Vocal-motor interference eliminates the memory advantage for vocal melodies

Emily A. Wood, Joseph Rovetti, Frank A. Russo

ARTICLE INFO

Keywords:
Embodied cognition
Sensorimotor simulation
Motor system
Memory
Music
Singing

ABSTRACT

Spontaneous motor cortical activity during passive perception of action has been interpreted as a sensorimotor simulation of the observed action. There is currently interest in how sensorimotor simulation can support higher-up cognitive functions, such as memory, but this is relatively unexplored in the auditory domain. In the present study, we examined whether the established memory advantage for vocal melodies over non-vocal melodies is attributable to stronger sensorimotor simulation during perception of vocal relative to non-vocal action. Participants listened to 24 unfamiliar folk melodies presented in vocal or piano timbres. These were encoded during three interference conditions: whispering (vocal-motor interference), tapping (non-vocal motor interference), and no-interference. Afterwards, participants heard the original 24 melodies presented among 24 foils and judged whether melodies were old or new. A vocal-memory advantage was found in the no-interference and tapping conditions; however, the advantage was eliminated in the whispering condition. This suggests that sensorimotor simulation during the perception of vocal melodies is responsible for the observed vocal-memory advantage.

1. Introduction

For several decades, we have known that producing and perceiving actions are closely linked. For instance, cortical motor areas involved in executing actions are also activated by observing actions (Iacoboni, 2009; Iacoboni & Mazziotta, 2007). Moreover, a number of studies have shown that patients with lesions to the motor system have impairments in recognizing actions (for review see Urgesi, Candidi, & Avenanti, 2014). Because observed actions are represented in part by the motor commands that produce them, spontaneous cortical motor activation in the absence of any overt movement by the observer has been interpreted as a sensorimotor simulation of the perceived action (Gallese, 2007; Jeannerod, 2001).

The cortical circuit responsible for simulation spans from early sensory processing in the superior temporal sulcus, which in turn relays information to parietal regions (including the inferior parietal lobule; IPL), as well as premotor areas and the inferior frontal gyrus (IFG). Together, this fronto-temporo-parietal system has been called the action observation network (AON). The AON overlaps substantially with but is broader than the human mirror neuron system, and has been proposed as the system responsible for observational learning (Cross, Kraemer, Hamilton, Kelley, & Grafton, 2009), responding not only during action production, but also during passive perception of action (Buccino et al., 2001; Caspers, Zilles, Laird, & Eickhoff, 2010; Rizzolatti & Craighero, 2004; Rizzolatti & Fogassi, 2014). In addition to the AON, primary motor cortex is responsive to both action perception and production (Caetano, Jousmäki, & Hari, 2007; Hari et al., 1998; Kilner & Lemon, 2013; Palmer, Bunday, Davare, & Kilner, 2016). Primary motor activity during action perception is thought to be a downstream effect of activation in motor planning areas (Kilner, Marchant, & Frith, 2009; Pineda, 2008), since primary motor and premotor/IFG areas are reciprocally connected (Dum & Strick, 2005; Matelli, Camarda, Glickstein, & Rizzolatti, 1986). Although primary motor cortex is not typically considered part of the AON, it is nonetheless considered part of the simulation of action, as it can be active during both perception and production (Jeannerod, 2001).

1.1. Spontaneous sensorimotor simulation of singing

Spontaneous simulation of action is more likely to occur the more easily that action can be mapped onto an individual’s motor repertoire (Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, 2006; Cross,
corroborate this interpretation by showing that pupil dilation is greater as being attributable to the enhanced arousal arising from processing & Peretz, 2019) or to liking ratings for melodies (Weiss et al., 2012). Weiss, Vanzella, et al., 2015b), and individuals with deficits in pitch seven years of age (Weiss, Schellenberg, et al., 2015a), piano players Weiss, Vanzella, et al., 2015b). This memory advantage has been de-
Trehub, 2017; Weiss, Trehub, & Schellenberg, 2012; Weiss, Trehub, Weiss, Schellenberg, Peng, & Trehub, 2019; Weiss, Schellenberg, & Trehub, 2017; Weiss, Trehub, & Schellenberg, 2012; Weiss, Trehub, & Schellenberg, & Habashi, 2016; Weiss, Schellenberg, et al., 2015a; Weiss, Vanzella, et al., 2015b). This memory advantage has been de-
scribed in several populations, including undergraduate psychology students (Weiss, Trehub, & Schellenberg, 2012), children as young as seven years of age (Weiss, Schellenberg, et al., 2015a), piano players (Weiss, Vanzella, et al., 2015b), and individuals with deficits in pitch perception (amusia; Weiss & Peretz, 2019). The vocal-memory advan
tage appears to be unrelated to one's ability to process pitch (Weiss & Peretz, 2019) or to liking ratings for melodies (Weiss et al., 2012).

