



## Cues for self-recognition in point-light displays of actions performed in synchrony with music

Vassilis Sevdalis\*, Peter E. Keller

*Music Cognition and Action Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*

### ARTICLE INFO

#### Article history:

Received 7 July 2009

Available online 9 April 2010

#### Keywords:

Action perception

Movement kinematics

Self-recognition

Audiovisual perception

Music

### ABSTRACT

Self-other discrimination was investigated with point-light displays in which actions were presented with or without additional auditory information. Participants first executed different actions (dancing, walking and clapping) in time with music. In two subsequent experiments, they watched point-light displays of their own or another participant's recorded actions, and were asked to identify the agent (self vs. other). Manipulations were applied to the visual information (actions differing in complexity, and degradation from 15 to 2 point-lights within the same clapping action) and to the auditory information (self-generated vs. externally-generated vs. none). Results indicate that self-recognition was better than chance in all conditions and was highest when observing relatively unconstrained patterns of movement. Auditory information did not increase accuracy even with the most ambiguous visual displays, suggesting that judgments of agent identity depend much more on motor cues than on auditory (action-generated) or audiovisual (synchronization) information.

© 2010 Elsevier Inc. All rights reserved.

### 1. Introduction

Most of the actions humans perform rely on multisensory information from the environment. For example, music performance and perception are activities that occur in multisensory contexts. Music making requires precise timing, coordination and motor control for planning and executing one's own movements, and for predicting the intentions and actions of others when playing in ensembles (Keller, 2008). In order to act appropriately, concurrent information from different sensory modalities must be efficiently processed simultaneously in space and time. Music listening is often accompanied by spontaneous body movements, even in young children. In everyday situations, the synchronization of body movements with music is a common activity, even for people without formal training, whether they are dancing, marching, or simply clapping in time during a concert. Does this contextual multisensory information help a person to know about his or her actions and the actions of others, or is the body movement alone sufficient? This study examines whether the dynamic relationship between an individual's movements and externally generated auditory information, as well as auditory information generated by the action itself, provide cues for self-recognition during the observation of impoverished visual displays of actions performed in synchrony with music.

It has been proposed that people understand their own and others' actions by means of action simulation, that is by mapping observed movements onto their own action system (Jeannerod, 2006; see also the related principle of common coding, Hommel, Müssele, Aschersleben, & Prinz, 2001). Evidence from single cell recordings in macaque monkeys (Gallese, Fadiga,

\* Corresponding author. Address: Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1A, 04103 Leipzig, Germany. Fax: +49 341 99 40 113.

E-mail address: [sevdalis@cbs.mpg.de](mailto:sevdalis@cbs.mpg.de) (V. Sevdalis).

Fogassi, & Rizzolatti, 1996) and imaging studies of humans (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Grèzes, Armony, Rowe, & Passingham, 2003) suggests that there is a close correspondence between neural activity while observing, imagining or executing the same action (for a review, see Rizzolatti & Craighero, 2004). Furthermore, embodied simulation and embodied cognition accounts (Gallese, 2007; Wilson, 2002) claim that social cognition and action understanding are grounded in the behavioral matching and neuronal overlap between action execution and action perception.

Behavioral evidence for a close relationship between perception and action has been obtained in a number of studies of self-recognition (for a review, see Knoblich, 2008). Self-recognition has been investigated mainly by focusing on either visual or auditory modalities. In the visual modality, designs employing observation of hand gestures (Daprati, Wriessnegger, & Lacquaniti, 2007), drawing movement trajectories (Knoblich & Prinz, 2001) and various full body movements (Loula, Prasad, Harber, & Shiffrar, 2005) have been used for investigating self-recognition. In the auditory modality, self-recognition has been examined by having individuals listen to the sounds of their own or others' clapping (Flach, Knoblich, & Prinz, 2004) and by having piano players discriminate between their own and others' musical performances (Repp & Knoblich, 2004). Studies of overt coordination with self- vs. other-generated stimuli (Flach, Knoblich, & Prinz, 2003; Keller, Knoblich, & Repp, 2007) have provided further support for the claim that the perception of agent identity is grounded in motor processes, specifically, in action simulation (Jeannerod, 2003, 2006). For example, in the music domain, Keller et al. (2007) found that pianists were not only able to recognize their own performances, but were also able to synchronize better when playing duets with their own previous recordings than with another pianist's recordings. Thus, visual and auditory cues to agent identity are provided by the idiosyncratic ways in which individuals move due to personal biomechanical constraints and past experience and training.

Neurophysiological and neuroimaging evidence suggests that the perception–action links are implemented on a neuronal level (Blakemore & Decety, 2001; Decety & Grèzes, 1999; Jeannerod, 2001; Rizzolatti & Craighero, 2004). In the domains of music (Bangert & Altenmüller, 2003) and dance (Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005; Cross, Hamilton, & Grafton, 2006; Cross, Kraemer, Hamilton, Kelley, & Grafton, 2009), it has been shown that perception–action links become more robust as a function of training and expertise, leading to stronger activation of brain areas that are related to action planning and motor performance in experts relative to novices (Haslinger et al., 2005; Hauelsen & Knösche, 2001). Additional evidence for shared auditory and motor processing networks in musical activities comes from studies by Lahav, Saltzman, and Schlaug (2007), Lindenberger, Li, Gruber, and Muller (2009), Mutschler et al. (2007) and Zatorre, Chen, and Penhune (2007).

