Acta Psychologica 152 (2014) 67–74

Contents lists available at ScienceDirect

Acta Psychologica

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

Know thy sound: Perceiving self and others in musical contexts

Vassilis Sevdalis [a,*], Peter E. Keller [b]

[a] Institute of Psychology, German Sport University Cologne, Germany
[b] The MARCS Institute, University of Western Sydney, Sydney, NSW, Australia

ARTICLE INFO

Article history:
Received 16 December 2013
Received in revised form 7 April 2014
Accepted 7 July 2014
Available online xxxx

PsycINFO classification:
2300 (Human Experimental Psychology)
2500 (Physiological Psychology & Neuroscience)
2200 (Psychometrics & Statistics & Methodology)

Keywords:
Perception
Action
Self-recognition
Music
Cue integration
Embodied cognition

ABSTRACT

This review article provides a summary of the findings from empirical studies that investigated recognition of an action's agent by using music and/or other auditory information. Embodied cognition accounts ground higher cognitive functions in lower level sensorimotor functioning. Action simulation, the recruitment of an observer's motor system and its neural substrates when observing actions, has been proposed to be particularly potent for actions that are self-produced. This review examines evidence for such claims from the music domain. It covers studies in which trained or untrained individuals generated and/or perceived (musical) sounds, and were subsequently asked to identify who was the author of the sounds (e.g., the self or another individual) in immediate (online) or delayed (offline) research designs. The review is structured according to the complexity of auditory–motor information available and includes sections on: 1) simple auditory information (e.g., clapping, piano, drum sounds), 2) complex instrumental sound sequences (e.g., piano/organ performances), and 3) musical information embedded within audiovisual performance contexts, when action sequences are both viewed as movements and/or listened to in synchrony with sounds (e.g., conductors' gestures, dance). This work has proven to be informative in unraveling the links between perceptual–motor processes, supporting embodied accounts of human cognition that address action observation. The reported findings are examined in relation to cues that contribute to agency judgments, and their implications for research concerning action understanding and applied musical practice.

Contents

1. Introduction

Human beings experience a vast amount of auditory information in their everyday environments, such as the growl of thunder or the horns of cars, the voices of colleagues or the snoring of neighbors. Fortunately, certain sounds take the form of music. Music has the potential to be pleasurable in an aesthetic sense, but it can also be used as a means to investigate human capabilities related to it: these may include the production of sounds and their perception — the latter related to processes of identification and recognition of physical sound properties. Sometimes, the generator and receiver of the sound are one and the same person. This review article examines relationships between sounds generated by individuals' actions and the recognition of these sounds and the individuals who produce them via listening.

Certain performing arts, such as music and dance, include a prominent auditory component in the form of sounds generated by, or produced as an accompaniment to, the performers' actions. In these cases, being able

* Corresponding author at: German Sport University Cologne, Institute of Psychology, Am Sportpark Müngersdorf 6, 50933 Cologne, Germany. Tel.: +49 221 4982 5721; fax: +49 221 4982 8320.
E-mail address: v.sevdalis@dshs-koeln.de (V. Sevdalis).

http://dx.doi.org/10.1016/j.actpsy.2014.07.002
0001-6918/

to articulate a distinctive performance style is considered an asset and a quest pursued through long-term practice. Indeed, the actions of skilled and novice musicians and dancers have distinctive individual characteristics, that stem from anatomical constraints and different learning histories (e.g., Repp, 1992, 1995; Sevdalis & Keller, 2012). In the music domain, differences in individual variation of performance execution can occur both within an individual and between individuals. Discrepancies and commonalities in performance displays are nevertheless perceptually detectable, based on cues such as tempo, expressive timing, and dynamics (e.g., loudness). In perceptual experiments, for example, this has been demonstrated in jazz musicians accurately detecting whether the same piano melodies are improvised or imitated (Engel & Keller, 2011), in subjective judgments of similarity between different performances of the same piece (Timmers, 2005), and in aesthetic judgments of averaged individual music performances (Repp, 1997). Thus, accuracy in perceptual tasks that require identifying (di)similarities between actions is attainable when the sole information source is the sound alone.