Weiss and colleagues have interpreted the vocal-memory advantage as being attributable to the enhanced arousal arising from processing predispositions for conspecific vocalizations. In one study, the authors corroborate this interpretation by showing that pupil dilation is greater for vocal melodies compared to piano melodies (Weiss et al., 2016). However, there was no direct evidence put forward that arousal was the factor driving differences in melody recognition between timbres, even though pupil dilation did differ between timbres in the expected di-
rection.

An alternative and more specific hypothesis for the vocal-memory advantage is that it arises from stronger sensorimotor simulation during the perception of singing compared to non-vocal melodies. If perceived vocal sounds are spontaneously mapped onto the motor commands that produce them more robustly than equivalent non-vocal sounds, this could lead to a vocal-motor memory trace that reinforces subsequent vocal melody recognition. In other words, the perception of a vocal melody may give rise to an additional vocal-motor code beyond the auditory code that would be generated for an equivalent non-vocal melody.

It has been suggested that engaging multiple sources of sensory information at encoding can improve memory. For instance, the addi-
tion of explicit motor engagement to written or pictorial information facilitates memory performance, whether it be from speaking out loud (MacLeod, Gopie, Hourihan, Neary, & Ozbuko, 2010), acting out mo-
tions (Engelkamp & Zimmer, 1989; Masumoto et al., 2006), or drawing (Wammes, Meade, & Fernandes, 2016). Furthermore, evidence shows that sensorimotor representations in the absence of any overt move-
ment during encoding can also lead to an enhancement in memory (Apel, Cangelosi, Ellis, Goslin, & Fischer, 2012; Baumeister, Rumiait, & Foroni, 2015; Bernardi, Schories, Jabusch, Colombo, & Altenmüller, 2013; Decloe & Obhi, 2013; Downing-Doucet & Guérard, 2014; Guérard & Lagacé, 2014; Iani et al., 2018; Naish, Barnes, & Obhi, 2016; Pezzulo, Barca, Bocconi, & Borgi, 2010; Shethani & Pulvermüller, 2013). For example, the use of mental practice (i.e., motor imagery) supports succesful learning of new music (Bernardi et al., 2013), and simulation during perception of action phrases supports recall for those actions (Iani et al., 2018).

These findings are also consistent with theories of embodied cog-
nition, which have long proposed that perceptual states associated with action observation become integrated into memory, including sensor-

imotor representations. During retrieval of action, these representations become activated, resulting in the simulation of the sensorimotor states associated with the original action (Barsalou, 2008; Glenberg, 1997; Iani, 2019; Kent & Lamberts, 2008). Consequently, retrieval of action should be impaired by blocking the same sensorimotor resources as those that were active during observation. Indeed, one study found that motor interference of the hands during encoding led to a memory im-
pairment for hand-related action words, while motor interference of the feet during encoding led to memory impairment for foot-related action words (Shebani & Pulvermüller, 2013). In another study, interference of expressive facial muscles during perception of emotional and neutral words resulted in impaired recognition of emotional items, but not neutral words (Baumeister et al., 2015).

If the memory advantage for vocal melodies as described by Weiss and colleagues is attributable to stronger simulation of vocal melodies compared to non-vocal melodies, then the advantage should be elimi-
nated by interfering with simulation of singing during encoding. In the case of phonation (including singing), production is largely associated with larynx motor cortex (Brown, Ngan, & Liotti, 2008). As such, it is reasonable to assume that a laryngeal vocal-motor task would be required to disrupt spontaneous sensorimotor simulation of singing.

1.3. The current study

To test whether the observed memory advantage for vocal melodies is attributable to stronger sensorimotor simulation during perception of vocal compared to non-vocal melodies, we replicated Weiss and col-
leagues’ original melody recognition task but included a condition in which we disrupted simulation of singing during encoding. We achieved this by having participants isochronously whisper the syllable
“la” while encoding melodies. Selection of this task was based on literature showing that whispering restricts laryngeal muscle activity (Tsunoda, Ohta, Niimi, Soda, & Hirose, 1997), effectively preventing simulation of phonation. In addition, we wanted the participants to engage in a dynamic vocal task (i.e. constantly moving), as opposed to something like static humming, during which they could have imitated the contours of the perceived melody through subtle inflections of pitch.

In an exposure phase of the experiment, participants listened to target vocal and piano melodies in three different interference conditions: no-interference, whispering (vocal-motor interference), or tapping (non-vocal motor interference). The no-interference condition is a replication of the original studies that demonstrated the memory advantage (e.g., Weiss et al., 2012). The whispering condition was meant to disrupt sensorimotor simulation during perception of singing. The tapping condition allowed us to address potential concerns about effects being attributable to the cognitive demand of the whispering condition (Aleman & van’t Wout, 2004; Gupta & MacWhinney, 1995; Hall & Gathercole, 2011; Larsen & Baddeley, 2003; Macken & Jones, 1995; Reisberg, Smith, Baxter, & Sonenshine, 1989). Following the exposure phase, participants completed a test phase in which they judged target and foil melodies as old or new.