Self-recognition studies employ various methods, paradigms and designs, including online (concurrent) vs. offline (delayed) tasks, full body movements vs. fine movements of body effectors and goal-directed vs. nongoal-directed actions. Furthermore, self-recognition has been studied by using point-light displays (Loula et al., 2005), the rubber hand illusion paradigm (Botvinick & Cohen, 1998), neuroimaging (for reviews see Legrand & Ruby, 2009; Lenggenhager, Smith, & Blanke, 2006) and more recently, virtual reality paradigms (Lenggenhager, Tadi, Metzinger, & Blanke, 2007). The main methodological technique in self-recognition experiments is to observe and compare self- vs. other-generated action signals. The crucial question is how signals that stem from oneself or another person are monitored and used in order to disambiguate the identity of the bodies and the origin of the actions.

The interplay between efferent and afferent signals is considered a key factor in the sense of agency – the experience that oneself is the cause of an ongoing action (Gallagher, 2000, 2005). Recent evidence suggests that the sense of agency may depend especially on efferent (motor) signals from one's own actions (Engbert, Wohlschläger, & Haggard, 2008; Tsakiris & Haggard, 2005; Tsakiris, Haggard, Franck, Mainy, & Sirigu, 2005). Although efferent information can modulate awareness of an action's timing, the sensory processing of self-generated events and action attribution (Tsakiris & Haggard, 2005), the role of afferent signals in the conscious phenomenal experience of agency remains unclear (Tsakiris, 2008). The meaning of afferent signals for perception is ambiguous because afferent signals may be either self- or externally-generated (i.e., refferent or ex-fferent; see also Knoblich & Repp, 2009; Repp & Knoblich, 2007). However, if the pattern of afferent information is such that it seems familiar or evokes resonance in the observer's motor system, it may be sufficient for self-recognition.

Most of the studies that aimed at investigating the role of motor cues in on-line judgments of agency (e.g., Daprati et al., 1997; Sirigu, Daprati, Pradat-Diehl, Franck, & Jeannerod, 1999; Van den Bos & Jeannerod, 2002) have used simple actions of specific body effectors (e.g., finger flexion, hand tapping). By contrast, studies of off-line self-recognition often have used more complex, whole-body actions. Even when human action is depicted by just a few point-lights (Johansson, 1973; for a review, see Blake & Shiffrar, 2007), observers can identify themselves, their friends or strangers from these point-light movement trajectories (Beardsworth & Buckner, 1981; Cutting & Kozlowski, 1977; Jokisch, Daum, & Troje, 2006; Loula et al., 2005; Prasad & Shiffrar, 2009). When walking movements were used, participants were able to recognize themselves and their friends equally well (Cutting & Kozlowski, 1977) and there was only a small advantage for self vs. friend recognition in a study by Beardsworth and Buckner (1981). In both studies, however, recognition rates were barely above chance. In the Loula et al. (2005) study, where a number of different actions were displayed, self-recognition was better than recognition of a friend's movements, although agent recognition was still not above chance for some actions such as walking. Recognition accuracy was action dependent, with the rich kinematic information contained in actions such as dancing affording a higher degree of recognition. In the Jokisch et al. (2006) study, distinction of self and other in walking patterns was achieved, although recognition for other's actions deteriorated when the actions were seen in profile rather than in a frontal viewpoint. By applying viewpoint manipulations, Prasad and Shiffrar (2009) obtained results that led them to conclude that

visual experience with one's own movements is of less importance in self-identification, which supports the hypothesis that motor resonance in action perception enables self-recognition.

The current study investigates self-recognition of actions performed by non-musicians in synchrony with music. Coordination with music may carry information about personal action styles in the way in which movements are timed relative to the music's rhythmic structure. Listening and/or moving to music may be like interacting with a virtual partner (Leman, 2007), and this interaction may be at the core of musical experience (Overy & Molnar-Szakacs, 2009). In research conducted outside the music domain, contextual cues (e.g., scenes surrounding the action, presence of another agent) have been previously shown to have an impact on the way actions are both perceived and performed (Iacoboni et al., 2005; Sebanz, Knoblich, & Prinz, 2005). By implementing a research paradigm that combines visual (point-light) displays of full body movements with auditory information (self- or externally-generated), we aim at elucidating the influence of contextual multisensory cues on judgments of agent identity, which have been previously studied in either the visual or the auditory modality alone. Our goal is therefore to assess self-recognition in more ecologically valid, multisensory contexts, which are typical for music-related behaviors. Furthermore, the actions considered are representative of realistic behaviors that vary in complexity or, more specifically, in the degrees of freedom that constrain movement kinematics: dancing, walking, and clapping in time with music.

We hypothesized that, in addition to intrinsic information that stems from an agent's movements in space and time, extrinsic information about the relation of movements to the music may potentially affect self-recognition performance. In other words, by examining self-recognition in point-light action displays presented together with auditory information, we test the hypothesis that contextual information about synchrony between movements and music improves recognition relative to information about movement kinematics alone. Specifically, music provides a temporal reference frame against which movement timing can be gauged. If such information were to contribute to self-recognition, it would be most likely to do so with relatively constrained actions (walking and clapping) whose agents were often not reliably identified in previous studies. This would be consistent with the inverse effectiveness principle (Alais & Burr, 2004; Meredith & Stein, 1983), which states that an observer's perception can compensate for the reduced quality of a stimulus in one modality by making use of information in another modality. Based on results of the study by Loula et al. (2005) and basic psychophysical premises (e.g., Weber–Fechner law), we assumed that recognition would be at a high level for dancing even in the visual-only condition, leaving little room for boosted performance due to additional sources (e.g., synchronization cues).

