active only during voice perception. This direct speech effect of direct speech style was later replicated by Alderson-Day et al. (2020).

Despite many similarities between inner speech and overt speech, inner speech appears to be more than overt speech without a motor component (Geva et al., 2011). In 1947, Smith and colleagues treated a healthy participant with curare (a poison that

effectively paralyses the muscles of the body). Remarkably, although the participant lost the ability to speak, he remained able to understand speech and to answer questions by contracting his left eyebrow. This case indicated that thought, and inner speech, can be produced without articulation (Perrone-Bertolotti et al., 2014). More recently, some researchers have reported discrepancies between inner and overt speech and argued that solely relying on overt speech to understand inner speech might result in misleading conclusions (Geva et al., 2011). Oppenheim and Dell (2008) carried out a study comparing tongue-twister errors between inner and overt speech. Lexical bias was observed in both inner and overt speech tasks, but the phonemic similarity effect was only evident in overt speech. The results showed that inner speech might have impoverished phonemic, but not lexical, representations. However, Brocklehurst and Corley (2009) did observe phonemic similarity effects in both inner and overt speech. This suggests that inner speech may be heterogeneous and more complex than a "weakened form of overt speech".

Indeed, some studies have found overall greater brain activations in inner speech than in overt speech. For example, Shuster and Lemieux (2005) observed greater activations of left middle frontal gyrus, right postcentral gyrus, and right cerebellum in inner speech relative to overt speech. Using lesion analysis, Geva and colleagues (2011) investigated cerebral correlates of inner speech in patients with chronic post-stroke aphasia. These patients performed both inner speech tasks (rhyme and homophone judgements) and overt speech tasks (oral reading tasks). Using voxel-based lesion-symptom mapping (a method for analysing relationships between lesion sites and behavioural deficits on a voxel-by-voxel basis), performance in inner speech tasks (relative to overt speech tasks) was correlated with

lesions in left inferior frontal gyrus and white matter regions adjacent to left supramarginal gyrus. This suggests that those brain areas may play more important roles in inner speech than overt speech, whilst regions such as bilateral motor and premotor cortex (Palmer et al., 2001) appeared to be more important for overt speech than inner speech.

Taken together, current literature has reported a close but complex relationship between inner and overt speech. Inner speech shares some of the acoustic, linguistic, and structural features with overt speech. Yet, the precise involvement of the motor system in inner speech remains debatable. This may be partly due to the various tasks that have been used to elicit inner speech. Some tasks involved explicit phonemic judgements (e.g., Filik & Barber, 2011; Geva et al., 2011), which may introduce additional task-driven effects on the motor system. Other tasks attempted to induce inner speech ecologically, which may depend less on covert articulation (e.g., Oppenheim & Dell, 2008; Shuster & Lemieux, 2005). One such task is silent reading.

Inner Speech in Silent Reading

Hearing inner speech during silent reading is a common experience. Huey (1908) argued that inner speech is important for learning to read and is more prominent when reading unfamiliar words. Empirical research has demonstrated inner speech in silent reading via syllable-length/phonetic-length effects and visual tongue-twister effects (Abramson & Goldinger, 1997; Hanson et al., 1991; Klapp, 1971; McCutchen & Perfetti, 1982). For instance, Abramson and Goldinger (1997) examined the effects of phonetic length variations, namely vowel and word-initial consonant length, in visual word recognition. They found that phonetically longer words (e.g., *bad*) were recognised more

slowly than phonetically shorter words (e.g., *bat*). Since these words were matched in a range of lexical properties such as word frequency and word length, the differences in response times could only be explained by phonetic lengths, i.e., how long it would take to pronounce these words aloud. The results suggest that phonological representations of words are not abstract. They are best characterised as inner speech as they share perceptual features (phonetic lengths) with overt speech. In an eye-tracking study, Ashby and Clifton (2005) examined the effects of lexical stress (e.g., *re**cord*** vs. ***re**cord*), as an auditory feature of spoken language, in silent reading. Lexical stresses are pronounced longer than unstressed syllables in speech. They observed a similar lexical stress effect in silent reading – readers spent more time reading words with two stressed syllables than words with one stressed syllable when other linguistic variables were controlled for. This is another example of inner speech in silent reading. It supports the implicit prosody hypothesis that people mentally construct a prosodic contour during silent reading Fodor (1998).

Not only does inner speech share phonological features with overt speech, but it could also contain speaker-specific information (Alexander & Nygaard, 2008; Filik & Barber, 2011; Kurby et al., 2009). For example, Alexander and Nygaard (2008) first played audio-recorded conversations to participants. Participants familiarised themselves with two talkers' voices in fast vs. slow speaking rates. They then asked the participants to silently read text materials that were supposedly written by either the fast or the slow talker. The researchers found that it took participants longer to read the same materials when they believed the materials were written by the slow talker rather than the fast talker. The results suggest that readers may engage in a type of auditory imagery (i.e., inner speech) that

encodes talkers' speaking rates or, more generally, speaker-specific information. However, it should be noted that for easy texts, such effects were only found significantly in oral reading tasks but not in silent reading tasks.

Furthermore, participants with self-reported high imagery skills changed their silent reading rates regardless of text difficulty. Participants with low imagery skills, however, only showed such effects in silent reading of difficult, but not easy, texts. This implies that readers may not always generate talker-specific inner speech during silent reading.

Some have argued that talker-specific inner speech may depend on the priming of specific voices before reading. Such inner speech may be a form of encouraged auditory imagery, as opposed to being an integral part of the silent reading process. However, Yao and Scheepers (2011) demonstrated similar inner speech-related modulations of silent reading rates without encouraging participants to imagine specific voices. Interestingly, the effect was found only in silent reading of direct quotes.

Inner Speech when Reading Direct Quotes

Previous studies have examined semantic and syntactic differences between direct speech (e.g., *David said, "This movie is awesome!"*) and indirect speech (e.g., *David said that the movie was awesome*; Banfield, 1973; Clark & Gerrig, 1990; Coulmas, 1986; Li, 1986; Partee, 1973). Coulmas (1986) defines that direct speech is reported from the reported speaker's perspective while indirect speech is reported from the reporter's own perspective. However, Clark and Gerrig (1990) focused on the distinct pragmatic functions of direct and indirect speech. Direct speech provides vivid demonstrations of *how* a sentence was spoken.

Indirect speech provides descriptions of *what* was spoken. As a stylistic device, direct speech is intended to make stories lively and enable the addressee to experience how it was spoken directly. For instance, reporters may depict the speaker's accent (e.g., North, South, Scottish), tone of voice (e.g., soft, loud), emotional state (e.g., surprise, sadness), and even nonverbal behaviours of speakers (e.g., gestures, facial expressions, postures). Indirect speech, in comparison, provides a mere description of what was said without providing vivid paralinguistic demonstrations like its direct counterpart (Yao & Scheepers, 2015). Taken together, in verbal communication, it is assumed that direct speech is reported in more expressive prosody than indirect speech.

Indeed, Yao (2011) found that readers tended to report direct speech texts in a more vivid way than indirect speech texts when reading aloud. In this study, the author examined whether readers would spontaneously adjust their pitch according to the emotional arousal of the speaker when reading direct or indirect speech texts aloud. The results showed that for direct speech texts, readers increased their voice pitch as the speaker's emotional arousal increased. Such a pattern was not observed when reading aloud indirect speech texts. There was also a larger variation of voice pitch in the direct speech condition than in the indirect speech condition, indicating that direct speech was read in a more varied voice than indirect speech. These findings showed that direct speech is indeed reported in more expressive prosody than indirect speech.

Investigations were then conducted to examine whether this vividness distinction of prosody could be extended to silent reading. As mentioned above, Yao and Scheepers (2011) reported that by manipulating talkers' speaking rates in context, participants' reading rates

were affected during the reading of direct speech quotations, but not indirect quotes. They explained that readers tended to engage in more vivid inner speech for direct speech quotes than for indirect speech quotes. Similarly, using eye-tracking, Stites et al. (2013) reported that readers spent more time on direct speech texts described as being said slowly (e.g., *said idly* or *declared lazily*) than quickly (e.g., *said energetically* or *announced excitedly*), but such effects were not observed on semantically-matched indirect speech texts.

Moreover, neuroimaging studies have observed stronger activations in the right superior temporal sulcus (STS) in the auditory cortex during silent reading of direct speech quotations than indirect speech quotations (e.g., Yao et al., 2011). Importantly, these areas of the STS were found to selectively respond to human voices (i.e., temporal voice areas (TVAs); Belin et al., 2000). Since silent reading does not involve any "bottom-up" auditory stimulation, the activations of TVAs may be best interpreted as "top-down" simulations of voice-related representations (i.e., inner speech).

These findings suggest that readers are more likely to simulate inner speech during silent reading of direct quotations than indirect speech. However, it remains unclear why readers would simulate inner speech in silent reading. What might the consequences of hearing inner speech be? One possibility may be that inner speech could help readers to understand the emotional states of the reported speaker.

Inner Speech and Emotional Prosody

Inner speech might have the capacity to convey emotions through certain perceptual features. In spoken communication, emotional prosody signals the speaker's emotional

states via modulations of vocal features such as loudness, pitch, and tempo (Banse & Scherer, 1996). For instance, “Joy” is characterised by increases in pitch, loudness, and tempo, whereas “Sadness” can be signalled by a reduction in these variables. Since inner speech can be characterised in acoustic features such as tempo (e.g., Alexander & Nygaard, 2008), it may also signal emotions like overt speech.

Moreover, Yao et al. (2011) has shown that inner speech in silent reading of direct quotes is associated with activation within the right TVA. Such a right-lateralised activation pattern has been suggested to play a role in decoding emotional prosody (Ethofer et al., 2009). The right STS has also been implicated in models of emotional prosody processing (e.g., Brück et al., 2011; Schirmer & Kotz, 2006). In both models, bilateral mid superior temporal cortex (m-STC) has been assumed to play a part in the extraction of acoustic features of prosodic information. These features are then integrated into emotionally significant prosodic representations in the right posterior STS (Brück et al., 2011) or the right anterior STS (Schirmer & Kotz, 2006) before reaching the frontal regions for emotional appraisal and/or amygdala for emotional responses. The neural overlaps between inner speech and emotional prosody imply that inner speech may have emotional consequences, which is the focus of the current project.

The Current Project

This project aims to explore the emotional consequences of inner speech in silent reading of direct speech quotations. The central hypothesis is that with the aid of inner speech, emotional information or representations can be more strongly activated when

people read direct speech quotations. Considering this assumption, inner speech can help readers to access the emotional states of reported speakers, and therefore to make faster and more appropriate emotional judgements.

To test this hypothesis, five experiments were conducted. Participants were instructed to silently read sentences and make emotional judgements. The first two experiments (Experiment 1 & 2) asked participants to judge categories of emotions and participants in the latter experiments (Experiment 3a, 3b, and 3c) to judge the intensity of emotions. Based on previous research on vocal emotions, pitch, loudness, and tempo all contribute to vocal emotions (e.g., Banse & Scherer, 1996). However, in vocal emotion communication, some researchers emphasised more on the contribution of vocal loudness or intensity (e.g., Pittam & Scherer, 1993). Vocal intensity can influence acoustic cues (e.g., higher intensity generates acoustic cues for intense emotions than less intense emotions); and was most often reported by people as factors to sense emotions of others in daily life (Juslin & Laukka, 2001; Planalp, 1996). Therefore, the first two experiments introduced vocal intensity as emotional cues to ideally affect inner speech.