Given our hypothesis that simulation contributes to the memory advantage for vocal melodies, we predicted an interaction between interference condition and timbre on melody recognition scores. We expected to replicate the vocal-memory advantage in no-interference and tapping conditions, but find that the advantage is reduced in the warning condition. This would suggest that the vocal advantage is owed, at least in part, to greater sensorimotor representations of vocal than non-vocal melodies resulting in more robust memory traces for vocal melodies. We also expected that there would be an overall decrease in melody recognition scores in both the tapping and whispering conditions compared to the no-interference condition due to the increase in cognitive demand.

2. Methods
2.1. Participants

Data are reported for 60 participants (53 females, seven males), who were recruited from the Ryerson Psychology undergraduate participant pool and from the community. Participants ranged in age from 18 to 53 years old (M = 23.27, SD = 6.47). Their general musical sophistication was measured using the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen, Gingras, Musil, & Stewart, 2014), which is used to assess musicality in the general population on a scale ranging from 18 to 126. Participants’ scores ranged from 41 to 107 (M = 71.43, SD = 14.55), indicating that they varied greatly in their musicality. An additional 15 participants took part in the study but were replaced for the following reasons: failing to understand or follow task instructions (n = 3); scoring below chance (as averaged across all conditions) on the recognition test (n = 4); or anticipating that their memory for the melodies would be tested (n = 8). We rejected participants who anticipated an upcoming memory test because our process of interest was spontaneous simulation. Participants who anticipated a memory test may have been more likely to rehearse all melodies intentionally. Doing so might have recruited motor areas (Halpern & Zatorre, 1999; Leaver, Van Lare, Ziellinski, Halpern, & Rauschecker, 2009), thus providing a vocal-motor code for both vocal and piano melodies, and effectively undermining the intent of our manipulation.

The sample size of 60 was determined on the basis of an a priori power analysis done using the R package “ANOVApower” (Lakens & Caldwell, 2019). This sample size was sufficient to detect an interaction between interference condition and timbre with 80% power (α = 0.05) given the following specifications: a moderate effect size for timbre (d = 0.5) in the no-interference and tapping conditions, and recognition scores that are moderately correlated across conditions (r = .5). The effect size was a conservative estimate, as previous studies have found this recognition advantage for vocal melodies to be much larger (e.g., d = 0.92; Weiss et al., 2017).

2.2. Stimuli

Stimuli consisted of 48 British and Irish folk melodies ranging from 12.8 to 20.5 s in duration, realized in a female vocal timbre sung on “la,” as well as a piano timbre. These are the same melodies that have been used in similar studies demonstrating the memory advantage for vocal melodies over non-vocal melodies (Weiss et al., 2017; Weiss, Schellenberg, et al., 2015a). For each participant, 24 of these melodies were randomly assigned to the vocal timbre and the other 24 were assigned to the piano timbre. Of the 24 melodies within each timbre, 12 melodies were randomly assigned to be targets (presented in both the exposure phase and the test phase) and the other 12 were assigned to be foils (presented only in the test phase). Of the 12 targets within each timbre, four were assigned to the no-interference condition, four were assigned to the tapping condition, and four were assigned to the whispering condition. Thus, across both timbres, there were eight melodies in each interference condition (four melodies × two timbres). For the purpose of practice, an additional eight folk melodies were generated and recorded using Protools Version 12.8.0.

2.3. Procedure and apparatus

After arriving at the lab and receiving general information about the study requirements, participants were asked to provide written informed consent. They were then seated in a double-walled sound attenuated chamber (Industrial Acoustics Co., Bronx, NY), with an Apple iMac 27” computer monitor placed approximately 60 cm in front of them. The experiment was programmed using Psychophysics Toolbox Version 3 (Brainard, 1997) in MATLAB Version 9.5. Melodies were presented through Koss SB40 headphones at approximately 70 dB SPL.

Participants received instruction on how to execute whispering and tapping trials before engaging in practice under the supervision of a research assistant. They first practiced these conditions as stand-alone tasks, and then while listening to practice melodies. In the whispering condition, participants were instructed to isochronously whisper the syllable “la” as quickly as they could while still being comfortable. To prevent prolonged gaps in whispering that could permit sensorimotor simulation, participants were advised to repeat the syllable “la” during inhalation as well as during exhalation. In the tapping condition, participants were instructed to repeatedly tap their index finger on the desk in front of them at a rate approximately matching their production of “la” during the whispering condition. In both the tapping and whispering conditions, the instruction to perform the task at a consistently fast rate prevented participants from synchronizing their tapping or whispering to the beat of the melodies, which ranged from approximately 63–130 beats per minute (BPM). With these instructions, tapping and whispering as observed in practice were consistently in excess of 180 BPM. Furthermore, participants were asked to perform whispering and tapping as quietly as possible, so as to not interfere with perception of the melodies. In the no-interference condition, participants were instructed to listen to each melody as they normally would, with no concurrent task. Once participants could comfortably and correctly perform these conditions, they were presented with six practice trials in a random order: two no-interference, two tapping, and two whispering (with each consisting of one vocal and one piano melody). A prompt on the screen before each melody informed participants of the interference condition. If participants performed all conditions correctly, they advanced to the experimental task. If not, they were given feedback on their performance and were asked to redo the six-melody practice.