## 2. Experiment 1

The main aim of Experiment 1 was to examine whether contextual multisensory information has an impact on self-recognition. With regard to visual information, based on the results of previous research (Loula et al., 2005), we expected that self-other discrimination would improve with decreasing constraints on movement, and, therefore, would be most accurate for dancing and least accurate for clapping. With regard to contextual information, we wanted to examine whether concurrent auditory information about the dynamic relationship between an individual's movements and music provides informative cues for self-recognition.

### 2.1. Method

#### 2.1.1. Participants

Fourteen adults (8 females; aged 21–29; mean age: 24.2 years) participated in the study in return for financial compensation. All of them reported to have normal hearing and normal or corrected to normal vision. None of them had previous experience with point-light displays. Participants did not know each other and were not informed about the experimental hypotheses.

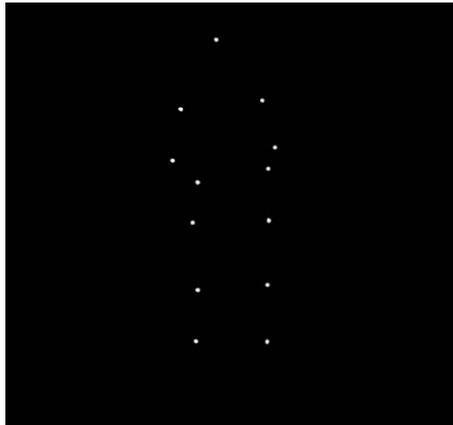
#### 2.1.2. Design

The experiment was carried out in two sessions, in a within subjects design, separated by 1–2 months. The sessions were named 'action session' and 'perception session'. The independent variables of interest in the perceptual experiment were action (dancing, walking, clapping) and modality (presence vs. absence of auditory information).

#### 2.1.3. Action session

**2.1.3.1. Materials.** Excerpts from three musical pieces of different styles ('drum and bass' dance, jazz and folk) were used as stimuli. The pieces were 'Apache', in a version performed by the Incredible Bongo Band, 'In the Mood' by the Glenn Miller Orchestra, and 'Madan' by Salif Keita. The excerpts were in .wav format, and their duration was 60–70 s. The excerpts had a clear beat and were selected on the basis that they would represent a variety of genres and that their tempi would allow the execution of all the three actions (clapping, walking and dancing) at a comfortable rate.

**2.1.3.2. Equipment and procedure.** Participants executed each of the three different actions in time with each of the three musical excerpts. Performances were recorded by a Vicon motion capture system (Vicon – Oxford, UK). Ten cameras were placed at relatively equal distances from the center of the room. Data acquisition with a sampling frequency of 200 fps



**Fig. 1.** Point-light depiction of an agent with 13 markers attached at the head and the main joints of the body.

was controlled by Vicon Nexus software. At the start of each musical excerpt, a digital signal was sent to a trigger panel, thus enabling the precise onset time of the audio signal to be recorded along with the motion capture data. Performances were also videotaped by a digital video camera (SONY HDR-HC9).

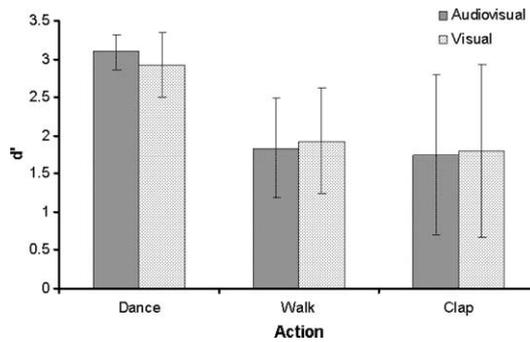
Participants came to the lab individually and were told that they would be required to clap, walk and dance in synchrony with the beat of three musical excerpts and that their movements would be recorded. They were informed that this experiment was not a test to judge their dancing abilities and that they should simply listen to the music and execute the actions in time to it. Music was used in order to exercise control over the timing of action execution. If participants were invited to clap or walk without an external referent (i.e., the beat of music), then their actions would presumably be much more diverse and variable in tempo. The treatment of the participants complied with all the ethical guidelines of the Max Planck Institute for Human Cognitive and Brain Sciences, where the research was conducted. All participants signed a consent form for their participation before the experiment began. Thirteen reflective markers were attached to the participants' bodies, at the head and at the main joints (shoulders, elbows, wrists, hips, knees and ankles) (see Fig. 1). For the clapping action, one extra marker was attached on the proximal phalanx of each index finger to enhance visibility and clarity of the clapping action.

Participants first listened to one musical excerpt to become familiar with it. Then they executed the three different actions, one at a time, for the entire duration of the excerpt. The desired time intervals of claps and walking steps were indicated by an experimenter and were the same for all the participants. The actions took place within the same starting positions and walking paths, which were clearly indicated by white tape on the floor (clapping: clapping while standing at the center of the room; walking: walking 5.50 m from the one end of the room to the other and then back; dancing: starting position at the center of the room and available dancing space of 280 × 280 cm). After all the actions had been executed for the first musical excerpt, the same procedure was repeated for the second and the third musical excerpts. The orders of the excerpts and of the actions were randomized. The duration of the action session did not exceed 1 h.<sup>1</sup>