Both experiments tried to manipulate inner speech by using different verbs that implied different levels of vocal intensity (e.g., ‘shouted’ signals high intensity, ‘whispered’ signals low intensity, ‘said’ signals neutral intensity). In Experiment 1, participants silently read direct speech, indirect speech, and narrative sentences. After each sentence, they were asked to choose one out of six facial expressions that best depicts the emotional state of the protagonist in the sentence. This paradigm was motivated by research on the interaction of emotional prosody and emotional faces. In our daily communication,

we often integrate information from vocal and visual channels to make effective emotional judgements. Early research has proposed a shared processing mechanism for vocal and visual emotional information and demonstrated that emotional prosody could influence the processing of facial expressions (de Gelder & Vroomen, 2000; Gelder et al., 2006). Recent event-related potential (ERP) studies (e.g., Paulmann & Pell, 2010) have also provided neurophysiological evidence that emotional prosody and facial expressions are strongly associated. I, therefore, reasoned that inner speech, or covert emotional prosody, can also influence face processing and how readers choose facial expressions. Specifically, I predicted that individuals might make faster and more consistent (i.e., tend to choose the same facial expression for each sentence) judgements during silent reading of direct rather than indirect speech. As for the effects of vocal intensity on response times and consistency, I expected that there would be larger differences in both factors between direct speech and indirect speech conditions when reading high- or low-intensity sentences than neutral sentences. This was on account of the possible uncertainty from more applicable choices for high- and low- intensity sentences (e.g., individuals may understand *"Stacey looked at him and barked, 'The car is blue'"* as angry or fearful, both of which are characterised with high intensity as suggested by Banse & Scherer (1996), while neutral facial expression was provided for them when reading neutral sentences).

Experiment 2 replicated the paradigm in Experiment 1 but replaced facial expressions with emotion words. This was because, compared with choosing emotion words, choosing facial expressions may introduce more variations in response times, as processing facial expressions needs extra time. Hence, Experiment 2 remained the same hypotheses as

for Experiment 1, but it might reveal stronger effects than observed or new effects that were not found in the former experiment.

Experiment 3a, 3b, and 3c examined participants' ratings of emotional intensity on neutral quotations. Different from the first series of experiments, only reporting styles were manipulated. Although these experiments kept reported verbs as neutral, the sentences themselves could be ambiguous because little context was provided (e.g., *Holding the map, David said, "We need to go left"* or *Putting down her book, she replied that she heard that story last week*). Therefore, participants had to make their judgements by detecting limited emotional information when reading these sentences. I hypothesised that with the help of inner speech, readers would have stronger emotional information when they read direct speech quotations, as compared to indirect speech quotations. In these experiments, participants were asked to rate "how emotional the protagonist is". By asking this question, I tested participants' perceptions of emotional arousal (i.e., emotional intensity) of the protagonists. Ideally, Experiment 3a would establish advantaged effects in the direct speech condition, captured by responses times and perceived levels of emotional arousal (i.e., faster response times and higher emotional ratings). To further examine whether the effect was driven by inner speech or other non-speech factors, the last two experiments introduced secondary tasks (articulatory suppression tasks vs. finger tapping tasks). The rationale was that articulatory suppression tasks blocked inner speech (and its effects on participants' performance), which consequently masked the difference between direct speech and indirect speech conditions. The difference of design between Experiment 3b and 3c was the difficulty of the interference tasks (verbally repeating "1, 2, 3, 4" vs. "bah, bah, bah…";

four-finger tapping vs. one-finger tapping). For both experiments, I predicted that there would be no differences in responses times or emotional ratings between direct speech and indirect speech conditions.

Experiment 1

In Experiment 1, each participant silently read 42 sentences and chose the most appropriate facial expression for each sentence. Reporting styles (direct speech vs. indirect speech; hereinafter referred to as DS and IS, respectively) and intensity (high, neutral, or low) of the sentences were manipulated. Non-speech (narrative; hereinafter referred to as NS) condition was added as a control group (also neutral). This results in a 2 (Reporting Style, Direct, Indirect) x 3 (Voice Loudness, high, low, neutral) + 1 Control (non-speech) experimental design. The intensity of the sentences was manipulated by the verbs that described how the reported speakers spoke (see Table 1). For neutral sentences, I hypothesised that the DS condition, as compared to the IS and NS conditions, would be associated with faster RTs (a quantitative effect) and most consistent choices of facial expressions (a qualitative effect). I expected that in the DS condition, readers could make emotional judgements based on emotional cues provided by the contexts/verbs and inner speech. The emotional cues provided by inner speech are likely to be consistent with those provided by the contexts, same way as the contextual manipulations on talkers' speed influence the tempo of inner speech accordingly (e.g., Yao & Scheepers, 2011). In the other two conditions, readers would make judgements based on the contexts only. With more emotional information in the DS condition, it would be easier for readers to judge the emotions, and they would be more likely to choose facial expressions that fit the reported

speakers well, as compared with the other two conditions. Therefore, they would respond faster and give more consistent choices in the DS condition. In terms of the effect of intensity on RTs, when readers read high- or low-intensity sentences, they may interpret it as several emotions and thus spend more time on this uncertainty. For example, readers may interpret *"Stacey looked at him and barked, 'The car is blue'"* as angry or fearful, both of which are characterised with high intensity as suggested by Banse and Scherer (1996). Conversely, it would be easier for them to make their choices for neutral sentences because neutral facial expressions were provided. Moreover, literature has shown that in spoken language, speech prosody can help listeners recognise emotions independent of semantic cues (Pell et al., 2011). Similarly, I predicted that inner speech would facilitate emotional processing. Therefore, it would reduce the uncertainty of high- or low-intensity sentences in DS, leading to larger differences in both response times and consistency between DS and IS conditions, as compared with neutral sentences.

Participants

A statistical power analysis was performed for sample size estimation, based on a hypothetical medium effect size (i.e., Cohen's $d = 0.5$; Cohen, 1988). With an $\alpha = 0.05$ and power = 0.8, G*Power (Faul et al., 2009) generated the sample size required with this effect size was 42. Therefore, 42 participants were recruited in this experiment. All 42 (19 male and 23 female) participants were native English speakers with no reported learning/reading/speaking disabilities. Participants averaged 30.2 years in age (SD = 12.0 years). The experiment lasted approximately 15 minutes. After the experiments, participants either were paid £2 or were awarded a course credit for their participation as compensation.

The experiment was approved by the ethics committee of the University of Manchester (Research ethics code: 16248).

Stimuli

42 scenarios were used in this experiment (see below example sentences). Character (i.e., the reported speakers) names were different across scenarios. For each scenario, sentences were written in different reporting styles (DS vs. IS) and different intensity (high, neutral, or low). There were seven conditions in total for each scenario, including the NS condition as a control group (42×7 sentences in total). The quoted utterances (underlined sentences) within each item were identical among high, low, and neutral conditions and were virtually the same between DS and IS conditions. All the scenarios and conditions were combined, and seven stimulus lists were constructed (six scenarios per condition per list). In each list, the order of the sentences was randomised. Each list was randomly allocated to six participants.

Intensity	RS	Sentence
High	DS	*Turning towards her friend, Carol **screamed**, "I already know that"*
	IS	*Turning towards her friend, Carol **screamed** that she already knew that.*
Low	DS	*Turning towards her friend, Carol **sighed**, "I already know that"*
	IS	*Turning towards her friend, Carol **sighed** that she already knew that.*
Neutral	DS	*Turning towards her friend, Carol **replied**, "I already know that"*
	IS	*Turning towards her friend, Carol **replied** that she already knew that.*
	NS	*Turning towards her friend, Carol **realised** that she already knew that.*

Table 1. Example sentences. RS = reporting style.

Twelve facial stimuli were taken from a validated database (Paulmann & Pell, 2009; Pell, 2005; Pell, 2005; Pell, 2002). All facial stimuli were static, colour photographs of facial expressions posed by actors (half by a male actor and half by a female). Each actor posed six facial expressions, including angry, fearful, happy, neutral, sad, and surprised. The facial stimuli were cropped into 200mm×263mm and were grouped into two sets (Figure 1A). They were paired with each sentence according to the character's gender and were presented below the sentences.

Figure 1. Female and male sets of facial stimuli. Each set includes angry, fearful, happy, neutral, sad, and surprised in sequence.

Procedure

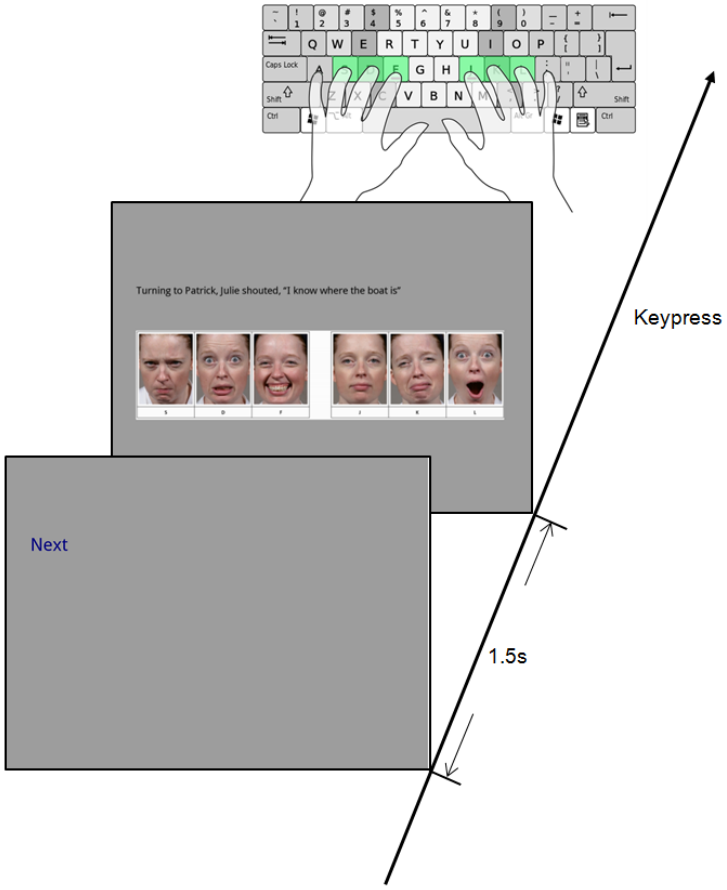

Figure 2. Experimental procedures (Experiment 1).

The experiment was conducted in a quiet laboratory setting. All participants gave informed consent before the experiment. They were then seated in front of a computer monitor, which presented the experiment using OpenSesame 3.1 (Mathôt, Schreij, & Theeuwes, 2012). The experiment began with seven practice tests before the actual tests. In both tests, each trial started with a visual cue ‘Next’ for 1000 ms, followed by a sentence and a set of six facial expressions underneath the sentence. Participants were instructed to read the sentence silently, with their fingers placed on six keys on the keyboard, respectively. Their task was to choose a facial expression that best depicted the protagonist’s emotional state by pressing a corresponding key (i.e., left hand: the index finger on the F, the middle

finger on the D, and the ring finger on the S; right hand: the index finger on the J, the middle finger on the K, and the ring finger on the L). Each button press triggered the next trial (Figure 1B), and their choice and response time were recorded. To control for the influence of handedness, for half of the participants, the order of the facial expressions was angry, fearful, happy, neutral, sad, and surprised; for the other half, the order was mirrored, that is, surprised, sad, neutral, happy, fearful, and angry.

Results and Discussion

RTs were trimmed in two steps: First, based on inspection of the RT distribution, RTs longer than 15000 ms were eliminated from the data, which removed certain extreme values. Next, RTs longer or shorter than three standard deviations from the mean RT of each condition were eliminated from the rest. This removed 3.3% of the total values. The remaining RTs were analysed in generalised linear mixed-effects models (GLMMs) using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2017). Two models were constructed. To examine the effects of intensity and the interaction between intensity and reporting style, I fitted a 2 x 3 GLMM (Model 1). The two reporting style conditions were deviation-coded into a DS-IS contrast (DS = 0.5, IS = -0.5). The three intensity conditions were dummy-coded into two contrasts: the contrast between high intensity and neutral (H-N) and the contrast between low intensity and neutral (L-N). A maximal random-effect structure was employed (Barr et al., 2013). To examine whether emotional activation is also enhanced in IS but to a lesser extent than DS, I compare emotional activations in DS, IS, and NS conditions in emotionally neutral contexts (Model 2). The fixed-effect structure consisted of two dummy-coded contrasts: the contrast between the DS and the IS conditions

(DS-IS contrast) and the contrast between the DS and the NS conditions (DS-NS contrast). Again, a maximal random-effect structure was employed, with *subject* and *item* as crossed random factors. All the *p* values were obtained using *the lmerTest* package (Kuznetsova et al., 2017) in R.