After successfully completing the practice, participants proceeded to
the exposure phase of the experiment, in which they listened to 24 melodies, consisting of four melodies in each of six interference condition-timbre pairings (e.g., tapping-voice). These melodies were presented in a random order. Participants were not informed that they would later be tested on their memory for these melodies. After listening to each melody, participants also rated how much they liked them on a 5-point scale ranging from 1 (dislike) to 5 (like; Weiss, Vanzella, et al., 2015b). These liking ratings ensured that participants were engaged with the task.

After finishing the exposure phase, participants completed a background questionnaire consisting of basic demographic questions as well as the Gold-MSI. This questionnaire took participants 5–10 min to complete. They then completed the test phase of the experiment, in which they listened to the same 24 melodies from the exposure phase (targets) as well as 24 new melodies (foils). After listening to each melody, participants rated them on familiarity using a 7-point scale ranging from 1 (definitely new) to 7 (definitely old; Weiss et al., 2012). Following the test phase, participants completed an additional questionnaire in which they indicated whether they understood and comprehended with all task instructions and whether they anticipated a memory test. Additionally, all participants were asked whether they could hear themselves tap or whisper during the encoding phase, and the final 42 participants were also asked to rate how loudly they could hear themselves whisper and tap on a scale from 1 (you could not hear yourself) to 5 (you could clearly hear yourself). Participants were then debriefed before exiting the lab.

3. Results

All statistical analyses were completed using R Version 3.6.1 (R Core Team, 2019). Analysis of variance (ANOVA) was performed using the “ez” package, with Greenhouse-Geisser adjustment of degrees of freedom applied to those effects that failed Mauchly’s test of sphericity. The calculation of effect sizes for follow-up t-tests was done using the “lqra” package.

3.1. Recognition scores for melodies

Six recognition scores were calculated for each participant. These were calculated by taking the average recognition rating for old melodies (targets) of the same timbre (e.g., piano; Weiss et al., 2012). Fig. 1 plots recognition scores as a function of interference condition and timbre. A repeated-measures 3 (interference condition) × 2 (timbre) ANOVA found a significant main effect of interference condition, F(2,118) = 7.26, p = .001, η² = 0.030, but not of timbre, F(1,59) = 1.86, p = .18, η² = 0.006. A paired-samples t-test found that recognition scores did not differ in the tapping-piano and whispering-piano conditions, t(59) = 0.16, p = .87, d = 0.021, which was in line with our expectation that the tapping and whispering conditions would be comparable with regard to cognitive demand. Furthermore, the 42 participants who rated how loudly they could hear themselves during the tapping condition (M = 2.01/5, SD = 1.06) and whispering condition (M = 1.94/5, SD = 0.83), showed no difference between these two conditions, t(41) = 0.38, p = .71, d = 0.059.

There was a significant interaction between interference condition and timbre on recognition scores, F(2,188) = 4.83, p = .010, η² = 0.014. Follow-up paired-samples t-tests assessed the simple effect of timbre on recognition scores in each interference condition. On the basis of the large number of studies showing a consistent directionality in recognition scores with respect to timbre (i.e., the vocal-memory advantage), these t-tests were one-tailed for the no-interference and tapping conditions. In the no-interference condition, recognition scores were significantly higher for vocal melodies than for piano melodies, t(59) = 2.58, p = .006, d = 0.33. This advantage for vocal melodies was also present in the tapping condition, t(59) = 1.69, p = .048, d = 0.22, but there was no effect of timbre on recognition scores in the whispering condition, t(59) = .83, d = 0.13.

A correlational analysis found that general musical sophistication did not predict overall recognition scores (as averaged across all conditions), r(58) = −0.03, p = .82. General musical sophistication also did not predict the extent to which participants recognized vocal melodies better than piano melodies. To assess this, we obtained a difference score for each participant based on recognition scores for vocal melodies less recognition scores for piano melodies in each condition. The correlation between the difference score and musical sophistication was not significant in the no-interference condition, r(58) = −0.023, p = .86, in the tapping condition, r(58) = −0.068, p = .61, or in the whispering condition, r(58) = −0.13, p = .31.