From a theoretical point of view, the above examples showcase the capacity of body movements and their effects to shape cognitive operations, a fundamental premise of embodied cognition approaches (e.g., Grafton, 2009; Wilson, 2002). Such approaches converge on the assumption that high-level cognitive functions are grounded in low-level sensorimotor functions. According to this rationale, the importance of actions becomes particularly apparent if one considers their potential to be the means of enacting upon the environment (Herwig, Beisert, & Prinz, 2013). Essentially, performing and perceiving actions constitute an individual's means to interact with the environment and with other individuals. The functions of actions, thus, go beyond their motor components, and extend to cognitive and affective ones. Indeed, at both neurophysiological and behavioral levels, evidence is mounting that the coupling between action perception and action execution is boosted with increases in the degree to which an individual has physical experience in performing an action (e.g., Schubert & Semin, 2009; Sevdalis & Keller, 2011b).

Although research in action understanding has traditionally focused on the visual modality (but see Shams & Kim, 2010, for a review on the modulation of vision by auditory information), recent work has highlighted the importance of the auditory modality (for a review, see Aglioti & Pazzaglia, 2010). The significance of audition becomes particularly evident if one considers the diversity of activities that include primarily or solely the auditory channel, such as speaking, singing, and instrumental music performance. Evidence suggests that action-related sounds activate premotor areas in the human brain (Gazzola, Aziz-Zadeh, & Keysers, 2006), and that training on a musical instrument induces differences in somatosensory, auditory, and motor cortical brain functions and structures (e.g., Bangert et al., 2006; D'Ausilio, Altenmuller, Olivetti Belardinelli, & Lotze, 2006; Lahav, Saltzman, & Schlaug, 2007; Münte, Altenmüller, & Jäncke, 2002). Similar findings regarding the effects of action-related sounds were obtained when sounds of actions were presented to congenitally blind individuals (Ricciardi et al., 2009). Taken together, these results suggest that auditory–motor mappings are established in the brain and support the mutual influences between auditory and motor processes.

One specific class of actions that are particularly well suited for investigating these auditory–motor brain mappings and their behavioral counterparts are those produced by oneself. The perception of self-produced actions and their sounds benefits from the fact that, in this case, the observer's action system has specialized proprioceptive knowledge that is based on direct motor experience (Wilson & Knoblich, 2005). Agency, or being the agent of an action, refers to the feeling of being in control of one's actions and their effects (Pacherie, 2012; Repp & Knoblich, 2007). In the case of perceiving one's own sounds, the auditory–motor mappings have the potential to share a common code, that is, to become matched on the level of common auditory–motor representations (cf., Herwig et al., 2013). This matching of sensory and motor features of actions allows an individual to use his or her motor system to simulate an observed action, which can then be used to determine authorship (e.g., self or others) based on sensorimotor discrepancies and similarities between the simulation and the action (cf. Hommel, Müsseler, Aschersleben, & Prinz, 2001; Prinz, 1990). Such self-recognition capacities appear early in human life, between the first and second years (for reviews, see Butterworth, 1992; Rochat, 1998), are shared with other species (for a review, see Byrne & Bates, 2010), and are considered to be a constituent part of social cognition (Decety & Sommerville, 2003). However, the primary modality employed in self-recognition experiments remains the visual one (for a review, see Suddendorf & Butler, 2013). In spite of cross-modal and unimodal designs having been employed to assess self-recognition based on auditory information in an increasing number of studies (e.g., seeing and hearing one's name, Platek, Thomson, & Gallup, 2004; listening to auditory signals generated by one's footsteps, Menzer et al., 2010), a comprehensive account that deals with music-related actions is still wanting.