The results of Model 1 showed that participants made faster judgements in the neutral conditions than in the high intensity or in the low-intensity conditions (both $p < 0.001$; Table 2 & Figure 3). An interaction was observed – the RT's difference between the DS and the IS conditions was larger in the high-intensity condition than in the neutral condition. However, no interaction was found in the low-intensity condition. Model 2 discovered that participants responded significantly faster to DS sentences than IS or NS sentences (both $p < 0.001$; Table 3). The post hoc test showed that readers also made their decisions faster when reading IS than NS sentences ($p < 0.001$; Table 3). Effect sizes (i.e., Cohen's *d*) were also calculated for each effect (Table 2 & 3).

Fixed effects	**Estimate**	**S.E.**	***T***	***p***	**Cohen's *d***
DS-IS	-324.82	10.21	-31.82	**<0.001**	0.277
L-N	272.20	15.94	17.08	**<0.001**	0.229
H-N	295.39	13.25	22.29	**<0.001**	0.226
DS-IS × L-N	3.14	11.68	0.27	0.79	0.002
DS-IS × H-N	-155.45	16.18	-9.61	**<0.001**	0.094

Table 2. The GLMM's estimates of RT (Model 1, Experiment 1). L = Low; N = Neutral; H = High.

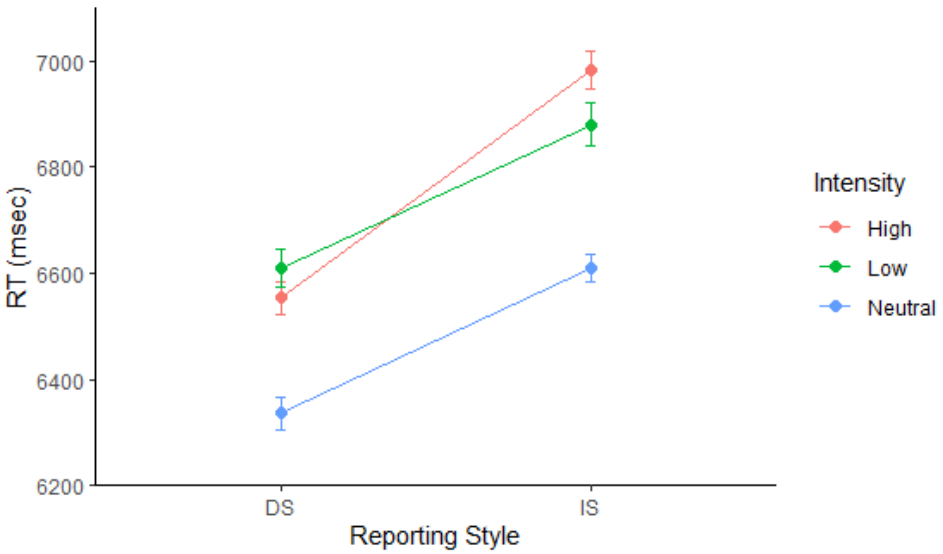

Figure 3. The mean RTs by condition (Experiment 1). The error bars represent the 95% confidence intervals for the means per condition.

Fixed effects	**Estimate**	**S.E.**	***T***	***p***	**Cohen's *d***
IS-DS	306.97	22.38	13.72	**<0.001**	0.229
NS-DS	535.61	22.27	24.05	**<0.001**	0.359
NS-IS*	228.84	33.53	6.83	**<0.001**	N/A

Table 3. The GLMM's estimates of RTs (Model 2, Experiment 1). IS-DS: the contrast between the IS and the DS conditions; NS-DS: the contrast between the NS and the DS conditions; NS-IS: the contrast between the NS and the IS conditions, generated from post hoc tests using *pairs* function.

To examine the consistency of the responses, the below analysis counted how many types of responses were elicited for each condition and generated a SD of the percentages of responses over six categories. These SDs reflect the consistency of the responses – a low SD indicates the numbers of responses for each category (i.e., each facial expression) tend to be closed (i.e., every category has its votes), therefore inconsistent; a high SD, in contrast, suggests the choices are consistent. To examine whether consistency advantages are

available in the DS condition, SDs for each condition were coded in a LMM (Model 3). Same as Model 2, the fixed-effect structure included main effects of H-N, L-N, and DS-IS and interactions (H-N x DS-IS and L-N x DS-IS). The model employed a maximum random-effect structure, and *p* values were estimated by *the lmerTest* package (Kuznetsova et al., 2017) in R.

As shown in Table 4, only the main effect of reporting style on the consistency of responses was found ($p = 0.02$), indicating a lower SD (i.e., less consistent) for DS than IS sentences. There was no effect of intensity or significant interaction between reporting style and intensity. Effect sizes were calculated for each effect (Table 4).

Fixed effects	**Estimate**	**S.E.**	***T***	***p***	**Cohen's *d***
DS-IS	-0.10	0.04	-2.43	**0.02**	0.290
L-N	0.01	0.05	0.21	0.84	0.029
H-N	-0.07	0.05	-1.55	0.12	0.219
DS-IS × L-N	-0.06	0.09	-0.64	0.52	0.182
DS-IS × H-N	-0.02	0.09	-0.17	0.87	0.047

Table 4. The LMM's estimates of SDs (Model 3, Experiment 1). L = Low; N = Neutral; H = High.

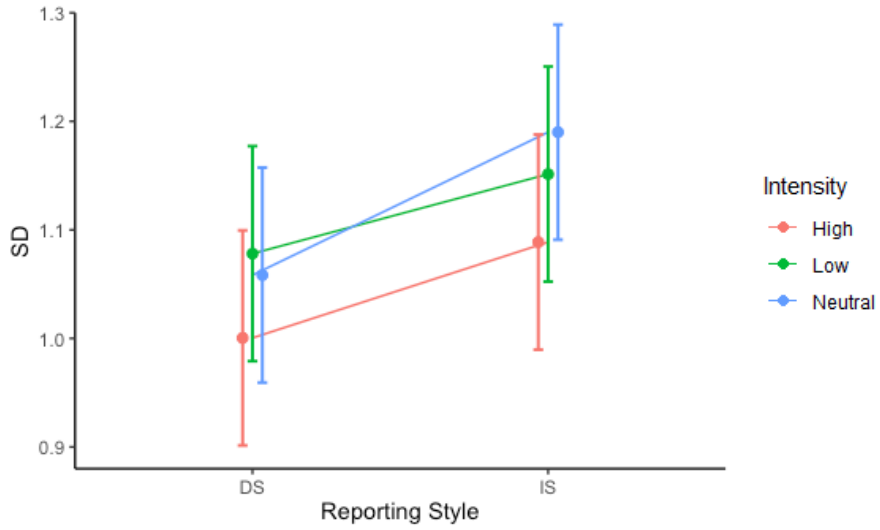

Figure 4. The mean SDs by condition (Experiment 1). The error bars represent the 95% confidence intervals for the means by condition.

Taken together, this experiment found that participants were faster to choose a facial expression after reading direct speech than indirect speech or narrative sentences. It supports our hypothesis that hearing inner speech in silent reading of direct quotes helps readers to access the emotional states of the protagonist, resulting in faster response times. The experiment also observed an RT's advantage for neutral sentences than those in high- or low- intensity as expected. This implies that in this experimental paradigm, readers can respond faster for neutral sentences when a neutral facial expression is provided but spend more time reacting to high- or low-intensity sentences due to more possible interpretations. As previously mentioned, high intensity can signal many different types of emotions, whereas neutral intensity is likely to signal neutral emotion only (Banse & Scherer, 1996). Therefore, there would be more uncertainty when individuals made decisions for high- or low- intensity. This was supported by the interaction observed – reporting style tended to

have a larger effect on RTs in the high-intensity condition than in the neutral condition. This might be attributed to inner speech's help in reducing uncertainty in the high-intensity condition and thus supports the prediction that inner speech facilitates emotional processing.

The analysis of participants' responses showed less consistency for quotations in DS than in IS, which is opposite to the expectation above. No other qualitative effect was found. This could be explained by how the experiment was set up. Since all the quotations were neutral, they did not contain any emotional information. The emotional judgements therefore can only be determined by the contexts and possibly inner speech that participants heard in DS condition. Theoretically speaking, because the direct quotations themselves are neutral in content, they might elicit inner speech that is in neutral prosody. Such inner speech could likely be ambiguous for making emotional judgements. Therefore, with more emotional information, participants can make their judgements faster but might give a wider range of answers, regardless of the intensity of sentences.

Experiment 1 raised a few theoretical and methodological issues. First, this experiment was motivated by studies on the integration of emotional prosody and facial expressions, which have indicated a shared mechanism between them. However, this does not necessarily mean there would be a similar relationship between 'inner speech prosody' and facial expressions. Second, facial expressions require extra processing time before an emotional judgement can be made. Hence the response times I collected measured not only the time needed to make an emotional judgement but also the extra time required to process facial expressions in the first place. Because the interpretation of facial expressions varies from individual to individual, it inevitably adds noise to the response times. To reduce such

noise, I decided to replace facial expressions with emotion words in Experiment 2. As participants no longer need to interpret facial expressions, the response times should accurately reflect the time needed for emotional judgements. To increase statistical power, I also expanded the list of scenarios in Experiment 2, adding 49 new scenarios to the 42 scenarios used in Experiment 1.

Experiment 2

In Experiment 2, each participant silently read ninety-one sentences and decided which emotional word was the most appropriate for the character. Independent variables were reporting styles (DS vs. IS) and intensity (high, neutral, or low) of the sentences, again leading to a 2×3 + 1 experimental design. Hypotheses were the same as those for Experiment 1.

Participants

Since the main difference between Experiment 1 and 2 was the options for responses (facial expressions to emotion words), Experiment kept the same sample size. All 42 (five male and 37 female) participants were native English speakers with no reported learning/reading/speaking disabilities. Participants were all undergraduate psychology students at the University of Manchester (mean age = 19.3 years and had not participated in Experiment 1). The experiment typically lasted 20 minutes. After the experiments, participants were all given two course credits for their participation. The experiment was approved by the ethics committee of the University of Manchester (Research ethics code: 16248).

Stimuli

Ninety-one scenarios were used in the experiment (49 new scenarios plus the scenarios used in Experiment 1). All the scenarios were manipulated in the same way as Experiment 1 (combining scenarios and conditions), constructing seven stimulus lists.

The facial expressions in Experiment 1 were replaced by the corresponding emotion words (i.e., angry, sad, fearful, neutral, happy, and surprised). They were shown below the sentences.