3.2. Liking ratings for melodies

Fig. 2 shows liking ratings for melodies as a function of interference condition and timbre. Liking ratings were analyzed using a repeated-measures 3 (interference condition) × 2 (timbre) ANOVA, which found a significant main effect of interference condition on liking ratings, F(2,118) = 6.03, p = .003, η² = 0.021. Follow-up paired-samples t-tests (with Holm-Bonferroni correction for multiple comparisons) found that liking ratings were lower in the whispering condition than in the no-interference condition, t(119) = −3.44, p = .002, d = 0.31, and in the tapping condition, t(119) = −2.88, p = .009, d = 0.26; however, there was no difference between liking ratings in the no-interference condition and in the tapping condition, t(119) = 0.76, p = .45, d = 0.070. There was no main effect of timbre on liking ratings, F(1,59) = 0.23, p = .63, η² = 0.001, nor was there an interaction between interference condition and timbre on liking ratings, F(2,118) = 1.30, p = .27, η² = 0.002.

4. Discussion

In the current study, we investigated whether the established memory advantage for vocal melodies over non-vocal melodies is attributable to stronger simulation of vocal melodies relative to non-vocal melodies. To test this idea, we compared recognition performance for vocal and piano melodies in different interference conditions during encoding, including one in which simulation of singing was disrupted...
(whispering) and two conditions in which simulation was left intact (tapping, no-interference). Our expectation was that vocal-motor interference during whispering would result in a reduction in the vocal-memory advantage compared to non-vocal motor interference (tapping) or to the absence of a motor task (no-interference).

4.1. Recognition scores for melodies

In line with our predictions, there was an interaction between interference condition and timbre on melody recognition scores. Follow-up t-tests revealed a significant vocal-memory advantage in the no-interference and tapping conditions, but no significant vocal-memory advantage was found in the whispering condition.

The vocal-advantage in the no-interference condition is a replication of past studies showing that vocal memories are recognized more accurately than non-vocal melodies in a recognition test that follows the encoding phase of the experiment (Weiss et al., 2012). The lack of a vocal-advantage in the whispering condition suggests that vocal-motor interference was successful, disrupting or eliminating the motor memory code that would otherwise be created for the incoming stream of vocal information. Because our recognition scores were based on confidence condition, it is possible that the initial vocal-memory advantage in the no-interference condition, or its elimination in the whispering condition, simply reflected changes in memory confidence rather than genuine changes in memory itself. To address this, we completed an additional analysis using d-prime scores, following the methods of Schellenberg and Habashi (2015). This analysis revealed that the vocal advantage persisted in the no-interference condition, with no advantage present in the whispering condition.

Importantly, the vocal-memory advantage also persisted in the tapping condition, which we assumed to be equivalent to whispering in terms of its cognitive demand (Gupta & MacWhinney, 1995; Larsen & Baddeley, 2003) while avoiding the disruption of vocal-motor activity thought to underpin the vocal-memory advantage. The assumption of equivalent cognitive demand was supported by the similarity of recognition scores between tapping and whispering conditions for the piano timbre. Taken together, these findings suggest that the whispering eliminated the vocal-memory advantage specifically because it disrupted vocal-motor activity, and not because of issues related to cognitive demand.

Another potential explanation for the observed findings is that whispering interfered with the perception of vocal melodies. If there was a substantial amount of spectral overlap between the vocal melodies and the participants’ whispering, then vocal melodies may not have been perceived as well in the whispering condition compared to the tapping and no-interference conditions (Ehmer, 1959; Moore & Glasberg, 1981). This could be true even though participants rated tapping and whispering as equivalent in their perceived loudness. We attempted to minimize perceptual masking by asking participants to whisper and tap as quietly as possible during the exposure phase, and by playing melodies via over-ear headphones at a comfortable level (70 dB SPL). However, to assess whether perceptual masking could have led to elimination of the vocal-advantage in the whispering condition, we completed an additional analysis with 38 participants who either (1) answered no to whether they could hear themselves whisper or tap, or (2) reported that they could not hear themselves in the whispering condition. This analysis revealed that the vocal-advantage persisted in the no-interference condition, t(37) = 2.84, p = 0.007, yet no advantage was present in the whispering condition, t(37) = 0.38. As such, it seems highly unlikely that the elimination of the vocal-advantage in the whispering condition was due to perceptual masking of singing.