What makes sounds – and especially music – significant? Sounds have the potential to cover a broad coverage of environmental events and can intrinsically occur in synchrony with actions. Although vision may often dominate the human sensorimotor landscape (Colavita, 1974; Posner, Nissen, & Klein, 1976), visual information is less important in activities when auditory information is the primary means of expression, such as music-related ones (e.g., instrument learning and performance). Music is an ancient and culturally widespread activity, naturally present in most people's lives, and practiced by individuals with varying levels of expertise. These characteristics render musical sounds ecologically valid stimulus materials that can be readily used in experimental contexts. In music performance, for instance, the production of complex sequences is typically accompanied by receiving instant auditory feedback for the actions one performs: this auditory information can be experimentally manipulated to test how it affects performance execution (e.g., Pfordresher, Keller, Koch, Palmer, & Yildirim, 2011). Audition has very accurate temporal resolution: for example, the threshold for auditory temporal order judgments is around 20 ms (Hirsch, 1959; Hirsh & Watson, 1996) and the threshold for the detection of auditory onset asynchronies can be as low as 2 ms (Zera & Green, 1993). Audition is often considered a more 'accurate' sense than vision in certain situations such as temporal processing or synchronization (Arrighi, Alais, & Burr, 2006; Repp & Penel, 2002, 2004), and is more developed than vision before and at birth (Robinson & Sloutsky, 2004). Thus, sounds possess unique qualities, ranging from low-level physical properties up to high-level social information.

However, musical sounds have an inherent ambiguity attached to them (McGuiness & Overy, 2011). An interesting characteristic of musical listening is that it can create an auditory landscape whose properties are fluid in nature (i.e., they change each moment as the music unfolds) and are to some degree unpredictable (i.e., due to this continuous temporal evolution). The auditory system has lower spatial resolution than the visual system: when visually observing objects or events, both spatial and temporal dimensions can be employed for perceptual decision-making. The embodied nature of music perception and production (in terms of auditory–motor overlap) can pose challenges: essentially, identifying the properties of auditory recordings and understanding the agent's communicative intentions entails simulating the properties of actions that generated them (Keller, 2008). As a consequence, it can be more difficult to attribute one's own agency to an auditory signal, in comparison to a visual one (Sevdalis & Keller, 2010).

Musical activities, such as coordinating with a co-performer and predicting upcoming events, require monitoring one's own and others' actions (Keller, 2014; Keller, Novembre, & Hove, in press), and rely on knowing one's own and others' actions and their effects (agency). Presumably, if there is an efficient self–other distinction of sounds, then self- and other-awareness increases, and, thus, self and other sounds can merge into a coherent Gestalt during performance. Considering the universality of musical behavior, multiple factors influence music execution and perception, ranging from physical sound properties to cognitive expectations, associated emotions, or skill level. Music is an ideal candidate for studying agency because it is rich in multidimensional cues with salient auditory and visual components. Essentially, alongside vision, audition is a primary sense that serves as a medium to connect the motor system with the external world, by receiving input and generating output. Despite these qualities of music, the potential contribution of music-related sounds to agency attribution has not received the attention it deserves.

The aim of this review is to fill this void, by reporting the results of empirical studies that used music-related sounds to investigate agent recognition in explicit self–other judgments. This review does not cover studies discussed in other reviews, investigating aspects of self-recognition by auditory information stemming from the speech/voice (see Sidtis & Kreiman, 2012), singing (Prather & Mooney, 2004, for a comparative perspective), footsteps/gait (Kannape & Blanke, 2012; Larsson, 2014), or computer tones produced by button presses in laboratory tasks (Hughes, Desantis, & Waszak, 2013). The review includes sections on: 1) simple auditory information (e.g., clapping, piano, and drum sounds), 2) instrumental sound sequences (e.g., piano/organ performances) and 3) musical information embedded within audiovisual performance contexts in which action sequences are both viewed as movements and/or listened to as they occur in synchrony with sounds (e.g., conductors' gestures, dance). In the concluding remarks, we discuss the reported findings in relation to embodied accounts of human cognition that address action observation. Finally, we consider the reported findings in relation to the process of generating cues that contribute to agency judgments, their implications for research concerning action understanding, as well as applied musical practice.