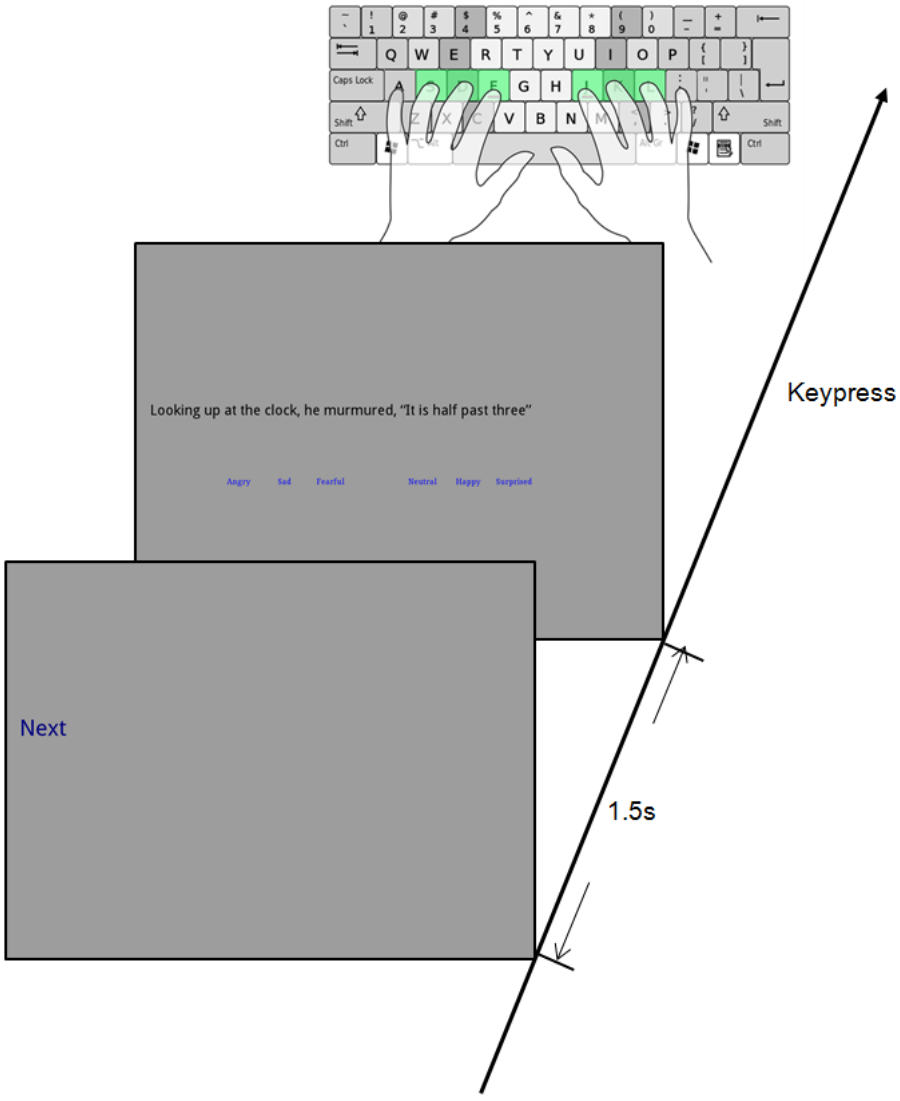

Figure 5. Experimental procedures (Experiment 2).

Procedure

The experiment was conducted in a quiet laboratory setting. All participants gave informed consent before the experiment. They were then seated in front of a computer monitor, which presented the experiment using OpenSesame 3.1 (Mathôt et al., 2012). The experiment began with an association training before the tests. Participants were asked to pass the training to memorise corresponding keys on the keyboard to each emotional word (e.g., ‘S’ for angry, ‘D’ for sad) with accuracy rates higher than 90%. After passing the training, participants started nine practice tests (six sentences were associated with one of the six emotions and the other three were ambiguous as the test sentences; three sentences were DS quotations, three were IS quotations, and the rest three were NS quotations) before the actual tests. In both tests, each trial started with a visual cue ‘Next’ for 1000 ms, followed by a sentence and six emotion words underneath the sentence. There was a break after every 30 trials in the actual test. Same as Experiment 1, participants read the sentence silently, with their fingers placed on six keys on the keyboard. Their task was to choose an emotion word, instead of facial expressions, that best depicted the protagonist’s emotional state by pressing a corresponding key (finger positions were the same as Experiment 1).

Each participant was asked to choose the most appropriate emotion words, instead of facial expressions, for 91 sentences and nine practice sentences. Before the tests, they were asked to pass an association training to memorise corresponding keys to each emotional word (e.g., ‘S’ for angry, ‘D’ for sad). All participants passed the training with an accuracy rate higher than 90%. During the experiment, there was a break after every

30 trials. Each button press triggered the next trial (Figure 4), and their choice and response time were recorded.

Results and Discussion

RTs were trimmed in two steps: same as in Experiment 1, extreme values (i.e., RTs longer than 15000ms) were removed from the data, based on inspection of the RT distribution. RTs longer or shorter than three standard deviations from the mean RT of each condition were then eliminated. The trimming procedure removed 1.9% of the total values. Two GLMMS of the remaining RTs (Model 4 and Model 5) were constructed in the same way as in Experiment 1.

The results of Model 4 revealed that, again, participants were faster to choose emotional words for neutral than high- or low-intensity sentences (both $p < 0.001$; Table 5). Two interactions were observed, implying that the difference of RTs between DS and IS conditions were larger for neutral than high- or low-intensity sentences (both $p < 0.001$; Figure 6). Model 5 replicated the findings of Model 1 in Experiment 2: participants responded faster after reading DS than IS or NS quotations (both $p < 0.001$; Table 5). Effect sizes were calculated for each effect (Table 5 & 6).

Fixed effects	**Estimate**	**S.E.**	***T***	***p***	**Cohen's *d***
DS-IS	-529.40	6.34	-83.57	**<0.001**	0.655
L-N	337.19	6.87	49.07	**<0.001**	0.397
H-N	352.16	8.36	42.11	**<0.001**	0.420
DS-IS × L-N	94.38	6.00	15.72	**<0.001**	0.107
DS-IS × H-N	165.27	10.05	16.45	**<0.001**	0.165

Table 5. The GLMM's estimates of RTs (Model 4, Experiment 2). L = Low; N = Neutral; H = High.

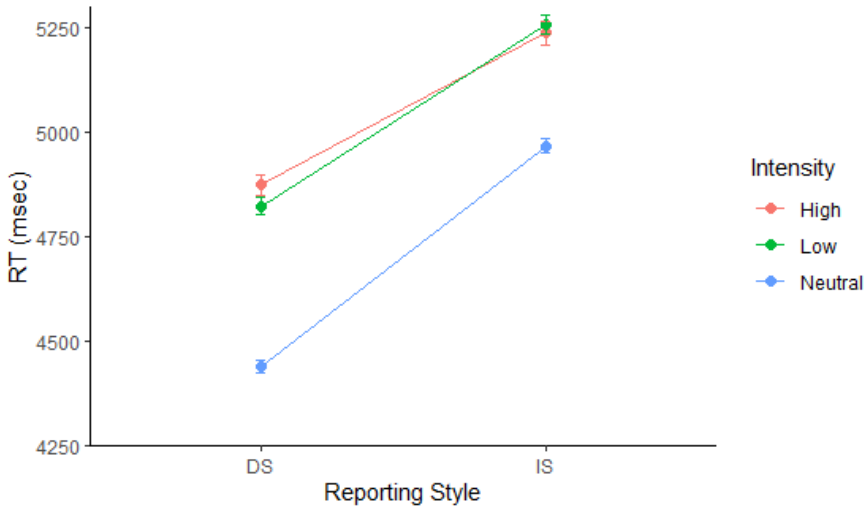

Figure 6. The mean RTs by condition (Experiment 2). The error bars represent the 95% confidence intervals for the means per condition.

Fixed effects	Estimate	S.E.	*T*	*p*	Cohen's *d*
IS-DS	630.23	12.66	49.78	**<0.001**	0.653
NS-DS	768.08	10.24	75.02	**<0.001**	0.787

Table 6. The GLMM's estimates of RTs (Model 5, Experiment 2). IS-DS: the contrast between the IS and the DS conditions; NS-DS: the contrast between the NS and the DS conditions.

To examine whether the assumption that choosing facial expressions adds noise to the response times was true, I compared the results of the two experiments. RTs of Experiment 2 were 2280 ms faster than RTs of Experiment 1 ($p < 0.001$). The SDs of RTs were smaller in Experiment 2 than those in Experiment1 (Table 7). These results supported the assumption.

Intensity	Speech	sdRT (Experiment1)	sdRT (Experiment2)
High	DS	2713	1927
High	IS	2840	2210
Low	DS	2730	2143
Low	IS	2831	2197
Neutral	DS	2747	2015
Neutral	IS	2635	2221
Neutral	NS	2902	2130
Total		2777	2320

Table 7. SDs of RTs by condition in Experiment 1 & 2. sdRT: SD of RTs.

Model 6 explored the consistency of responses by condition. In general, participants gave more consistent choices for IS than DS quotations ($p = 0.01$) and for neutral than high- or low- intensity sentences (both $p < 0.001$; Table 8). Interestingly, there was an interaction – when quotations were neutral, participants gave more consistent choices for IS than DS quotations, but when quotations were low, participants tended to respond in an opposite way (Figure 7). Effect sizes were calculated for each effect (Table 8).

Fixed effects	**Estimate**	**S.E.**	***T***	***p***	**Cohen's *d***
DS-IS	-0.15	0.06	-2.66	**0.01**	0.263
L-N	-0.53	0.07	-7.09	**<0.001**	1.471
H-N	-0.50	0.11	-4.71	**<0.001**	0.821
DS-IS × L-N	0.33	0.14	2.32	**0.02**	1.016
DS-IS × H-N	0.16	0.14	1.11	0.27	0.487

Table 8. The LMM's estimates of SDs (Model 6, Experiment 2). L = Low; N = Neutral; H = High.

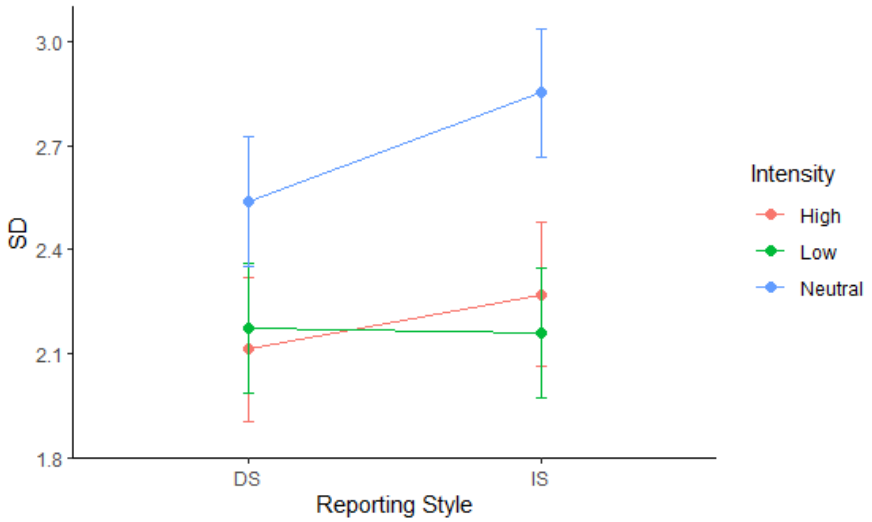

Figure 7. The mean SDs by condition (Experiment 2). The error bars represent the 95% confidence intervals for the means by condition.

Experiment 2 replicated the RT's advantage for the DS condition (i.e., readers were faster to make emotional judgements after reading DS than IS or NS sentences). This result supports a quantitative benefit of hearing inner speech on emotional judgements. The RT's advantage for neutral sentences was also observed as hypothesised. In contrast to Experiment 1, two interactions were found – reporting style tended to have a larger effect on RTs in neutral than high- or low- intensity, supporting the initial hypothesis. This conflict between Experiment 1 and Experiment 2 can be explained by the change of participants' options from choosing facial expressions to emotion words. As mentioned earlier, the paradigm of Experiment 1 was motivated by studies on emotional prosody, which might not necessarily suggest a similar relationship between 'inner speech prosody' and facial expressions. Also, choosing facial expressions adds noise to the RTs (as shown by the larger SDs of the RTs in Experiment 1 than those in Experiment 2), therefore reducing statistical

power, as compared to choosing emotion words. It is likely that choosing emotion words, instead of encouraging participants to imagine what facial expression the protagonist is making, is more eco-valid for silent reading experiments. Therefore, the results of Experiment 2 are more likely to be accurate.