#### 2.1.4. Perception session

**2.1.4.1. Materials.** Point-light movies were prepared using Final Cut Pro and QuickTime Software. The onset times of motion capture and audio data were synchronized by aligning the motion capture files of the agents in frontal view with the motion capture files of the trigger panel onset signals and with the audio files. From each combination of 3 actions × 3 musical excerpts, two 5-s excerpts were randomly selected and served as stimuli. The point-light movies displayed the performances of the three different actions (dancing, walking, and clapping) with or without the accompanying music (Audiovisual vs. Visual condition). Participants were matched in pairs according to gender and physical body proportions (i.e., for every 'self' there was just one 'other'). The matching process was carried out in the same way as in previously published studies (Loula et al., 2005). Matching participants according to gender and physical body proportions ensured that neither gender (Kozłowski & Cutting, 1977) nor weight (Runeson & Frykholm, 1983) could be used as the basis for participants' discriminations. Thus, seven matched pairs (4 female pairs and 3 male pairs) were created. In total, for every pair, 72 point-light movies were created that contained all the possible combinations of agents, actions, modalities, musical pieces, and different episodes of each action (2 agents × 3 actions × 2 auditory conditions × 3 pieces × 2 selections). From these movies, 36 movies depicted the 'self' and 36 movies depicted the 'other'. Additionally, 16 extra movies were created as practice trials (8 for 'self' and 8 for 'other'). The 5-s excerpts selected for practice trials were different from those in test trials.

<sup>1</sup> At the end of the session, participants were required to rate the familiarity of the musical pieces on a scale from 1 to 3 (anchors: unfamiliar–familiar). The dance piece received an average familiarity rating of 1.43 (SD = 0.76), the jazz piece 2.57 (SD = 0.76) and the folk piece 1.36 (SD = 0.63). In all subsequent analyses, no significant differences in performance were found across pieces and familiarity did not correlate with performance.



**Fig. 2.** Average  $d'$  scores in the two modality conditions for the three actions. Error bars indicate 95% confidence intervals.

**2.1.4.2. Equipment and procedure.** The same participants were invited back and watched the 5-s point-light movies on a computer monitor. The point-light figures were projected at 11.42 degrees of visual angle, their height was  $\sim 10$  cm and the viewing distance was about 50 cm for all the participants. The session started with the participants watching a block of eight practice trials (4 for 'self' and 4 for 'other', randomly selected from the 16 practice trials). Then each participant watched 144 point-light movies presented in random order (36 'self' and 36 'other' movies repeated across two blocks). Before each trial onset, a white fixation cross appeared at the center of the monitor, lasting for 1-s. The auditory information was delivered over two loudspeakers placed to the left and right in front of the participants, at a distance of 2 m. The task was to indicate by key press on a computer keyboard whether the depicted agent was oneself or another person. On the keyboard, the left arrow was labeled 'S' for 'selbst', indicating 'self', and the right arrow was labeled 'A' for 'andere', indicating 'other'. No feedback about correctness was provided after responses. The duration of the perception session did not exceed 1 h.

## 2.2. Results and discussion

Self-other discrimination was assessed by computing  $d'$ . This measure takes response bias into account by subtracting z-transformed false alarm rates ('self' judgments for 'other' displays) from hit rates (correct 'self' responses) (see Macmillan & Creelman, 1991). The results are shown in Fig. 2. Recognition accuracy was significantly better than chance ( $d' = 0$ ) for all actions: dance,  $t(13) = 22.60$ ,  $p < .001$ ; walk,  $t(13) = 6.25$ ,  $p < .001$ ; clap,  $t(13) = 3.54$ ,  $p < .01$ . In raw scores, when collapsed across all conditions and stimuli (modality conditions, musical pieces, and selections), correct recognition performance (i.e., 'self' as 'self' and 'other' as 'other') for the dancing action was 94.49% (SE = 3.09), for the walking action 79.46% (SE = 7.26), and for the clapping action 76.04% (SE = 9.28).

A  $3 \times 2$  repeated measures analysis of variance (ANOVA) was conducted on  $d'$  scores to test our hypotheses about the effects of Action (dance vs. walk vs. clap) and Modality (visual vs. audiovisual) on self-recognition. This analysis revealed a statistically significant main effect of Action,  $F(2, 26) = 3.36$ ,  $p = .05$ ,  $\eta_p^2 = .21$ . Paired  $t$ -tests revealed that the difference in performance between the dance and walk conditions was significant,  $t(13) = 3.91$ ,  $p < .01$ , while the difference between the walk and clap conditions was not significant,  $t(13) < 1$ , n.s. The main effect of Modality was not significant in the ANOVA, indicating no reliable differences in self-recognition between audiovisual and visual conditions,  $F(1, 13) < 1$ , n.s. The interaction between Action and Modality was likewise not significant,  $F(2, 26) < 1$ , n.s.

To summarize, recognition accuracy was highest for the most unconstrained movement patterns (i.e., for dancing), although agents were still recognized reliably in the case of the simplest action (clapping). The audiovisual presentation of point-light action patterns did not influence agent recognition for any action. Thus, the availability of additional cues about synchrony between movements and music did not improve recognition.

## 3. Experiment 2

The finding that auditory information did not boost performance in Experiment 1 suggests either that visual information about the action kinematics was sufficient for self-recognition or that there was no information to be gained from the relationship between the movements and the music. The information in the visual displays was quite rich because 13 point-lights (dancing and walking) or 15 point-lights (clapping) were used. Therefore, in the current experiment we tested whether agent recognition is still possible when parametric degradation is applied to the visual information. We hypothesized that auditory information might improve self-recognition when the visual information is severely impoverished.

Furthermore, another reason why auditory information did not boost performance in Experiment 1 may have been the fact that no action-specific auditory information was provided. Therefore, we added a condition in Experiment 2 in which another kind of auditory information was added, one that was specific to the actions (clapping sounds). Clapping sounds have been shown in past studies to contain cues to agent identity (Flach et al., 2004). Thus we now also tested the hypothesis

that self-generated sounds could serve as ancillary cues to agent identity, particularly when visual information is degraded. Clapping was the only action investigated in Experiment 2.

