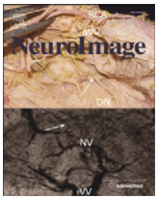
NeuroImage 67 (2013) 313–321

Contents lists available at SciVerse ScienceDirect

NeuroImage

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



Synchronizing with auditory and visual rhythms: An fMRI assessment of modality differences and modality appropriateness

Michael J. Hove [a],*, Merle T. Fairhurst [a], Sonja A. Kotz [a], Peter E. Keller [a,b]

[a] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[b] MARCS Institute, University of Western Sydney, Australia

ARTICLE INFO

Article history:
Accepted 26 November 2012
Available online 30 November 2012

Keywords:
Timing
Rhythm
Modality differences
Sensorimotor synchronization
Basal ganglia

ABSTRACT

Synchronizing movements with auditory beats, compared to visual flashes, yields divergent activation in timing-related brain areas as well as more stable tapping synchronization. The differences in timing-related brain activation could reflect differences in tapping synchronization stability, rather than differences between modality (i.e., audio-motor vs. visuo-motor integration). In the current fMRI study, participants synchronized their finger taps with four types of visual and auditory pacing sequences: *flashes* and a *moving bar*, as well as *beeps* and a *frequency-modulated 'siren'*. Behavioral tapping results showed that visuo-motor synchronization improved with moving targets, whereas audio-motor synchronization degraded with frequency-modulated sirens. Consequently, a modality difference in synchronization occurred between the discrete beeps and flashes, but not between the novel continuous siren and moving bar. Imaging results showed that activation in the putamen, a key timing area, paralleled the behavioral results: putamen activation was highest for beeps, intermediate for the continuous siren and moving bar, and was lowest for the flashes. Putamen activation differed between modalities for beeps and flashes, but not for the novel moving bar and siren. By dissociating synchronization performance from modality, we show that activation in the basal ganglia is associated with sensorimotor synchronization stability rather than modality-specificity in this task. Synchronization stability is apparently contingent upon the modality's processing affinity: discrete auditory and moving visual signals are modality appropriate, and can be encoded reliably for integration with the motor system.

Introduction

Precise temporal coordination between action and different perceptual systems is crucial for interacting with a dynamic environment. Precise visuo-motor integration is needed to catch a ball (or catch dinner) and audio-motor integration is needed to synchronize movements with music. Empirically, the temporal integration of action and perception is commonly examined in tasks requiring finger tapping to an isochronous pacing sequence. Previous neuroimaging and behavioral studies of tapping have established strong modality differences between audio-motor and visuo-motor synchronization. However, the vast majority of previous studies used only *flashing* visual stimuli, and flashes are known to yield poorer synchronization performance than auditory stimuli (Repp, 2005). Visuo-motor synchronization improves significantly with moving stimuli (Hove and Keller, 2010; Hove et al., 2010; Iversen et al., submitted for publication). In the present study, we tested whether previously observed activation differences reflect modality per se (i.e., audio-motor vs. visuo-motor integration), or differences in synchronization performance.

Neuroimaging studies have uncovered divergent neural activation patterns for visuo-motor versus audio-motor synchronization. Differences extend well beyond primary sensory areas into regions implicated in the brain's timing networks, including the basal ganglia, supplementary motor areas (SMA), and cerebellum (e.g., Buhusi and Meck, 2005; Coull et al., 2011; Macar et al., 2002; Schwartze et al., 2012). Direct comparisons of audio-motor and visuo-motor synchronization reported activation in different areas of the cerebellum (Jäncke et al., 2000; Penhune et al., 1998). Additionally, audio-motor, but not visuo-motor synchronization, yielded significant activation in the SMA (Jäncke et al., 2000; Penhune et al., 1998). In a meta-analysis on 38 neuroimaging studies of finger-tapping, striking differences between audio- and visuo-motor synchronization were uncovered in the putamen of the basal ganglia; synchronization with auditory, but not visual stimuli, consistently activated the putamen (Witt et al., 2008). The putamen is a key area for beat and rhythm processing (Coull et al., 2011; Grahn and Rowe, 2009; Kotz et al., 2009; Teki et al., 2011; Wiener et al., 2009). A recent study comparing audio and visual beat perception showed more putamen activation and more sensitive beat perception for auditory than for visual stimuli; nevertheless, within the visual condition the degree of putamen activation predicted beat sensitivity (Grahn et al., 2011).

---

* Corresponding author at: Max Planck Institute for Human Cognitive and Brain Sciences, Research Group "Music Cognition and Action", Stephanstrasse 1a, 04103 Leipzig, Germany. Fax: +49 341 9940 2614.
E-mail address: michaeljhove@gmail.com (M.J. Hove).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2012.11.032

Taken together, neural activation differences observed between auditory and visual modalities in synchronization and beat perception have important implications. For example, it has been argued that auditory rhythms induce an internal rhythm that guides movement, whereas visual rhythms do not generate an internal rhythm (Jäncke et al., 2000). Additionally, modality differences have provided evidence that time is represented in a distributed network rooted in sensorimotor processes, rather than subserved by a centralized clock mechanism (Jantzen et al., 2005). Furthermore, differences in neural activations could support an auditory specialization for encoding temporal information (e.g., Welch and Warren, 1980).