As for the responses, the analysis found more effects in Experiment 2 than in Experiment 1. Same as in Experiment 1, participants' judgements were less consistent for DS than IS quotations, which could be explained by more possible emotional information in the DS condition. The interaction suggested different effects of reporting style on neutral or low-intensity conditions. Individuals made more inconsistent choices in the DS than in the IS conditions when the sentences were neutral, which was the opposite way when they read low-intensity sentences. Following our hypothesis, this means that hearing inner speech in the DS conditions added more uncertainty to readers' minds when reading neutral sentences rather than low-intensity sentences. The possible reason was that when the sentences were ambiguous (as neutral sentences), the inner speech tended to be ambiguous, adding more uncertainty for neutral sentences than low-intensity sentences. Nevertheless, this uncertainty did not influence how fast they made the judgements. That is, hearing inner speech facilitates the speed of emotional activation, but it might add ambiguity to the intensity of emotional activation when reading relevant ambiguous texts (e.g., neutral quotations).

The above two experiments found both possible quantitative and qualitative effects of inner speech in the DS conditions, as compared to the IS conditions. While the

RT's advantage in the DS conditions supports the hypotheses of the benefit of hearing inner speech on emotional judgements, the qualitative effects were not as proposed. To more directly test whether hearing inner speech in reading DS quotations helps readers make emotional judgements, the following three experiments simplified the experimental paradigm. Instead of asking participants to choose appropriate emotional faces or words, Experiment 3a, 3b, & 3c asked them to give emotional ratings (specifically, emotional arousal ratings). These experiments reduced the IV – intensity and kept the quotations neutral. The quantitative effect would be, again, the RTs' advantage in DS conditions than IS conditions; the qualitative effect would be higher emotional ratings for DS than IS quotations. This expectation is based on the idea that hearing inner speech can help readers access the emotional state of the protagonist, hence having more emotional renounce and giving higher emotional ratings.

Experiment 3a

In Experiment 3a, participants were instructed to read one hundred scenarios and rate how emotional the protagonist is in each story. Reporting styles (DS vs. IS + NS) were manipulated. In the DS condition, based on the findings above, I expected that inner speech would help readers access the emotional states of the protagonists and therefore respond faster. By simplifying the paradigm, I believed that the qualitative effects would be observed. That is, individuals would give higher emotional ratings (i.e., intensity ratings) with more information detected from contexts and inner speech in the DS condition, as compared to the other two conditions.

Participants

The sample size was estimated by a statistical power analysis in G*Power (Faul et al., 2009), with a hypothetical medium effect size (Cohen's *d* =0.5), α = 0.05 and power = 0.8. The needed sample size with the effect was 36. However, based on previous studies, I expected a higher effect size and the experiment recruited only 21 participants. Admittedly, the sample didn't reach the benchmark, which would need to be addressed in future studies. All 21 (eight male and 13 female) participants were native English speakers with no reported learning/reading/speaking disabilities. Participants averaged 27.6 in age (SD = 6.4) and had not participated in Experiment 1 or 2. The experiment typically lasted 20 minutes. After the experiments, participants were given two course credits or paid £2 for their participation. The experiment was approved by the ethics committee of the University of Manchester (Research ethics code: 16248).

Stimuli

This experiment presented one hundred scenarios, consisting of 60 critical scenarios and 40 scenarios acting as fillers (see Table 9). Only reporting styles (DS vs. IS) were manipulated in this experiment. NS condition was used as a control group. Instead of using speaking verbs in different intensity (e.g., 'screamed' as high intensity, 'sighed' as low intensity) in previous experiments, Experiment 3a used neutral verbs (e.g., said) only, indicating neutral context, to avoid the issues reported in the previous experiments. However, participants may tend to give low emotional ratings for all the scenarios chronically after scenarios in neutral intensity recurring frequently, consequently narrowing the possible differences between conditions. Therefore, fillers (scenarios with high/low-

intensity contexts and emotional/neutral quotations) are needed to counterbalance participants' emotional ratings. For this purpose, fillers consisted of 30 stories with high-intensity speaking verbs and ten stories with low-intensity speaking verbs (as participants may give low ratings for stories in low-intensity as well). Fillers were also written in different reporting styles (half in DS; half in IS) with quotations in different intensity of emotion (half emotional quotes; half neutral quotes), forming the fillers part of the stimulus lists. There were three stimulus lists, including the part of the 60 critical scenarios that were constructed in a similar way as in the previous experiments and the fillers part. Participants were randomly presented with one of the lists.

	Intensity	**RS**	**Quotations**	**Scenarios**
Critical	Neutral	DS	Neutral	*Holding the map, David said, "We need to go left"*
		IS	Neutral	*Holding the map, David said that they needed to go left.*
		NS	Neutral	*Holding the map, David thought that they needed to go left.*
Fillers	High	DS	Emotional	*Counting the crowd, he yelled, "I'll fire whoever is absent"*
			Neutral	*Mary turned to her sister and roared, "There's no cereal left"*
		IS	Emotional	*As her eyes lit up, she exclaimed that the food was excellent.*
			Neutral	*She turned to Pam and shouted that the cardigan was blue and green.*
	Low	DS	Emotional	*Staring at the couple, Bernard whimpered, "I feel like a loser"*
			Neutral	*Entering the café, Christine murmured, "All the tables are full"*
		IS	Emotional	*Olivia turned to her friend and breathed that he had not passed the exam.*
			Neutral	*Recognising the dress, Liz whispered that it belonged to her daughter.*

Table 9. Example stories.

Procedure

Participants gave informed consent before the experiment in a laboratory setting. The experiment was programmed using OpenSesame 3.1(Mathôt et al., 2012) and presented to participants with a computer monitor. Participants were asked to read short stories silently for comprehension and indicate how emotional the protagonists were in the stories on a four-point scale (0 = not emotional at all, 1 = slightly emotional, 2 = fairly emotional, 3 = highly emotional). They gave the ratings by pressing corresponding number keys on the number pad (right thumb on key 0, index finger on key 1, middle finger on key 2 and ring finger on key 3). Each key press triggered the next trial and their ratings, as well as response times, were recorded. Between trials, there was a 500 milliseconds' break.

Results and discussion

RTs were trimmed before analysis. With the interest of the experiment on inspection of the RTs, RTs longer than 10000 ms were removed from the data because participants were instructed to respond as quickly as possible, and 10000ms would be sufficient for them to read and make subsequent emotional judgements. Next, RTs more than three standard deviations from the mean RT of each participant by condition were deleted. This removed 6.3% of total values. To analyse the effect of reporting style on RTs, Model 7 was constructed using the *lme4* (Bates et al., 2015) package in R. Speech factors (DS, IS, NS) were deviation-coded with the DS as the baseline. The model included a maximal random-effect structure, with *subject* and *item* as crossed random factors (Barr et al., 2013). All the *p* values were calculated using *the lmerTest* package (Kuznetsova et al., 2017) in R. As shown in Table 10, participants responded faster when they read scenarios in DS (4440 ms)

than in IS (4703 ms) or NS (4733 ms). Effect sizes for each effect were presented (Table 10).

Fixed effects	**Estimate**	**S.E.**	***t***	***p***	**Cohen's *d***
IS-DS	263.11	20.91	12.58	**<0.001**	0.299
NS-DS	292.73	13.04	22.44	**<0.001**	0.339

Table 10. The GLMM's estimates of RTs (Model 7, Experiment 3a). IS-DS: the contrast between the IS and the DS conditions; NS-DS: the contrast between the NS and the DS conditions.

A LMM of ratings was fitted to examine whether scenarios in DS were rated more emotional than scenarios in IS or NS. The same deviation coding and model structure were adopted in the LMM of ratings. Results were shown in Table 11. Participants rated DS scenarios higher (0.64) than IS scenarios (0.52). However, the difference between ratings for DS and NS conditions was not observed.

Fixed effects	**Estimate**	**S.E.**	***t***	***p***	**Cohen's *d***
IS-DS	-0.12	0.04	-2.85	**<0.01**	0.148
NS-DS	-0.05	0.06	-0.93	0.36	0.063

Table 11. The LMM's estimates of ratings (Model 8, Experiment 3a). IS-DS: the contrast between the IS and the DS conditions; NS-DS: the contrast between the NS and the DS conditions.

The above findings confirmed the first hypothesis regarding the RTs and partially supported the second hypothesis on the emotional ratings. The result of faster RTs in the

DS condition was consistent with previous experiments. It supported the possibility that hearing inner speech in the DS condition assists people to make faster emotional judgements by accessing the emotional states of the characters in the scenarios. Moreover, participants perceived DS scenarios to be more emotional than IS scenarios. The rating difference suggested that a more expressive implicit prosody may be projected onto DS than IS scenarios. Scenarios written in DS might help participants to simulate the settings and speculate the emotional states of characters because of hearing inner speech. Nevertheless, there was no significant emotional difference between the DS and the NS conditions.

The experiment used the NS condition as a control group to rule out the possibility that the effects observed above were from the differences between speech reporting style and narrative style. While the RTs' difference was found between the two conditions (DS-IS) as expected, no significant emotional difference suggests that perhaps the NS conditions might not be that meaningful as a control group to understand the emotional consequences of inner speech. This is because scenarios are written in NS (e.g., *Holding the map, David thought that they needed to go left*) may encourage speculations of the emotional states, similarly as when readers processing DS scenarios (e.g., *Holding the map, David said, "We need to go left")*, and therefore masks the difference between the DS and NS conditions. The difference between the two was also not the interest of the current project. Therefore, in future experiments, NS conditions were removed.

RT and emotional ratings' differences between the DS and the IS conditions support the hypothesis but it remained unclear whether the differences were driven by differences in inner speech or other non-inner speech differences between the two conditions. For

instance, the subordinate clauses in IS texts may be syntactically more complex to process, thereby increasing the reaction times and interfering with emotional judgement, as compared with DS texts. To examine whether inner speech was the key that led to the differences, Experiment 3b was designed to see if the differences remained when inner speech was controlled.

Experiment 3b

To prove that the direct speech effects on RTs and emotion judgements were driven by increased inner speech, Experiment 3b aimed to control these effects by adding secondary tasks (an articulatory suppression task vs. a finger tapping task). The underlying theory is that adopting a secondary task that suppresses subvocal articulation and therefore interferes with or blocks inner speech can examine the deficiencies on a primary task (Alderson-Day & Fernyhough, 2015). Articulatory suppression tasks have been widely used in previous studies to interfere with inner speech (e.g., Baldo et al., 2005; Lidstone et al., 2010). These studies also included nonverbal tasks as comparisons (e.g., finger tapping) to control the general secondary effects and to reveal the specific effects of inner speech (Alderson-Day & Fernyhough, 2015). Hence, to confirm that the effects observed in the previous experiments were because of inner speech, Experiment 3b adopted the dual-task method (an emotional judgement task + an articulatory suppression task vs. an emotional judgement task + a finger tapping task).

Theoretically, if the DS effects were driven by inner speech, they would be diminished when they make emotional judgements during articulatory suppression but not during finger tapping. That is, there would be differences in RTs and Ratings between the

DS-IS contrast (i.e., faster RTs and higher ratings in the DS than the IS conditions as in Experiment 3a) when participants conduct finger tapping, but the differences would be masked during articulation.

Participants

24 Participants (20 female and four male) were recruited from the University of Manchester, including students and staff, aging from 18 to 42 (mean age = 21.5, SD = 2.8). They were all native English speakers with no reported learning/reading/speaking disabilities. They had not participated in any of the above experiments. Typically, the experiment lasted 20 minutes and participants were awarded either two course credits or £2 as compensation. The experiment was approved by the ethics committee of the University of Manchester (Research ethics code: 16248).