### 3.1. Method

#### 3.1.1. Participants

Twelve out of the fourteen adults that participated in Experiment 1 could be contacted and were invited back to participate in Experiment 2, which took place 2–3 months later.

#### 3.1.2. Design

The independent variables were the visual information available in the clapping displays (original full-body condition and three degradation levels) and the additional auditory information provided (none, music only, and music + clapping sounds).

#### 3.1.3. Materials

The clapping performances from the action session of Experiment 1 were used to prepare stimuli for Experiment 2. We introduced three degradation levels to the visual information available. In the degraded stimuli, only the body effectors that were involved in the execution of the clapping action were retained (i.e., the arms) and all the other markers were discarded. This degradation process resulted in four visual conditions (see Fig. 3): full body (15 markers), shoulders–elbows–wrists–fingers (8 markers), elbows–wrists–fingers (6 markers), and fingers only (2 markers).

Auditory information was also manipulated to produce two auditory conditions in addition to a silent condition. One condition contained only the music (dance, folk or jazz) that initially accompanied the action (Music) and the other contained the sounds of each participant's clapping along with the music (Music + Clap).

From each participant's clapping action, three 5-s excerpts were randomly selected to serve as point-light movies. In order to reduce the familiarity with the stimuli, these selected excerpts were different than those used in Experiment 1. The same matched participant pairs as in the previous experiment were used. In total, for every pair, 216 new point-light movies were created that contained all the possible combinations of agents, visual degradation conditions, auditory conditions, musical pieces, and different episodes of each action (2 agents  $\times$  4 degradation levels  $\times$  3 auditory conditions  $\times$  3 pieces  $\times$  3 selections). From these movies, 108 movies depicted the 'self' and 108 movies depicted the 'other'.

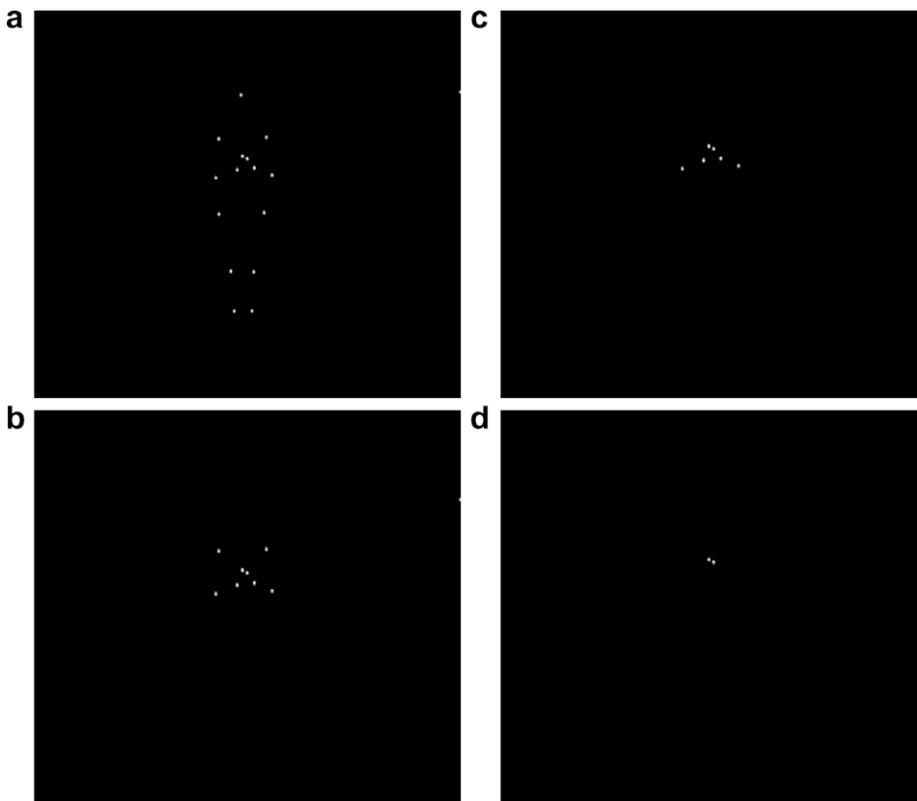
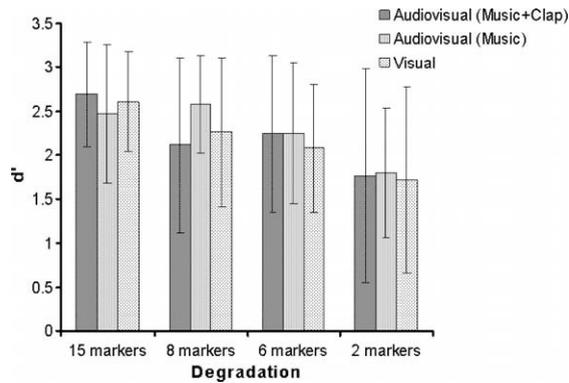


Fig. 3. Point-light depiction of an agent with 15, 8, 6 and 2 markers attached at the head, the main joints, and the fingers.



**Fig. 4.** Average  $d'$  scores in the three modality conditions for the four degradation levels. Error bars indicate 95% confidence intervals.

### 3.1.4. Equipment and procedure

The experiment started with the participants watching a block of 8 practice trials, selected randomly from 16 practice trials. In total, each participant watched 216 point-light movies on a computer monitor, presented in random order (108 'self' and 108 'other' movies). All other aspects of the procedure were as in Experiment 1.