Given these activation differences in timing circuits, it is perhaps unsurprising that a strong behavioral advantage has also been observed for audio-motor over visuo-motor synchronization. Rhythmic finger tapping is much more accurate with auditory stimuli than with flashing visual stimuli (e.g., Chen et al., 2002; Dunlap, 1910; Kolers and Brewster, 1985). Stable synchronization is possible at much faster rates with auditory than with visual sequences (Repp, 2003). In a target-distracter paradigm, when auditory beeps and visual flashes are presented in competition with each other, participants' movement timing is dictated by the auditory stimuli, regardless of volition (Repp and Penel, 2004). Finally, the serial dependence between inter-tap intervals, which intimates underlying timing processes (e.g., Vorberg and Wing, 1996), differs between audio- and visuo-motor synchronization: Inter-tap intervals in audio-motor synchronization typically alternate between short and long intervals (a negative lag1 autocorrelation), which suggests active error correction (Semjen et al., 2000); whereas synchronization with flashing visual stimuli typically has a positive or non-negative lag1 autocorrelation, which suggests weak (or absent) tap-to-tap error correction (Chen et al., 2002; Hove and Keller, 2010; Hove et al., 2010). Together these results suggest different underlying processes for synchronizing with audio- versus flashing visual sequences.

However, nearly all imaging and behavioral evidence for differences between visual and auditory synchronization used *flashing* visual stimuli. While flashes may offer the most similar control for auditory beeps in terms of temporal onset/offset (and no additional confounding factors), they lack ecological validity in that the visual system rarely processes or acts upon purely temporal information devoid of spatial translation. The visual system has considerably lower temporal resolution than audition (e.g., Holcombe, 2009), and thus is severely handicapped in synchronizing with discrete temporal stimuli. Vision excels at processing spatial, rather than temporal information (e.g., Bertelson and Radeau, 1981; Welch and Warren, 1980). When a visual stimulus contains spatiotemporal information (rather than purely temporal information), action timing to intercept that moving stimulus can be very precise (Bootsma and van Wieringen, 1990).

In a series of recent finger-tapping studies, we have shown that synchronization timing improves dramatically with spatiotemporal visual stimuli, compared with purely temporal flashing stimuli. In one study, participants tapped along with flashing visual stimuli and with visual images that alternated between a high and low position creating apparent motion. Synchronization was considerably more stable with the apparent motion stimuli than the flashes (Hove and Keller, 2010). In another study, participants tapped along with visual flashes, fading stimuli, and stimuli that moved frame-by-frame at a linear velocity, as well as an auditory metronome. Synchronization with the moving stimuli was much better than with flashing or fading stimuli; however, an auditory advantage was still observed, especially at very fast tempi (300 and 240 ms IOI), (Hove et al., 2010). In both these studies, the moving visual stimuli also yielded *negative* lag1 autocorrelations, suggesting that error correction was occurring. Additionally, in a target-distracter study that presented moving visual stimuli in competition with auditory beeps, the moving visual stimuli attracted movement timing as much as auditory stimuli, thus erasing the auditory dominance previously observed over flashes (Hove et al., in press). Together, these studies demonstrate that motion increases the temporal reliability of visual encoding (cf. Ernst and Bülthoff, 2004) and thus facilitates precise visuo-motor integration.

The foregoing suggests that the 'modality differences' observed between auditory and flashing visual stimuli should be interpreted cautiously: It is unclear if the previously reported differences in brain activation truly reflect differences between *modality* (i.e., audio-motor vs. visuo-motor integration) or simply the poor performance with flashing visual stimuli due to their less precise temporal encoding. The significant improvement in visual synchronization with moving stimuli encourages the re-examination of established modality differences. Are differences in neural timing circuits substantially reduced with improved visuo-motor synchronization (or degraded audio-motor synchronization)?

In the present fMRI study, participants synchronized their finger taps with four types of visual and auditory pacing sequences: visual stimuli were *flashes* and a *moving bar*, and auditory stimuli were *beeps* and *frequency modulated 'sirens'*. Within the visual modality, synchronization should improve for the moving bar stimuli compared to the flashes due to the spatial processing advantage. Within the auditory modality, synchronization should degrade for the siren compared to the discrete beeps, since the siren's continuous presentation should reduce or blur the neural encoding of its target compared to the beep's discrete target (cf. Barsz et al., 2002). Thus, these stimuli can disentangle synchronization performance from modality. Critically, neural activation in key timing areas such as the putamen should vary with the stability of synchronization performance, rather than being dictated by modality. We expect to replicate previously observed modality differences only for the discrete stimuli (beeps versus flashes), whereas the modality differences should be substantially less pronounced with the continuous stimuli (siren versus moving bar). Accordingly, we anticipate an interaction between modality and discrete/continuous stimulus structure for both behavioral synchronization performance and neural activation in time-sensitive areas such as the putamen.

Material and methods

Participants

Fourteen right-handed volunteers (7 women) aged 24 to 34 years ($M = 27.7 \pm 3.0$ years) participated in the experiment.[1] Participants were paid for their participation and gave informed consent. Participants had a range of musical training ($M = 7.9$ years; $SD = 9.5$); this did not affect tapping performance ($ps > .5$).

2.2. Experimental protocol

Participants lay supine in the fMRI scanner and tapped their right index finger on hard plastic surface embedded in a custom MR-compatible air-pressure response device. Participants were instructed to synchronize their finger taps along with four different types of isochronous pacing sequences at two tempi. The study employed a 2 (modality: auditory, visual) × 2 (style: discrete, continuous) × 2 (tempo: slow 600 ms IOI, fast 400 ms IOI) within-subjects design (see Fig. 1). Including two tempi ensured that participants attended to the stimuli and did not simply tap along with one memorized tempo. Auditory pacing sequences were presented over headphones (MR Confon, Magdeburg, Germany) and visual pacing sequences were presented via a projector (SANYO PLC-XP50L). Taps were recorded and pacing sequences were presented via a PC running Presentation software (Neurobehavioral Systems). The presentation program syncs to the refresh rate and allows consistent timing in

---

[1] Two additional volunteers participated in the experiment, but failed to synchronize with the pacing sequences in more than half of the trials in the scanner; therefore they were excluded and their imaging data were not analyzed.