Stimuli

The current experiment adopted stimuli from Experiment 3a (NS sentences were removed). There were ninety scenarios, including 60 critical and 30 scenarios (as fillers). There were also ten sentences for practising the dual tasks. Together, four stimulus lists were constructed in the same way as in Experiment 3a. Participants were randomly allocated to read one of the lists.

Procedure

The experiment was conducted in a quiet laboratory setting. All participants gave informed consent before the experiment. They were then seated in front of a computer

monitor, which presented the experiment using OpenSesame 3.1. Same procedure as Experiment 3a was adopted. Participants were instructed to silently read sentences and give their emotional ratings by pressing number keys (0-3). Each number keypress triggered the next trial. However, different from Experiment 3a, participants were asked to do additional tasks concurrently when they did the reading tasks. Participants were randomly instructed to either engage in an articulatory suppression task (i.e., saying "1,2,3,4" repeatedly) or in a finger tapping task (i.e., tapping their left little finger upon ring finger, middle finger, and index finger and repeating in the same order) while they read at their preferred speed. Each participant did half trials during articulatory suppression and another half trials during finger tapping. Participants practised the tasks before the testing phases. Responses and response times were recorded.

Results and discussion

The data trimming process was the same as in Experiment 3a: RTs longer than 10000 ms were deleted; RTs over three standard deviations from the mean RT of each participant by condition were removed from the data. The trimming procedure discarded 2.6% of the total values. A GLMM of RTs was constructed with a maximal random-effect structure, with *Speech* and *Task* as fixed factors, and *subject* and *item* as random factors. All the *p* values were estimated with the *lmerTest* package (Kuznetsova, 2017) in R.

The full model (Model 9) estimates were presented in Table 12. Both main effects of reporting styles and secondary tasks were observed. Participants made faster emotional judgements on DS sentences than IS sentences; they also responded faster during articulation suppression than during finger tapping. The interaction between reporting styles

and secondary tasks was significant (See Figure 8). As shown in the figure, the RT difference between the DS and the IS conditions was smaller when participants performed the articulatory suppression task than the finger-tapping task. That is, the articulatory suppression task restrained the DS effect on RTs as hypothesised.

Fixed effects	**Estimate**	**S.E.**	***t***	***p***	**Cohen's *d***
DS-IS	-303.09	12.09	-25.06	**<0.001**	0.477
AS-FT	-87.42	9.69	-9.03	**<0.001**	0.123
Speech × Task	136.74	10.44	13.10	**<0.001**	0.176

Table 12. The GLMM's estimates of RTs (AS = articulatory suppression condition, FT = finger tapping condition, Speech × Task = interaction between reporting styles and secondary tasks; Model 9, Experiment 3b)

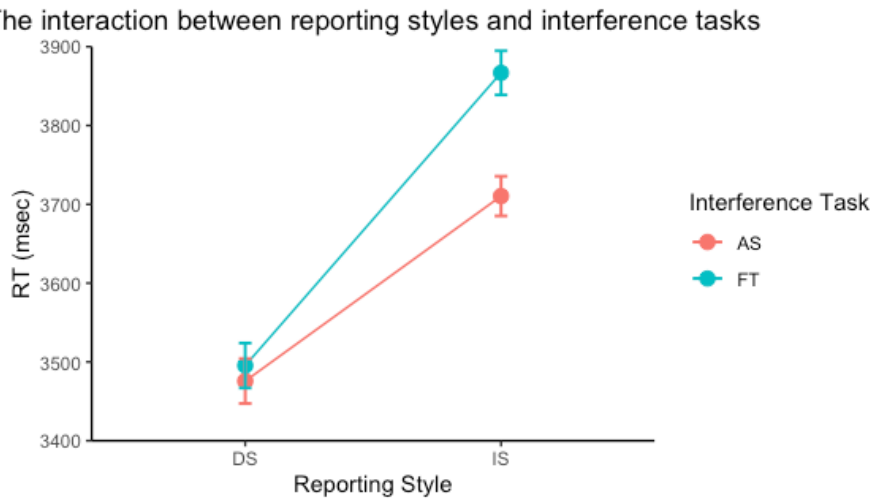

Figure 8. The interaction between reporting styles and interference tasks.

A LMM of ratings with a maximal random-effect structure was fitted to test the DS effects, secondary task effects, and the interaction between reporting styles and secondary tasks. Results were presented in Table 13. There were no significant effects.

Fixed effects	**Estimate**	**S.E.**	***t***	***p***	**Cohen's *d***
DS-IS	0.07	0.05	1.37	0.186	0.086
AS-FT	-0.05	0.06	-0.75	0.459	0.053
Speech × Task	-0.03	0.10	-0.29	0.777	0.031

Table 13. The LMM's estimates of ratings (Model 10, Experiment 3b).

As expected, the overall RT difference between the DS and the IS condition was found, indicating a DS effect in processing time. The experiment also found the RT difference between the two interference tasks. This might suggest that the articulatory suppression task less interfered with the primary task (i.e., making emotional judgements) than the finger-tapping task in the experiment. Therefore, the effect of secondary task interference needed to be tested. The interaction observed supported the hypothesis as well. The RT difference between the DS and the IS conditions was reduced when individuals conducted the AS task rather than the FT task, which might be a result of the suppression of inner speech in the AS condition.

Regarding emotional ratings, no effects were found, rejecting the hypothesis. However, as the experiment recruited only 24 participants, the sample size might be too small to reveal the effect. Alternatively, this could be causally affected by the secondary

task interference. Articulatory suppression involved “1, 2, 3, 4” recitation, which could intervene the “0, 1, 2, 3” buttons that participants pressed for ratings; tapping left fingers might meddle with right finger pressing. Therefore, the secondary tasks interference possibly masked the rating difference of the DS-IS contrast between the two interference conditions. In other words, the secondary tasks might be too confusing for participants to undertake, which affected the emotional judgements that they made concurrently.

To examine the latter speculation, Experiment 3c replicated the above experiment with two easier interference tasks.

Experiment 3c

Experiment 3c aimed to reproduce the effects on RTs found in Experiment 3b and test whether easier secondary tasks would reveal the expected differences in ratings between the AS and FT conditions. Specifically, I expected that participants would respond faster and perceive higher emotional arousal for DS than IS quotations in general. Moreover, it was hypothesised that faster RTs and higher ratings would be discovered in the DS than the IS conditions when participants conducted the FT task, but not when they performed the AS task.

Participants

24 native English speakers with no reported learning/reading/speaking disabilities participated in the experiment. There were 22 female and two male participants, aged 19.0 on average (SD = 0.7). They had not participated in any previous experiments of this project. The experiment typically lasted 20 minutes. Participants were awarded with two course credits or paid £2 after the experiment.

Stimuli

The stimuli were the same as in Experiment 3b.

Procedure

Participants followed the same procedure as in Experiment 3b with one exception – different interference tasks. Instead of repeating "1, 2, 3, 4", participants were asked to recite "bah, bah, bah…" following metronome beats (100 beats per minute). The FT task was altered from four-finger tapping to one-finger tapping – left index finger tapping the spacebar on the keyboard at the beats of the metronome (100 beats per minute). The metronome beats were introduced to remind participants of the secondary tasks. Each participant practised five times for each secondary task before the testing phase. RTs and ratings were recorded.

Results and discussions

Following the same trimming procedure, RTs longer than 10000 ms were removed; RTs over three standard deviations from the mean RT of each participant by condition were deleted from the data. 2% of the total values were removed. To analyse the effects on RTs, a GLMM (Model 11) was constructed with *speech* and *task* as fixed factors and *subject* and *item* as random factors following a maximal random-effect structure. *lmerTest* package in R was used to estimate *p* values.

Table 14 presented the model estimates. There was a main effect of *speech* on RTs. Different from Experiment 3b, no effect of *task* was found. The interaction between *speech* and *task* was significant, but in the opposite direction against the interaction observed in Experiment 3b (Figure 9). The RT difference between the DS and the IS conditions was

larger when participants undertook the AS task than the finger task. This easier AS task (as compared to the AS task in Experiment 3b) strengthened the DS effects on RTs, which was contradictory to the hypothesis.

Fixed effects	**Estimate**	**S.E.**	***t***	***p***	**Cohen's *d***
DS-IS	-421.64	12.60	-33.48	**<0.001**	0.657
AS-FT	7.85	11.20	0.70	0.484	0.011
Speech × Task	-121.50	10.85	-11.20	**<0.001**	0.155

Table 14. The GLMM's estimates of ratings (Model 11, Experiment 3c).

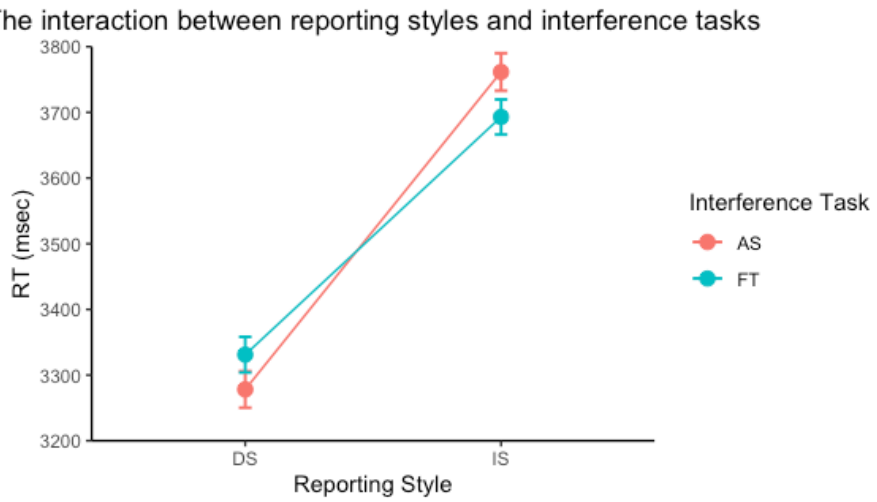

Figure 9. The interaction between speech and task.

A LMM (Model 12) of ratings with a maximal random-effect structure was constructed to test the effects of *speech*, *task*, and the interaction between them. However, again, no effect on ratings was found (Table 15).

Fixed effects	Estimate	S.E.	*t*	*p*	Cohen's *d*
DS-IS	0.04	0.03	1.33	0.184	0.054
AS-FT	-0.02	0.06	-0.38	0.706	0.030
Speech × Task	0.07	0.07	0.93	0.362	0.084

Table 15. The LMM's estimates of ratings (Model 12, Experiment 3c).

The results again confirmed the RTs' advantage in the DS condition, as compared with the IS condition. However, inconsistent with Experiment 3b, there was no difference between overall RTs between the AS and the FT conditions. This might suggest that in this experiment, the secondary tasks were equivalent in the difficulties of the tasks. The inconsistency reflected the different interference of the secondary tasks (either AS or FT). Ideally, the AS task should be as difficult as the FT task, but only because the AS task interferes with inner speech, the difference in the DS-IS contrast would be masked. That is, the expected version of secondary tasks would have no effects on RTs or ratings but interact with reporting styles. In this context, the tasks in Experiment 3c were more appropriate for testing the "inner speech hypothesis".

Nevertheless, the interaction of the two independent variables observed in this experiment conflicted with the one found in Experiment 3b. It presented a reversed result – instead of mitigating the effect of DS on RTs, the AS task enhanced it. A possible explanation was that the revised AS task might be too easy to perform suppression but instead facilitated inner speech. That is, the mouth actions might activate readers' Broca's

area and thus encouraged the production of inner speech, resulting in a faster RT in the DS than the IS conditions.