2. Literature review

2.1. Simple auditory information

Clapping is a frequent music-related activity that is often expressed as simple rhythmic music accompaniment by non-expert individuals or as complex sound production, such as in flamenco music, or even as a performance reward in the form of applause. In a pioneering study, Repp (1987) asked individuals who were known to each other to clap at a self-selected rate. In a subsequent session, the same individuals were asked to listen to sequences of 10 successive claps, lasting about 10 s, and identify who has been clapping (self vs. others). Recognition was poor but above chance for others' clapping, whereas self-recognition (own clapping) was much higher: almost half of the individuals were able to identify their own clapping among 20 different clapping excerpts. These results were interpreted in terms of the potential of clapping to convey individual characteristics through sound: recognition performance was thought to be based on the acoustic properties/consequences of hand configuration during clapping (e.g., sound spectrum), and on individual differences in tempo and timing, since these were uncontrolled by the self-selected clapping rate.

In later research (Flach, Knoblich, & Prinz, 2004), the acoustic and temporal parameters of clapping sounds were explored through further manipulations. Participants without any particular music expertise, who did not know each other, were recruited and asked to produce clap sequences of different degrees of rhythmic complexity. In a subsequent session, at least a week later, the same individuals listened to recorded sounds of their claps and were asked to identify the agent (i.e., self vs. others): half of the participants received full information about the claps (i.e., in their original acoustic format, with information about frequency content, duration, and amplitude available); the other half received only temporal information, by listening to clap sounds that were replaced with tones (beeps), which had constant duration and amplitude, so that only the temporal structure of the claps was maintained. The results showed that recognition performance was independent of the rhythmic complexity of the sequences and did not differ between the two groups, with self-recognition accuracy being preserved even when the claps were replaced by uniform tones retaining only the original temporal pattern that was produced. In another experiment, the role of tempo was further explored by having participants listen to pure tone sequences that were presented either at the original self-tempo or at another (matched) participant's tempo. Tempo proved to be an important cue for self-recognition, since recognition performance was above chance only for the untransformed tempo sequences and at chance for the transformed ones. In a final experiment, participants drew circles on a graphic tablet in synchrony with tone sequences reproducing either their own or another participant's clapping. Self-recognition did not depend on concurrent synchronization accuracy with the auditory sequences. Taken together these results suggested that temporal cues were more important than non-temporal cues, such as the sound frequency and amplitude of clapping, for self-recognition.

The previous studies investigated self-recognition in delayed (offline) paradigms, with action generation of sounds and their perception separated in time. Another line of experiments has investigated self-recognition in more dynamic (online) designs. In one study (Repp & Knoblich, 2007), this was made possible by a finger-tapping task. Musically trained participants were asked to generate finger taps in synchrony with computer tones at different tempi. Each tap produced a tone of pitch height E7.[1] In one condition, the tone sequences either started as self-controlled (i.e., the participants were generating the tones) and, at transition points that were unpredictable for the participants, they changed to computer control (i.e., the computer was generating the tones) or vice versa. The task for the participants was to detect the transition from self to computer control or vice versa (i.e., whether their taps were triggering the sounds they heard or whether they were synchronizing with sounds controlled by the computer). In the second condition, the tone sequences (including the self/computer transitions) were just passively listened to as a playback of the sounds recorded in the active condition. The results showed that overall performance was above chance, and that perceptual accuracy in detecting transitions was higher in the actively generated tapping sequences than in sequences that were passively listened to. Correct responses were facilitated by the presence of produced timing variability between taps and tones (e.g., mean absolute asynchrony, mean absolute change in asynchrony), and perceived temporal irregularities (e.g., mean and absolute difference of the target tone interonset interval of the sequence relative to the one heard in a given trial). There was, however, an asymmetry in biases to perceive the tones as self- or computer-generated. Participants were biased towards mistaking the computer-generated tones as being produced by themselves in the active condition, while human-produced tones were more likely to be attributed to the computer in the passive condition.