The tapping condition had a weaker vocal-advantage (d = 0.22) than the no-interference condition (d = 0.33). One possible interpretation of this finding is that tapping led to partial vocal-motor interference. Although the participants were using their hands rather than their voice, which likely activate distinct and non-overlapping regions of primary motor cortex (i.e., hand vs. larynx; see Fig. 2, Lotze et al., 2000), it is possible that tapping led to a spread of activation, leading to partial suppression of a vocal-motor memory code. For instance, pre-verbal infants display an overt natural coupling between rhythmic babbling and upper limb activity that has been interpreted as arising from spread of activation (Iverson & Thelen, 1999). From this spread of activation account, we might speculate that the vocal-advantage would have been stronger in the tapping condition had we instructed participants to tap with their foot rather than their hand. The foot region lies at the opposite end of primary motor cortex and may therefore be less likely to interfere with motor activity in laryngeal areas. However, this account seems somewhat unlikely in view of inhibitory transcranial magnetic stimulation (TMS) studies that have selectively targeted articulatory and hand areas of primary motor cortex during speech perception. For instance, inhibitory TMS over the lip area has been shown to influence the discriminability of near-neighbor phonemes, whereas TMS over the hand area does not (Möttönen & Watkins, 2009; Smalle, Rogers, & Möttönen, 2015). Another explanation for the weaker vocal-advantage in the tapping condition compared to the no-interference condition is that attentional resources were divided between the two concurrent tasks (tapping and listening to melodies), resulting in less resources being allocated to sensorimotor simulation. Indeed, studies have shown that the divided attention resulting from performing two concurrent tasks leads to a decrease in the capacity to simulate (Bach, Peatfield, & Tipper, 2007; Chong, Cunnington, Williams, & Mattingley, 2009; Gowen, Bradshaw, Galpin, Lawrence, & Poliaffk, 2010; Puglisi, Leonetti, Cerri, & Borroni, 2018; Puglisi et al., 2017).

In line with our predictions, our analysis also revealed a main effect of interference condition, with superior melody recognition performance in the no-interference condition compared with tapping and whispering conditions. This finding can be interpreted with respect to cognitive demand. The no-interference condition was the only condition that did not involve a demanding secondary task, which would have enabled more resources to be allocated to encoding. Supporting this interpretation, speech perception studies have found that the more cognitive resources are challenged during listening, the worse recognition is for speech in a subsequent memory test (Pichora -Fuller, Schneider, & Daneman, 1995).
One further consideration is that the observed vocal-memory advantage was much smaller in the current study compared with prior studies that have documented it, even in the no-interference condition. For example, in Weiss et al. (2017), the mean melody recognition score for voice was just over 3 and the mean score for piano was just under 2, with a corresponding effect size of $d = 0.92$. By comparison, the no-interference condition of the current study found a mean melody recognition score for voice of 1.86 and a mean score for piano of 1.41, with a corresponding effect size of $d = 0.33$. The reduced vocal-memory advantage observed in the present study may have been due to the mere presence of a secondary task on some trials in the exposure phase. The presence of a secondary task may have led participants to focus less attention on melodies regardless of interference condition. This reduced attention may have attenuated the strength of the motor memory code, and consequently, the extent of the memory advantage that is normally observed for vocal melodies.

Like prior studies, the vocal-memory advantage was not found to be dependent on musicality (Weiss, Vanzella, et al., 2015b). One interpretation of this finding is that the occurrence of sensorimotor simulation depends on general experience with vocal-motor control (including speech) rather than specific experience producing vocal melodies. Indeed, even individuals with deficits in pitch perception (amusia) show a vocal-memory advantage (Weiss & Peretz, 2019). However, unlike previous studies, we did not find an effect of musicaity on overall memory for melodies (Weiss, Vanzella, et al., 2015b). This may be because our participants did not have the same range of musical ability as did the participants in Weiss, Vanzella, et al. (2015). We measured general musical competency in our sample using the Gold-MSI, whereas Weiss, Vanzella, et al. (2015) used a design in which highly-trained musicians were compared with non-musicians.

It is worth noting that Weiss and colleagues previously considered a simulation account of the vocal-memory advantage, reasoning that, if sensorimotor memory traces can support melody recognition, highly trained piano players should show a piano-memory advantage. However, trained pianists do not show a piano-memory advantage, but instead show a vocal-memory advantage equivalent to that observed in non-musicians (Weiss, Vanzella, et al., 2015b). Nevertheless, these results obtained with pianists are consistent with results of prior studies that have examined sensorimotor simulation in highly trained musicians during perception of tones produced from their primary instrument. These prior studies have shown that simulation only occurs during perception of previously learned motor sequences, rather than novel sequences (D’Ausilio, Altenmüller, Olivetti Belardinelli, & Lotze, 2006; Lahav, Saltzman, & Schlaug, 2007; Novembre et al., 2014). In contrast, simulation of singing appears to be spontaneous, as it occurs in response to unknown melodies (Lévêque & Schon, 2013; Lévêque & Schön, 2015). This spontaneous simulation of vocal timbres, but not non-vocal timbres, may be owed to the vast and early experience that humans have integrating vocal perception and production, which would be more extensive than experience had with musical instruments. Thus, because the pianists were not familiar with the piano melodies in Weiss, Vanzella, et al.’s (2015) experiment, they did not possess the motor plans necessary to execute them. As such, spontaneous sensorimotor simulation of these piano melodies would have been unlikely despite the pianists’ training.