No effects on emotional arousal were found, as in the former experiment, rejecting the hypothesis. Again, the explanation could arise from the interference tasks, as this was the largest difference of experimental design between Experiment 3a and the other two. When comparing the results of the three experiments, I observed that, in general, readers responded slower in Experiment 3a (3830 ms) than in Experiment 3b (3320ms) and Experiment 3c (3256 ms). One assumption was that with the secondary tasks added, participants might feel urged to made decisions and therefore assigned ratings based on the "external" emotional cues (i.e., the context) only, regardless of the aid of inner speech. Another way to interpret it was that both AS and FT tasks in Experiment 3b & 3c interfered with participants' perception of emotional arousal. AS tasks might, as hypothesised, halt the production of inner speech and thus mask the difference of ratings between DS and IS conditions. On the other hand, FT tasks could affect how readers made emotional judgements due to its general auditory interference (the emotionless reciting sounds in both experiments) and simply confusing finger pressing (in Experiment 3b) for quotations in whichever reporting style. In either case, this needs further studies to reveal the truth.

General Discussion

The present project aims to explore the emotional consequences of hearing inner speech in silent reading of direct speech quotations. Specifically, the above five experiments attempted to examine how inner speech influenced the interpretation of a reported speaker's emotional state in written stories. I predicted that hearing inner speech helped readers to

access the emotional states of the reported speakers, and the consequences could be a quantitative effect on response times (i.e., faster RTs) and qualitative effects on emotional judgements (i.e., making more appropriate choices). In other words, inner speech could improve the speed as well as the intensity of emotional activations. The project partially supported this central hypothesis.

All the experiments demonstrated a quantitative boost to emotional judgements after reading direct speech quotations. In Experiment 1, the faster RTs were observed in the DS than the IS and the NS conditions, which was then replicated in Experiment 2. That is, an RT advantage appeared in the DS conditions when readers judge categories of emotions (regardless of choosing facial expressions or emotion words). This effect of reporting style remained significant while readers judge the intensity of emotions in Experiment 3a, 3b, and 3c. Importantly, Experiment 3b & 3c introduced interference tasks (AS vs. FT) to confirm whether the RTs' difference between the DS and the IS conditions arose from differential processing of speech or non-inner speech discrepancies, such as linguistic content or grammar. Experiment 3b proved that inner speech could be the key that caused the difference by presenting an interaction between reporting style and task – indicating the AS task (supposed to block inner speech) reduced the difference. However, an opposite interaction was reported when easier secondary tasks were adopted in Experiment 3c. It was possible that the AS task in the latter experiment was too easy to suppress but rather enhanced inner speech by the pre-activation of individual's Broca's area and consequently the increased production of inner speech.

I also hypothesised that hearing inner speech could assist participants in making more appropriate emotional judgements. The qualitative effects could be reflected by either more consistent choices (Experiment 1 & 2) or higher perceived emotional arousal (Experiment 3a, 3b, and 3c) in the DS than the IS conditions. The first two experiments adopted the intensity of the quotations as the other independent variable besides reporting style. It was expected that there would be more consistent choices for DS than IS quotations overall and the intensity of the quotations would interact with this effect. However, both experiments observed a main effect of reporting style on the choices' consistency, yet in the contradictory direction (i.e., more consistent choices were given to IS quotations). The unexpected finding suggested that hearing inner speech might not just simply increase the intensity of emotional activation in these experiments, or it did, but because of the vivid presentation, it might add more possibilities when readers made emotional judgements. In Experiment 1, there was a lack of interaction between reporting style and intensity in emotional judgements. The lack of qualitative effects may be explained by the neutral contents of the quotations (e.g., "It's *half past three*"). Because quotations themselves do not carry any emotional information, participants' emotional judgements were determined predominantly by the contexts. As there was no time limit for responding, participants would eventually reach the same emotional interpretation based on the same contexts between the DS and IS conditions, despite hearing inner speech in the DS conditions. To confirm whether there was interaction, Experiment 2 replaced facial expressions with emotion words to reduce the additional efforts in processing facial expressions. An interaction was found – readers gave more inconsistent choices in the DS than IS conditions for neutral sentences than low-intensity sentences. This could mean that inner speech

generated in the DS conditions tended to be ambiguous that added more uncertainty to the neutral sentences, but not when they read low-intensity sentences. However, this could not explain why there was no interaction between neutral sentences and high-intensity sentences on this matter.

This led to my concern that the experimental setup in the above two experiments may be ineffective in testing inner speech's 'qualitative' emotional consequence. Another problem was the use of the 'neutral' labels in both experiments. By increasing the statistical power, Experiment 2 revealed the effect of intensity – participants responded faster for neutral sentences than high- or low-intensity sentences. This proved what I expected, i.e., with 'neutral' as an option, it was easier for participants to make decisions for neutral sentences. However, since the quotations were neutral, participants may be inclined to make 'neutral' responses rather than other emotional responses. Indeed, in both experiments, the percentages of 'neutral' responses were the highest in both the DS and the IS conditions. They were also top-chosen emotion labels for low-intensity sentences. In Experiment 2, neutral even owned the second-highest percentages for high-intensity sentences. Considering high-intensity sentences are supposed to increase the responses to high arousal emotions, this result was surprising. It should be noted that, in both experiments, after the experiment session, many of the participants commented that some trials were too hard to make judgements or to understand the emotions of the characters without detailed contexts. And for these trials, they chose 'neutral' as they thought 'neutral' labels could fit in most cases. As a result, the 'neutral' labels may also be responsible for the unexpected findings on the qualitative effects. That is, the tendency of choosing the

‘neutral’ labels may mask potential qualitative effects of other emotion labels and lead to the interaction effect between reporting style and intensity on ‘neutral’ responses.

On the other hand, Experiment 3a presented the expected effect – higher emotional ratings for DS than IS quotations. Nevertheless, the effect was absent when interference tasks were added in Experiment 3b & 3c. The interference tasks, either AS tasks intended to interfere with the production of inner speech or FT tasks being nonverbal tasks as control, both masked the qualitative effect. This was possible considering AS tasks block inner speech while FT tasks could have auditory meddling or confusing finger pressing with number ratings. This could be examined by comparing several interference tasks with different difficulties in future research.

Overall, the project indicated faster response times when individuals read DS quotations, which could be one of the emotional consequences of hearing inner speech. Varied results of the possible qualitative effect suggested that inner speech might help when readers made emotional judgements but not simply enhance the intensity of emotional activations.

The findings are in line with previous studies. Most importantly, the present findings support the proposal that hearing inner speech has emotional consequences. Neuroimaging evidence supports the hypothesis. Neuroimaging studies have reported an overlapping activation (i.e., the right TVA) between inner speech (e.g., Yao et al., 2011) and emotional prosody studies (e.g., Ethofer et al., 2009). Additionally, the observed quantitative effects support the vividness distinction between the two reporting styles in silent reading (i.e.,

readers tend to engage in more vivid inner speech in DS texts than IS texts; Yao et al., 2011). As the sentences were virtually the same between the DS and the IS conditions, the advantage of RTs in the DS condition is believed to be the consequence of hearing inner speech in the DS condition because the vivid inner speech facilitated emotional judgements. Moreover, research has implied that inner speech, like overt speech, also has acoustic features such as pitch and tempo (e.g., Alexander & Nygaard, 2008; MacKay, 1992). Such acoustic features may constitute an ‘inner speech prosody’ that conveys emotional information. The current findings also agree with this statement. As the quantitative effects on RTs are believed to be attributed to inner speech, inner speech should contain emotional information, which made emotions more accessible and consequently made emotional judgements easier for the participants. The emotional information brought by inner speech could only be from the context. In other words, the quantitative effects are best interpreted as effects of ‘inner speech prosody’.

The major strength of the present project is that it extended our knowledge of inner speech, supporting the possibility that hearing inner speech has emotional consequences. The current studies are the first to show ‘quantitative’ emotional consequences of hearing inner speech. This has important implications for the impact of using speech quotations on written communications. Using DS quotations instead of IS quotations may help addressees to understand the emotion conveyed by the quotations more efficiently. This would be helpful when writers, for instance, try to communicate certain emotions to readers. The present studies also provide a solid behavioural foundation for future research on the neural correlates of emotional processing during silent reading of DS quotations.

However, the qualitative effect or how inner speech influences the nature of emotional activations remains uncertain due to the diverse findings. Experiment 1 & 2 reported less consistency for DS than IS quotations, indicating hearing inner speech might add uncertainty to one's emotional judgements, while Experiment 3a observed a higher perceived emotional intensity of DS than IS quotations, suggesting inner speech might increase the intensity of emotional activations. One way to interpret it is that hearing inner speech can enhance both vividness of emotional information and intensity of emotional activations. Nevertheless, neither the use of intensity of quotations as a factor nor the use of interference tasks did not reveal interactions that were in line with the assumption. Larger sample sizes could reveal the real qualitative effects. It could also suggest defects in experimental setups as mentioned above, which could not properly examine the effects. Finally, there could be task-driven effects in the above studies. In the experiments, participants were asked to make explicit emotional judgements. Engaging in emotion-matching or emotional-rating tasks, participants may be encouraged to imagine the emotional aspects of the scenarios. It was thus difficult to determine whether the quantitative boosts to RTs were driven by spontaneously generated inner speech or by explicit imagery of emotions that was encouraged by the tasks.

The current project has several theoretical implications on the possible cognitive and emotional consequences of hearing inner speech in reading. The quantitative effect of direct speech suggests that inner speech may help faster access the emotional state of the protagonist, therefore facilitating emotional judgements in reading. Such emotional facilitation implies an important role of inner speech in reading process, which can be

possibly linked to theory of mind (ToM) in reading – the ability to understand mental states of characters (Alderson-Day et al., 2020). Research has reported the significance of ToM when readers follow the mental states of the speakers, such as beliefs, intents, and emotions (e.g., Spreng, et al., 2009; Herman, 2008). However, there has been inconsistent evidence concerning the relationship between inner speech and ToM. In a neuroimaging study, Alderson-Day et al. (2020) tested whether the DS-IS contrast could be observed both in speech and thoughts (e.g., *David said, "We need to go left" (DS)- David said that they need to go left (IS)* vs. *David thought, "We need to go left" (DS) – David thought that they need to go left (IS)*). The study found elevated responses to direct rather than indirect speech in Voice-Selective auditory cortex, while no such effect was found for thoughts (Alderson-Day et al., 2020). Moreover, the same pattern was observed in ToM regions, indicating a correlation between inner speech and ToM (Alderson-Day et al., 2020). The inner speech-induced emotional facilitation is in line with the finding.

Inspired by Alderson-Day et al. (2020), I suggest that future studies could use the comparison between speech and thoughts to confirm whether the emotional facilitation is the consequence of hearing inner speech or merely differences between speech reporting (e.g., grammar or source memory). Future experiments could replicate the paradigm to investigate whether the difference of the direct speech effect between speech and thoughts remains when readers make emotional arousal ratings. Alternatively, future studies can keep the use of interference tasks, but a variety of tasks should be tested to examine whether difficulties of the tasks influence the effects differently.

The emotional facilitation caused by inner speech could be one possible response to 'why inner speech occurs'. The interpretation may help understand the circumstance of auditory verbal hallucinations in schizophrenia. Many have argued that AVHs are experiences that one's inner speech misattributed to a third party outside his/her brain (Barber et al., 2021; Vilhauer, 2016). Future studies are advised to extend the findings to AVHs.

Also, inner speech-induced emotional facilitation also supports the association between literary reading and higher empathy. Evidence has found that people who read literary text were more empathic than those who read the same text without its literary elements (Koopman, 2016). Meta-analyses have also reported a positive effect of fiction-reading on empathy (Dodell-Feder & Tamir, 2018; Mumper & Gerrig, 2017). Theoretically, when reading narrative texts, people simultaneously take on the character's perspective to understand the character. Readers infer what the character thinks and how they feel, and therefore understand more about the character (Koopman, 2018). In this way, literary reading could be considered as 'practice for inferring emotions and taking the perspectives of others in real life' (i.e., empathy; Koopman, 2018). The finding of emotional facilitation by inner speech is in accordance with the association between literary reading and empathy. Hearing inner speech in reading helps readers access characters' emotional states, leading to faster inferences about characters' emotions and consequently higher empathy. Future studies are suggested to introduce empathy as a factor to further understand the emotional consequences of hearing inner speech.