### 3.2. Results and discussion

Self-other discrimination was again assessed by computing  $d'$ . The results are shown in Fig. 4. Two participants were excluded from the analyses due to failure to comply with the experimental instructions.<sup>2</sup> For the remaining participants, recognition accuracy was significantly better than chance for all degradation levels:  $ts(9) > 4.14$ ,  $p < .01$ .<sup>3</sup> In raw scores, when collapsed across all other conditions, correct recognition performance with 15 markers was 91.30% (SE = 6.26), with 8 markers 87.04% (SE = 7.55), with 6 markers 85.00 (SE = 8.14), and with 2 markers 78.70% (SE = 10.29).

A  $4 \times 3$  repeated measures analysis of variance (ANOVA) was conducted on the  $d'$  scores to test our hypotheses about the effects of visual Degradation (15 vs. 8 vs. 6 vs. 2 markers) and auditory information (none, music, music + clap) on self-recognition. This analysis revealed only a statistically significant main effect of Degradation,  $F(3, 27) = 6.25$ ,  $p < .01$ ,  $\eta_p^2 = .41$ . A significant linear trend across degradation levels confirmed the reliability of the decrease in recognition from 15 to 2 markers,  $F(1, 9) = 10.32$ ,  $p = .01$ . The main effect of auditory information was not significant in the ANOVA,  $F(2, 18) < 1$ , n.s., and the interaction was likewise not significant,  $F(6, 54) < 1$ , n.s.

To summarize, recognition accuracy for clapping agents decreased with decreasing availability of kinematic information, although agents could still be recognized reliably in the case of highly degraded displays (2 markers). Furthermore, agent recognition was reliable irrespective of whether self-generated auditory information (clapping sounds) or externally-generated sound (music) accompanied the actions. Therefore, self-other discrimination was achievable based only on kinematic cues, without any additional support from auditory cues.

## 4. General discussion

The aim of this study was to investigate the relative contributions of multisensory spatial-temporal cues that can be used for explicit judgments of agent identity. Our experiments show that self-recognition was based exclusively on visual information about personal movement kinematics, with contextual cues about synchrony between movements and sounds conferring no discernable benefit, even in impoverished visual conditions. The current findings extend research on the links between action execution and action perception to actions performed in synchrony with music. The sensitivity of the action-observation network may depend on the nature of the action: Relatively unconstrained movements and spatially unoccluded actions may be particularly potent due to richness in visual information about personal styles of action execution.

<sup>2</sup> These participants had very large, negative values of  $d'$ , while all other participants had positive values of  $d'$ . This deviant performance indicates that they reversed their responses, responding 'self' when they should have responded 'other', and vice versa. Although we could have simply reversed these participants' responses, because the absolute values of their discrimination sensitivity indices were high and hence their ability to distinguish agents per se was good, we considered it more prudent to exclude them. If the analyses are carried out with the reversed scores, all the effects reported above are maintained. The apparent difference in  $d'$  scores for the clapping action between the Experiments 1 and 2 is due to the exclusion of these two participants in Experiment 2.

<sup>3</sup> In order to test whether the better-than-chance performance observed with 2-marker displays was partially attributable to priming by earlier presented displays of the given action with more markers, we extracted data for all 2-marker trials that occurred before any trials with a greater number of markers, and then computed recognition accuracy for each participant based only on these data (note that this could not be done in the case of one participant, for whom all 2-marker displays occurred after the other conditions). This analysis revealed that participants' self-other judgments were on average accurate for 73.3% of 2-marker displays that were encountered before seeing fuller displays of the clapping action. This accuracy rate was significantly better than chance (50%) ( $t(8) = 6.16$ ,  $p < .01$ ).

Under such conditions, self-recognition may be driven more by visual than by auditory cues. The high informativeness of visual cues is evidenced by the fact that self-recognition was possible even when simple actions were depicted as point-light displays and even in highly degraded displays (i.e., displays with high spatial occlusion). Furthermore, it should be noted that the results obtained in our experiments concern actions that were not directed towards objects or goals. Finally, participants in our study performed the discrimination task with minimal training, without any feedback, and without having any previous experience with point-light displays.

The actions used in our experiments are gross motor movements (dancing, walking and clapping), which may afford easier agent recognition in comparison to previous experiments that focused on fine motor movement recognition (Daprati et al., 1997, 2007; Sirigu et al., 1999; Van den Bos & Jeannerod, 2002). Furthermore, whereas dancing may enable larger individual differences, in both of our experiments, agents were recognized even when the 'self' and the 'other' were performing movements in a very similar way (clapping). Even the highly ambiguous displays (two point-lights clapping) may contain a sufficient amount of the information needed to recognize an agent. This information is provided by the agents' movements in space and time and not by differing limb lengths and body proportions, as these anatomical features remained constant while self-other discrimination varied across different actions (Experiment 1) and different versions of the same action (Experiment 2). The fact that agents were distinguishable between different actions (Experiment 1) and for different versions of the same action (Experiment 2) may also imply that self-related kinematic properties possess highly specified gestalts. Perceptual information can be extracted by particular configural properties specified in action patterns (see also Pinto & Shiffrar, 1999). The occluded elements can perhaps be 'filled in' in the mind of the observer, and, therefore, make the agent identification possible. This suggestion is consistent with previous studies (Loula et al., 2005), supporting the role of action simulation for self-recognition. Recognition may be based on visual perception of idiosyncratic information, patterned in specified action gestalts.

Minimal visual information has been shown to be sufficient for action perception in a wide variety of music-related and other contexts. In the music domain, it has been shown that inexpressive body movements (Davidson, 1993) and reductions in visual information (Davidson, 1994) may provide sufficient cues for detecting a performer's intentions when presented as point-light displays. Furthermore, performer's intentions can be detected from blurred video images of specific body effectors (Dahl & Friberg, 2007). In other domains, Kozłowski and Cutting (1977) have reported that minimal visual information is sufficient for identifying the gender of a point-light actor and Pollick, Paterson, Bruderlin, and Sanford (2001) reported that arm movements are informative about the agent's underlying affective state.