In a follow-up study (Knoblich & Repp, 2009), the above results were confirmed and extended. This time, university students without specific musical expertise served as participants, and drum sounds were used as stimuli, in similar detection experiments of self- or computer-generated sounds. In a first experiment, students showed worse detection accuracy than the musicians of the Repp and Knoblich (2007) study. In a second experiment, a model sequence of isochronous sounds was presented and participants had to reproduce it by tapping. The taps either triggered sounds or were computer controlled. In a third experiment, tempo differences between the model sequence and the computer reproduction were introduced. The reproduction interonset interval was either the same as the model interonset interval or different by +/−5% and +/−10%. In both experiments, the results showed that perceptual sensitivity in discriminating between self or computer control was greater in the active in comparison to the passive condition, but also

[1] The note E of the 7th octave on the piano keyboard.

greater bias in the active than the passive one (cf. Repp & Knoblich, 2007). Again, variability of asynchronies between tapping movements and their auditory consequences, and variability of the time intervals between target and heard sounds, were used as cues for self–other distinction.

A variant of the above paradigm has also been used with clinical populations including schizophrenia patients and putative psychotic syndrome patients (Hauser et al., 2011). The task was to reproduce a computer-generated series of 4-tone drum sound sequences on a drum pad. While tapping, participants heard either their own self-produced tones or a computer controlled reproduction of drum tones. As in previous experiments, different tempi for the interonset intervals of sequence reproduction were introduced. Perceptual accuracy (in detecting whether the sounds were self- or computer-generated) was generally above chance, but was lower for patients in comparison to controls, with no difference evident between schizophrenic and putative syndrome individuals. Tempo changes were more effective for increasing perceptual accuracy for patients than for controls. Patients exhibited greater bias to attribute events to self, but no difference in bias between schizophrenic and putative individuals was observed. Patients produced larger asynchronies than controls (but only when the sequence reproduction tempo decelerated). Additionally, correlations of identification accuracy (and bias) with indices of psychopathology (e.g., assessments of proneness to schizophrenia or prodromal symptoms of schizophrenia) were found. The authors concluded that, in comparison to typical development individuals, schizophrenia and putative syndrome individuals were characterized by impairments in agency discrimination, exaggerated self-attribution bias, associations with psychopathology indices, and beneficial effects from additional sensorimotor cues to agency (e.g., deceleration of sequence tempo).

2.2. Instrumental sound sequences

The richness and complexity of instrumental sounds in music performances have been aptly employed as a tool for investigating self–other distinction. In a developmental study, Delogu and Olivetti Belardinelli (2003/2004) asked 6–12 year old children to listen to the song ‘twinkle twinkle little star’, then practice it on a computer keyboard by pressing any key on it (e.g., different rows of keys produced different sound intensities), and then recorded their performances. After 4, 6, or 8 days the children were presented with 3 different versions and asked to recognize which melody they had produced themselves. Recognition accuracy overall reached 69%, with a non-significant trend to improve across age. No differences were observed between males and females and no differences regarding the time intervals between the two sessions (i.e., the number of days elapsed). All children used mostly dynamics (i.e., loudness) as cues to recognition, but also errors in performance and tempo were employed by the older ones.

In a study with expert adult performers, Repp and Knoblich (2004) had advanced pianists record mostly unfamiliar pieces, each with a duration of 15–20 s, on a regular or a silent keyboard. Two months later, the pianists were asked to identify their performances (1 out of 12), by giving ratings on a 1–5 scale (5 = yes). The results indicated that they gave their own performance significantly higher ratings than any other pianist's performance. In two follow-up tests (2–3 months later), edited performances were presented, where differences in tempo (overall rate), overall dynamic (intensity) level, and dynamic nuances were removed. Pianists' ratings did not change significantly: self ratings were significantly higher than other ratings. This suggests that the remaining information of expressive timing and articulation—variables associated with idiosyncratic phrasing and musical/stylistic preferences—was sufficient (and the most salient) for self-recognition. Absence of sound during recording did not have a significant effect, and the same was true for familiarity with the pieces, suggesting that episodic memory for sounds of particular performances cannot explain the self-recognition effect.

In another experiment (Keller, Knoblich, & Repp, 2007), skilled piano performers were tested in a virtual duet paradigm. Each pianist was asked to record one part from piano duet pieces. The pieces were totally unfamiliar to them and lasted around 1–2 min each. Two to three months later the pianists returned and played the complementary part of the duet piece in synchrony with either their own or other participants' recordings. They were also asked to identify which recording was their own. The results showed that self-recognition was above chance and synchronization errors (i.e., absolute asynchrony between keystrokes in separate parts and the standard deviation of asynchronies) were smaller and less variable with self-produced than with other-produced performances. Furthermore, there was a positive correlation between asynchrony indices and recognition accuracy, indicating that the better the participants were in synchronizing, the better they were in recognizing agency (i.e., of having played the music themselves).