Conclusion

This project explored the emotional consequences of hearing inner speech in silent reading of direct speech quotations. To achieve this aim, five experiments were conducted to examine how inner speech influenced the judgement of a reported speaker's emotional state in direct speech quotations. The central hypothesis was that readers could better access the emotional states of the reported speakers with the aid of inner speech and thus make faster (quantitative) and better (qualitative) emotional judgements. All the experiments discovered the RT's advantage in the DS conditions, indicating a quantitative benefit and supporting the hypothesis.

However, the investigation on the qualitative effect revealed mixed findings. Experiment 3a observed a higher perceived emotional arousal in the DS than IS conditions, supporting the hypothesis. However, later experiments (3b & c) did not replicate this effect, suggesting that future studies could consider better interference tasks. Both Experiment 1 & 2 indicated a lower consistency of the DS than the IS quotations, which was opposite to the expectations. This might imply that ambiguous inner speech generated could add more uncertainty to the neutral quotations.

The current project is the first to show the RT's advantage in silent reading of DS quotations, supporting the possibility that hearing inner speech has emotional consequences as hearing overt speech. It has implications for using the DS reporting style on written communications. It also provides a behavioural foundation for future neuroimaging studies on emotional processing when hearing inner speech. Future studies are advised to confirm

the established quantitive effect by using different interference tasks or comparisons between reporting styles of speech and thoughts (direct speech + indirect speech vs. direct thoughts + indirect thoughts) and inspect the possible qualitative effects.

References

Abramson, M., & Goldinger, S. D. (1997). What the reader's eye tells the mind's ear: silent reading activates inner speech. *Perception & Psychophysics*, *59*(7), 1059–1068. https://doi.org/10.3758/bf03205520

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*(5), 931–965. https://doi.org/10.1037/bul0000021

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. *Journal of Cognitive Neuroscience*, *32*(9), 1637–1653. https://doi.org/10.1162/jocn_a_01571

Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: talker-specific auditory imagery in reading. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(2), 446–459. https://doi.org/10.1037/0096-1523.34.2.446

Ashby, J., & Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*, *96*(3), B89-100. https://doi.org/10.1016/j.cognition.2004.12.006

Banfield, A. (1973). Narrative style and the grammar of direct and indirect speech. *Foundations of Language*, *10*(1), 1–39

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression.

Journal of Personality and Social Psychology, *70*(3), 614–636. https://doi.org/10.1037//0022-3514.70.3.614

Barber, L., Reniers, R., & Upthegrove, R. (2021). A review of functional and structural neuroimaging studies to investigate the inner speech model of auditory verbal hallucinations in schizophrenia. Translational psychiatry, 11(1), 1-12.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/doi.org/10.1016/j.jml.2012.11.001

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312. https://doi.org/10.1038/35002078

Brocklehurst, P., & Corley, M. (2009). Lexical bias and the phonemic similarity effect in inner speech. In *15th Annual Conference on Architectures and Mechanisms for Language Processing,* (p. 7–9 September 2009). Barcelona.

Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, *8*(4), 383–403. https://doi.org/10.1016/j.plrev.2011.10.002

Carroll, N. C., & Young, A. W. (2005). Priming of emotion recognition. *The Quarterly Journal of Experimental Psychology Section A*, *58*(7), 1173–1197. https://doi.org/10.1080/02724980443000539

Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, *66*(4), 764–805. https://doi.org/10.2307/414729

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Lawrence

Erlbaum Associates. *Hillsdale, NJ*, 20-26.

Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error biases in inner and overt speech: evidence from tongue twisters. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(1), 162–175. https://doi.org/10.1037/a0021321

de Gelder, B., Meeren, H. K. M., Righart, R., Stock, J. van den, van de Riet, W. A. C., & Tamietto, M. (2006). Beyond the face: exploring rapid influences of context on face processing. In *Progress in Brain Research* (Vol. 155, pp. 37–48). Spain: Elsevier. https://doi.org/10.1016/S0079-6123(06)55003-4

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by eye and ear. *Cognition and Emotion*, *14*(3), 289–311.

Diamond, E., & Zhang, Y. (2016). Cortical processing of phonetic and emotional information in speech: A cross-modal priming study. *Neuropsychologia*, *82*, 110–122. https://doi.org/10.1016/j.neuropsychologia.2016.01.019

Dodell-Feder, D., & Tamir, D. I. (2018). Fiction reading has a small positive impact on social cognition: A meta-analysis. Journal of Experimental Psychology: General, 147(11), 1713.

Ethofer, T., Van De Ville, D., Scherer, K., & Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology*, *19*(12), 1028–1033. https://doi.org/10.1016/j.cub.2009.04.054

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. Behavior research methods, 41(4), 1149-1160.

Filik, R., & Barber, E. (2011). Inner speech during silent reading reflects the reader's

regional accent. *Plos One*, *6*(10), e25782. https://doi.org/10.1371/journal.pone.0025782

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, *27*(2), 285–319.

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A. (2011). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain: A Journal of Neurology*, *134*(Pt 10), 3071–3082. https://doi.org/10.1093/brain/awr232

Hanson, V. L., Goodell, E. W., & Perfetti, C. A. (1991). Tongue-twister effects in the silent reading of hearing and deaf college students. *Journal of Memory and Language*, *30*(3), 319–330. https://doi.org/10.1016/0749-596X(91)90039-M

Herman, D. (2008). Narrative theory and the intentional stance. Partial Answers: Journal of Literature and the History of Ideas, 6(2), 233-260.

Huang, J., Carr, T. H., & Cao, Y. (2002). Comparing cortical activations for silent and overt speech using event-related fMRI. *Human Brain Mapping*, *15*(1), 39–53. https://doi.org/10.1002/hbm.1060

Huey, E. B. (1908). The inner speech of reading and the mental and physical characteristics of speech. In *The psychology and pedagogy of reading* (pp. 117–141). The Macmillan Company.

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilisation and decoding accuracy in vocal expression of emotion. *Emotion*, *1*(4), 381–412. https://doi.org/10.1037/1528-3542.1.4.381

Klapp, S. T. (1971). Implicit speech inferred from response latencies in same-different

decisions. *Journal of Experimental Psychology*, *91*(2), 262—267.

Koopman, E. M. E. (2016). Effects of "literariness" on emotions and on empathy and reflection after reading. Psychology of Aesthetics, Creativity, and the Arts, 10(1), 82.

Koopman, E. M. (2018). Does originality evoke understanding? The relation between literary reading and empathy. Review of General Psychology, 22(2), 169-177.

Kurby, C. A., Magliano, J. P., & Rapp, D. N. (2009). Those voices in your head: activation of auditory images during reading. *Cognition*, *112*(3), 457–461. https://doi.org/10.1016/j.cognition.2009.05.007

Li, C. N. (1986). Direct speech and indirect speech: A functional study. In F. Coulmas (Ed.), *Direct and indirect speech* (pp. 29–45). Berlin: Mouton de Gruyter.

MacKay, D. G. (1992). Constraints on theories of inner speech, 121–149.

Mar, R. A., Oatley, K., & Peterson, J. B. (2009). Exploring the link between reading fiction and empathy: Ruling out individual differences and examining outcomes.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

McCutchen, D., & Perfetti, C. A. (1982). The visual tongue-twister effect: Phonological activation in silent reading. *Journal of Verbal Learning and Verbal Behavior*, *21*(6), 672–687. https://doi.org/10.1016/S0022-5371(82)90870-2

Morin, A., El-Sayed, E., & Racy, F. (2015). Self-awareness, inner speech, and theory of mind in typical and ASD individuals: A critical review. Theory of mind: development in children, brain mechanisms and social implications. Nova Science Pub.

Mumper, M. L., & Gerrig, R. J. (2017). Leisure reading and social cognition: A meta-

analysis. Psychology of Aesthetics, Creativity, and the Arts, 11(1), 109.

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*(1), 528–537. https://doi.org/10.1016/j.cognition.2007.02.006

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 Pt 1), 182–193. https://doi.org/10.1006/nimg.2001.0779

Partee, B. H. (1973). The syntax and semantics of quotation. In S. R. Anderson & P. Kiparsky (Eds.), *A festschrift for Morris Halle* (pp. 410–418). New York: Holt, Reinhart and Winston.

Paulmann, S., & Pell, M. (2010). Contextual influences of emotional speech prosody on face processing: How much is enough? *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 230–242. https://doi.org/10.3758/CABN.10.2.230

Paulmann, S., & Pell, M. D. (2009). Facial expression decoding as a function of emotional meaning status: ERP evidence. *Neuroreport*, *20*(18), 1603–1608. https://doi.org/10.1097/WNR.0b013e3283320e3f

Pell, M. D. (2002). Evaluation of nonverbal emotion in face and voice: Some preliminary findings on a new battery of tests. *Brain and Cognition*, *48*(2–3), 499–504.

Pell, M. D. (2005). Nonverbal emotion priming: Evidence from the "Facial Affect Decision Task." *Journal of Nonverbal Behavior*, *29*(1), 45–73. https://doi.org/10.1007/s10919-004-0889-8

Pell, M. D. (2005). Prosody-face interactions in emotional processing as revealed by the facial affect decision task. *Journal of Nonverbal Behavior*, *29*(4), 193–215.

https://doi.org/10.1007/s10919-005-7720-z

Pell, M. D., Jaywant, A., Monetta, L., & Kotz, S. A. (2011). Emotional speech processing: disentangling the effects of prosody and semantic cues. *Cognition & Emotion*, *25*(5), 834–853. https://doi.org/10.1080/02699931.2010.516915

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. *Behavioural Brain Research*. https://doi.org/10.1016/j.bbr.2013.12.034

Pittam, J., & Scherer, K. (1993). Vocal expression and communication of emotion. In L. Michael & H.-J. Jeannette (Eds.), *Handbook of Emotions*. New York: The Guildford Press.

Planalp, S. (1996). Communicating emotion in everyday life. In *Handbook of communication and emotion* (pp. 29–48). Elsevier. https://doi.org/10.1016/B978-012057770-5/50004-7

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. https://doi.org/10.1016/j.neuroimage.2012.04.062

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. https://doi.org/10.1016/j.bandl.2004.07.007

Smith, S. M., Brown, H. O., Toman, J. E., & Goodman, L. S. (1947). The lack of cerebral effects of d-tubocuarine. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, *8*(1), 1–14.

Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. Journal of cognitive neuroscience, 21(3), 489-510.

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

Tian, X., Ding, N., Teng, X., Bai, F., & Poeppel, D. (2018). Imagined speech influences perceived loudness of sound. *Nature Human Behaviour*, *2*(3), 225–234. https://doi.org/10.1038/s41562-018-0305-8

Vercueil, L., & Perronne-Bertolotti, M. (2013). Ictal inner speech jargon. *Epilepsy & Behavior*, *27*(2), 307–309. https://doi.org/10.1016/j.yebeh.2013.02.007

Vilhauer, R. P. (2016). Inner reading voices: An overlooked form of inner speech. Psychosis, 8(1), 37-47.

Vygotsky, L. S. (1987). Thinking and Speech. The collected works of Lev Vygotsky (Vol. 1). New York, NY: Plenum Press. https://doi.org/10.1016/B978-0-444-41663-6.50012-9

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

Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, *23*(October), 3146–3152. https://doi.org/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. https://doi.org/10.1016/j.neuroimage.2012.01.111

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

Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, *23*(10), 3146–3152. https://doi.org/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. https://doi.org/10.1016/j.neuroimage.2012.01.111

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

Yao, B. (2021). Mental Simulations of Phonological Representations Are Causally Linked to Silent Reading of Direct Versus Indirect Speech. Journal of Cognition, 4(1), 6. https://doi.org/10.5334/joc.141