4.2. Liking ratings for melodies

We found a main effect of interference condition on liking ratings such that melodies in the whispering condition were rated lower compared with the other two conditions. Although we did not expect this result, it raises the question of whether the effects of vocal-motor interference extend beyond memory to other aspects of higher-order processing, such as the formation of preferences. One possibility is that interfering with the motor system during perception causes it to feel less fluent or familiar, therefore decreasing liking (Peretz, Gaudreau, & Bonnel, 1998). In any case, because there was no interaction between interference condition and timbre on liking ratings, the vocal-memory advantage is unlikely to be attributable to preferential liking. This is consistent with the conclusions of prior studies (Weiss & Peretz, 2019; Weiss, Schellenberg, et al., 2015a; Weiss et al., 2016; Weiss, Vanzella, et al., 2015b).

4.3. Implications and future directions

Overall, these results provide evidence that the memory advantage for vocal melodies is attributable to stronger sensorimotor simulation during the perception of vocal melodies compared to non-vocal melodies. This work adds to a body of literature demonstrating that simulation can support memory performance (Apel et al., 2012; Baumeister et al., 2015; Decloe & Obhi, 2013; Downing-Doucet & Guérard, 2014; Guérard & Lagacé, 2014; Naish et al., 2016; Shebani & Pulvermüller, 2013), similar to the effects of actual action execution (Engelkamp & Zimmer, 1989; MacLeod et al., 2010; Masumoto et al., 2006; Wammes et al., 2016). Furthermore, the results are in line with several studies showing evidence for spontaneous sensorimotor simulation during the perception of singing (McGarry et al., 2015; Pruitt et al., 2019; Royal et al., 2015), as well as stronger simulation during perception of vocal compared to non-vocal melodies (Lévêque et al., 2013; Lévêque & Schon, 2013; Lévêque & Schön, 2015; Whitehead & Armony, 2018).

We propose that the underlying neural basis of singing simulation is in the AON and primary motor cortex, which have been shown to be active during both production of singing as well as passive perception of singing (Caillain et al., 2006; Lévêque & Schon, 2015). More specifically, this would include activation of laryngeal-phonation premotor areas and, as a consequence, laryngeal-phonation primary motor areas (Brown et al., 2008). This activation of the motor system during the perception of singing may provide a vocal-motor memory code in addition to the auditory code that is also provided by the auditory system during the perception of non-vocal melodies. Primary motor cortex may be particularly important for providing this motor code, as suggested by TMS studies. For instance, Decloe and Obhi (2013) as well as Naish et al. (2016) showed that inhibitory TMS applied over areas of primary motor cortex corresponding to simulation of perceived action impaired participants’ recognition of stimuli associated with the action, indicating that motor cortical processing contributes to the formation of memory representations. Similarly, the formation of a vocal-motor memory code in the primary motor cortex is likely the basis of the memory advantage for vocal melodies. In the current study, we were able to eliminate this through the use of a vocal-motor interference task (i.e., whispering) that occupied the laryngeal-phonation primary motor areas. In turn, this likely disrupted the simulation of singing as well as the motor memory code it provides. In this case, vocal melodies, like non-vocal melodies, would only have an auditory memory trace to facilitate recognition, therefore eliminating their inherent memory advantage.

Another way of interpreting spontaneous sensorimotor simulation of singing is in terms of verbal working memory. The AON largely overlaps, structurally and functionally, with the dorsal stream of auditory processing (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). This processing stream is responsible for sensorimotor integration (e.g., during speech perception and production), and it is likely also the basis of verbal working memory (Hickok & Poeppel, 2000, 2004). According to this view, the superior temporal gyrus stores phonological information in short-term memory, the IFG maintains it via rehearsal, and area Spt—a part of the IPL,—serves as the auditory-motor integrator between these areas (Buchsbaum & D’Esposito, 2019; Hickok, Buchsbaum, Humphries, & Mufﬂuer, 2003). Thus, because singing is better able to spontaneously engage the vocal-motor system than non-vocal music (e.g., Lévêque & Schon, 2013, 2015), it may be more robustly represented in verbal working memory. This could in turn lead to...
a stronger representation of sung melodies in long-term memory and, consequently, better recognition (i.e., the memory advantage for vocal melodies). In this scenario, whispering eliminates this vocal-advantage by occupying verbal working memory, similar to the effect of articulator suppression.

It should be acknowledged that the current study did not measure the effects of sensorimotor simulation directly, but inferred it from our vocal-motor interference manipulations. The role of sensorimotor simulation in memory representations of actions could be assessed more directly in future work using brain imaging or stimulation techniques. For instance, further studies are necessary to confirm the exact neural substrates involved in spontaneous simulation of singing, as well as how simulation can support encoding and retrieval. A follow-up could also investigate whether simulation of laryngeal motor or premotor areas during perception of melodies is able to eliminate the vocal-memory advantage observed in a subsequent recognition task.