Movement kinematics may be considered in Gibson's terms (Gibson, 1979) as invariant structures that can specify agent identity. Even under audiovisual transformations of information, the core elements of the sense of being an action's agent remain relatively intact and stable over time. The kinematic specification of dynamics (KSD) principle (Runeson & Frykholm, 1983) may account for the fact that the movement kinematics in point-light displays, apart from specifying action properties, expectancies and intentions, may also specify person-related individual characteristics. Therefore, if person-related dynamics may be identified from movement kinematics alone, then the availability of further dynamic information that accompanies the action (synchronization with music) or specifies the action (clap sounds) does not have an additive effect on identification. In other words, extra dynamic cues added to person-related dynamic properties may be redundant and thus of no use for the perceptual judgments. In Brunswik's terms, the information that specifies the actions of the self (i.e., the cues available) is a source of noise or uncertainty. The task of the organism is to base its perceptual judgments on cue-based probabilistic inferences of the information available (Brunswik, 1956; Hammond & Stewart, 2001). In our experiments, during the inferential process of the 'self' identifying 'self' or 'other' based on visual and auditory signals, some cues are given higher perceptual weights and some may be ignored. When multiple cues were available for recognizing self vs. other, participants relied on the kinematic information. In other words, kinematic information was sufficient for self-recognition. This implies that judgments of self-recognition depend on the availability and reliability of the cues provided. Different combinations of cues may lead to different strategies that determine the weighting of sensory and motor signals. The current results suggest that, when combined with visual cues, contextual auditory cues were superseded. This finding may imply a hierarchy of cues for self-recognition. In our experiments, visually communicated kinematic information seems to be at the top of this hierarchy.

Why do participants ignore auditory information? Firstly, according to the modality appropriateness hypothesis (Welch & Warren, 1980), discrepancies between modalities are resolved in favor of the most precise or the most appropriate modality for the given task or situation. In our experimental task, visual information may have dominated over information from other modalities. Moreover, the current results seem to challenge the inverse effectiveness principle (Alais & Burr, 2004; Meredith & Stein, 1983). In our experiments, the reduction of spatial information available in the visual modality did not increase reliance upon additional auditory information (whether self- or externally-generated). The ability to detect self vs. other information from unimodal (soundless) point-light displays may suggest that the comparisons for agent identity across modality conditions (visual vs. audiovisual) are made using a modality-neutral metric. Judgements of agent identity may be reliable even if they are executed based on unimodal information. Furthermore, temporal information may be encoded and embodied motorically when biological motion is perceived. Auditory cues may be ignored when visually observing the self in action potentially because action cues may modulate auditory and visual information perception. Another explanation may be the robustness of the action signals per se. Kinematic patterns may provide salient information for efficient self-other discrimination, thus, rendering all other information redundant. The current findings suggest that self-recognition may be an inferential process determined by an optimal combination of the perceptual weights allocated to the cues available.

Despite the absence of effects of auditory information in our study, the role of audiovisual cues in the perception of agent identity in musical contexts deserves further attention in future research. For instance, it may be fruitful to assess how self-recognition is affected by systematic manipulations of the degree of temporal (in)congruence between movements and auditory information in terms of phase (lead/lag) and periodicity (tempo) relations. Such exploration of the potential cues that mediate perception and action may provide further insights about how humans understand the actions of the self and others. Communicated information is multimodally encoded and embedded in interactive environments. By investigating interaction settings from multiple perspectives, one may gain an understanding of the constituent mechanisms involved in interpersonal relations and identify how these mechanisms enable shared intentions, joint actions and cooperation.

## Acknowledgments

The research was supported by the Max Planck Society. The authors would like to thank Bruno Repp for comments on an earlier version of this paper, Jan Bergmann for programming and technical assistance, Kerstin Träger for technical support, and Regine Steinke and Juliane Zeiss for their help with data collection. The results of Experiment 1 are reported in brief form in the proceedings of *The Neurosciences and Music – III* (Sevdalis & Keller, 2009).