5. Conclusion

Spontaneous activity in motor cortex has been observed widely during the perception of a variety of human-produced actions in the absence of performing the action. This preferential engagement has been conceptualized as a sensorimotor simulation of the observed action because it represents actions by the motor commands necessary to actually execute the action. There is evidence that sensorimotor simulation is related to enhanced memory representations of objects related to action (Baumeister et al., 2015; Decloe & Obhi, 2013; Naish et al., 2016), but this effect is relatively unexplored in the auditory domain. In the current study, we investigated whether the source of the memory advantage for vocal melodies observed in several studies is stronger engagement of the motor system during the perception of vocal melodies relative to non-vocal melodies (sensorimotor simulation). Our results showed that a vocal-memory advantage was present in no-interference and tapping (non-vocal-motor interference) conditions, but was eliminated in the whispering (vocal-motor interference) condition. These results support the view that spontaneous sensorimotor simulation during the perception of vocal melodies is responsible for the observed vocal-memory advantage. We propose that the neural underpinnings of sensorimotor simulation, particularly primary motor activity, is able to provide an additional memory code for intentional actions, which can lead to an enhancement in the subsequent recognition of those actions.

Acknowledgments

We would like to thank Mike Weiss for graciously providing the stimulus files. We are also grateful to Elizabeth Earle for her tireless assistance in data collection and Fran Copelli for her support in setting up the experiment and in creating the graphical abstract.

Funding

Funding for this study was provided by a Discovery grant from the Natural Science and Engineering Research Council (2017-06969) awarded to F. A. Russo.

Declaration of Competing Interest

The authors declare no competing interest.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandc.2020.105622.

References

Alekman, A., & van’t Wout, M. (2004). Subvocalization in auditory-verbal imagery: Just a form of motor imagery? Cognitive Processing, 5(4), 228–231. https://doi.org/10.1007/s10339-004-0004-y.

Azel-Jad, K., Cangeltoli, A., Ellis, R., Goulain, J., & Fischer, M. H. (2012). Object affordance influences instruction span. Experimental Brain Research, 222(3), 199–206. https://doi.org/10.1007/s00221-012-2351-0.

Aziz-Zadeh, L., Sheng, T., & Gheytanchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. PLoS ONE, 5(11), 10.1371/journal.pone.0013853.https://doi.org/10.1371/journal.pone.0013853.

Bach, P., Peatfield, N. A., & Tipper, S. P. (2007). Focusing on body sites: The role of spatial attention in action perception. Experimental Brain Research, 178(4), 509–517. https://doi.org/10.1007/s00221-007-0764-4.

Barsalou, L. W. (2008). Grounded cognition. Annual Review of Psychology, 59(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.095639.

Baumeister, J. C., Rumitani, R. I., & Foroni, F. (2015). When the mask falls: The role of facial motor resonance in memory for emotional language. Acta Psychologica, 155, 29–36. https://doi.org/10.1016/j.actpsy.2014.11.012.

Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. Advances in Cognitive Psychology, 5(1), 69–83. https://doi.org/10.1080/15685304.2008.1006063.

Bernardi, N. F., Schories, A., Jabusch, H.-C., Colombo, B., & Altenmüller, E. (2013). Mental practice in music memorization: An ecological-empirical study. Music Perception: An Interdisciplinary Journal, 30(3), 275–290. https://doi.org/10.1525/mp.2013.30.3.275.

Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10(4), 433–436. https://doi.org/10.1163/156856897X00357.

Brown, S., Negan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. Cerebral Cortex, 18(8), 2037–2045. https://doi.org/10.1093/cercor/bhm131.

Caetano, G., Jousmäki, V., & Hari, R. (2007). Actor’s and observer’s primary motor cortex has been observed widely during the perception of a variety of human-produced actions in the absence of performing the action. This preferential engagement has been conceptualized as a sensorimotor simulation of the observed action because it represents actions by the motor commands necessary to actually execute the action. There is evidence that sensorimotor simulation is related to enhanced memory representations of objects related to action (Baumeister et al., 2015; Decloe & Obhi, 2013; Naish et al., 2016), but this effect is relatively unexplored in the auditory domain. In the current study, we investigated whether the source of the memory advantage for vocal melodies observed in several studies is stronger engagement of the motor system during the perception of vocal melodies relative to non-vocal melodies (sensorimotor simulation). Our results showed that a vocal-memory advantage was present in no-interference and tapping (non-vocal-motor interference) conditions, but was eliminated in the whispering (vocal-motor interference) condition. These results support the view that spontaneous sensorimotor simulation during the perception of vocal melodies is responsible for the observed vocal-memory advantage. We propose that the neural underpinnings of sensorimotor simulation, particularly primary motor activity, is able to provide an additional memory code for intentional actions, which can lead to an enhancement in the subsequent recognition of those actions.

Acknowledgments

We would like to thank Mike Weiss for graciously providing the stimulus files. We are also grateful to Elizabeth Earle for her tireless assistance in data collection and Fran Copelli for her support in setting up the experiment and in creating the graphical abstract.

Funding

Funding for this study was provided by a Discovery grant from the Natural Science and Engineering Research Council (2017-06969) awarded to F. A. Russo.

Declaration of Competing Interest

The authors declare no competing interest.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandc.2020.105622.