## References

- Alais, D., & Burr, D. (2004). The ventriloquism effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262.
- Bangert, M., & Altenmüller, E. O. (2003). Mapping perception to action in piano practice: A longitudinal DC-EEG study. *BMC Neuroscience*, *4*, 26.
- Beardsworth, T., & Buckner, T. (1981). The ability to recognize oneself from a video recording of one's movements without seeing one's body. *Bulletin of the Psychonomic Society*, *18*, 19–22.
- Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*, 47–73.
- Blakemore, S., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*, 561–567.
- Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*, 756.
- Brunswick, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.
- Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, *15*, 1243–1249.
- Cross, E. S., Hamilton, A. F. D. C., & Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *Neuroimage*, *31*, 1257–1267.
- Cross, E. S., Kraemer, D. J. M., Hamilton, A. F. D. C., Kelley, W. M., & Grafton, S. T. (2009). Sensitivity of the action observation network to physical and observational learning. *Cerebral Cortex*, *19*, 315–326.
- Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk – Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, *9*, 353–356.
- Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, *24*, 433–454.
- Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., et al (1997). Looking for the agent: An investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, *65*, 71–86.
- Daprati, E., Wriessnegger, S., & Lacquaniti, F. (2007). Kinematic cues and recognition of self-generated actions. *Experimental Brain Research*, *177*, 31–44.
- Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, *21*, 103–113.
- Davidson, J. W. (1994). What type of information is conveyed in the body movements of solo musician performers? *Journal of Human Movement Studies*, *6*, 279–301.
- Decety, J., & Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences*, *3*, 172–178.
- Engbert, K., Wohlschläger, A., & Haggard, P. (2008). Who is causing what? The sense of agency is relational and efferent-triggered. *Cognition*, *107*, 693–704.
- Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor Facilitation during action observation – A magnetic stimulation study. *Journal of Neurophysiology*, *73*, 2608–2611.
- Flach, R., Knoblich, G., & Prinz, W. (2003). Off-line authorship effects in action perception. *Brain and Cognition*, *53*, 503–513.
- Flach, R., Knoblich, G., & Prinz, W. (2004). Recognizing one's own clapping: The role of temporal cues. *Psychological Research*, *69*, 147–156.
- Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*, 14–21.
- Gallagher, S. (2005). *How the body shapes the mind*. Oxford: Oxford University Press.
- Gallese, V. (2007). Before and below 'theory of mind': Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*, 659–669.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.
- Grèzes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: An fMRI study. *Neuroimage*, *18*, 928–937.
- Hammond, K. R., & Stewart, T. R. (Eds.). (2001). *The essential Brunswik: Beginnings, explications, applications*. New York: Oxford University Press.
- Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., & Ceballos-Baumann, A. O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *Journal of Cognitive Neuroscience*, *17*, 282–293.
- Hauelsen, J., & Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, *13*, 786–792.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*, 849–878.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*, 529–535.
- Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *Neuroimage*, *14*, S103–S109.
- Jeannerod, M. (2003). The mechanism of self-recognition in humans. *Behavioural Brain Research*, *142*, 1–15.
- Jeannerod, M. (2006). *Motor cognition: What actions tell the self*. New York: Oxford University Press.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*, 201–211.
- Jokisch, D., Daum, I., & Troje, N. F. (2006). Self recognition versus recognition of others by biological motion: Viewpoint-dependent effects. *Perception*, *35*, 911–920.
- Keller, P. E. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective on the study of interaction* (pp. 205–221). Amsterdam: IOS.
- Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition*, *16*, 102–111.
- Knoblich, G. (2008). Bodily and motor contributions to action perception. In R. L. Klatzky, B. MacWhinney, & M. Behrmann (Eds.), *Embodiment, ego-space and action* (pp. 45–78). New York: Psychology Press.

- Knoblich, G., & Prinz, W. (2001). Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 456–465.
- Knoblich, G., & Repp, B. (2009). Inferring agency from sound. *Cognition*, 111, 248–262.
- Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, 21, 575–580.
- Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired action. *Journal of Neuroscience*, 27, 308–314.
- Legrand, D., & Ruby, P. (2009). What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychological Review*, 116, 252–282.
- Leman, M. (2007). *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press.
- Lenggenhager, B., Smith, S. T., & Blanke, O. (2006). Functional and neural mechanisms of embodiment: Importance of the vestibular system and the temporal parietal junction. *Reviews in the Neurosciences*, 17, 643–657.
- Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: Manipulating bodily self-consciousness. *Science*, 317, 1096–1099.
- Lindenberger, U., Li, S.-C., Gruber, W., & Müller, V. (2009). Brains swinging in concert: Cortical phase synchronization while playing guitar. *BMC Neuroscience*, 10, 22.
- Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 210–220.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, 221, 389–391.
- Mutschler, I., Schulze-Bonhage, A., Glauche, V., Demandt, E., Speck, O., & Ball, T. (2007). A rapid sound–action association effect in human insular cortex. *PLoS ONE*, 2, e259.
- Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception*, 26, 489–504.
- Pinto, J., & Shiffrar, M. (1999). Subconfigurations of the human form in the perception of biological motion displays. *Acta Psychologica*, 102, 293–318.
- Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82, B51–B61.
- Prasad, S., & Shiffrar, M. (2009). Viewpoint and the recognition of people from their movements. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 39–49.
- Repp, B. H., & Knoblich, G. (2004). Perceiving action identity – How pianists recognize their own performances. *Psychological Science*, 15, 604–609.
- Repp, B. H., & Knoblich, G. (2007). Toward a psychophysics of agency: Detecting gain and loss of control over auditory action effects. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 469–482.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192.
- Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, 112, 585–615.
- Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: Corepresenting stimulus–response mappings. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 1234–1246.
- Sevdalis, V., & Keller, P. E. (2009). Self-recognition in the perception of actions performed in synchrony with music. *Annals of the New York Academy of Sciences*, 1169, 499–502.
- Sirigu, A., Daprati, E., Pradat-Diehl, P., Franck, N., & Jeannerod, M. (1999). Perception of self-generated movement following left parietal lesion. *Brain*, 122, 1867–1874.
- Tsakiris, M. (2008). The self–other distinction: Insights from self-recognition experiments. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective to the study of interactions* (pp. 149–163). Amsterdam: IOP Press.
- Tsakiris, M., & Haggard, P. (2005). Experimenting with the acting self. *Cognitive Neuropsychology*, 22, 387–407.
- Tsakiris, M., Haggard, P., Franck, N., Mainy, N., & Sirigu, A. (2005). A specific role for efferent information in self-recognition. *Cognition*, 96, 215–231.
- Van den Bos, E., & Jeannerod, M. (2002). Sense of body and sense of action both contribute to self-recognition. *Cognition*, 85, 177–187.
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88, 638–667.
- Wilson, M. (2002). Six views on embodied cognition. *Psychonomic Bulletin & Review*, 9, 625–636.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory–motor interaction in music perception and production. *Nature Reviews Neuroscience*, 8, 547–558.