In addition to different performers playing different pieces, situations in which each performer plays different versions of the same piece have been employed. This was implemented in a study by Repp and Keller (2010), in which pianists played musical excerpts of about 20 s duration three times, and several months later were asked to identify whether pairs of recordings (containing two of their own performances or two performances by another arbitrarily matched pianist) were the same or different takes of the same excerpt. Pianists were above chance in recognizing their own performances, and equally good for self and other recordings in identifying whether the performance was identical or not. In a second experiment, the same pianists were asked to detect small local timing deviations (artificially introduced by the experimenters in the auditory stimuli): these were increments of interonset intervals, sounding as ‘hesitations’ in playing, and decrements of interonset intervals, sounding as ‘hastenings’ in playing. The results showed that the pianists were better in detecting these deviations in their own performances than in the performances of another pianist, but only when the deviations were placed at points of pre-existing difference in local timing (i.e., in places where the two performers' played with different expressive timing nuances). If the artificial deviations occurred in places where timing profiles were similar, and hence similar temporal expectations were entertained by self and others, then there was no self advantage. Hesitations were also easier (and marginally significantly faster) to be identified, than hastening, apparently because they sound more inappropriate (and are therefore easier to be perceived as perturbations), while hastening reflects a more valid expressive nuance.

A further step in investigating performer and listener variability was taken in a study by Gingras, Lagrandeur-Ponce, Giordano, and McAdams (2011). In this experiment professional organ players (prize and non-prize winners) produced two expressive and two inexpressive versions of the same Bach piece. Student listeners (one group with a little experience with music, and another group inexperienced with music) were then recruited and carried out a sorting task, of grouping together excerpts (of about 10–14 s duration) to performers. The results indicated that both musician and non-musician listeners performed above chance, with no differences between them in sorting accuracy. Both categories of listeners used expressive variations (tempo and articulation) for discriminating between performers. Still, prize winner performers were categorized more accurately than non-prize winners, and expressive performances were categorized more accurately than inexpressive ones, since more distinctiveness about an individual expressive pattern was available in these cases.

In a recent study (Van Vugt, Jabusch, & Altenmüller, 2013), such findings were extended, this time by using scales (i.e., stepwise ascending and descending sequences of pitches) rather than musical pieces. Listeners (pianists) were presented with pairs of ascending scales played on a piano. The task was to identify whether the scales were played by the same pianist or by different pianists. There were two versions of the scales, one 'veridical' and one 'magnified'. The latter version was edited so that it amplified all timing deviations, leading to more extreme keystroke timing variations. In the recognition test, only the magnified scale renditions were identified significantly above chance, indicating that timing irregularities are important for agent identification, even in simple musical materials (i.e., scales).

Finally, in a study that employed an online design, agency was investigated by providing auditory feedback to the performer while the musical piece was executed (Couchman, Beasley, & Pfordresher, 2012). Participants with little musical experience (i.e., piano training) performed novel melodies from memory on a keyboard. They heard either the normal feedback of their actions (i.e., unaltered pitches that were synchronous with keystrokes) or altered auditory feedback (i.e., with parametric or random variations of pitches and temporal lags). For instance, pitch lags were manipulated such that when a participant pressed a key, they heard the pitch corresponding to a note they had played earlier in the sequence; temporal delays were introduced by generating the tone after some time (e.g., 100, 200, 300, or 400 ms after the keypress). After each trial participants rated their experience of agency. The results of three experiments showed that pitch and synchrony alterations, both individually and in combination, have a significant effect on judgments of agency, with incremental variations leading to more ambiguous agency judgments than random ones (which were attributed to non-self actions). Although altered feedback also deteriorated the production of melodic sequences, it was found that agency judgments were independent of production errors, and also uninfluenced by the presence of a confederate who acted as an accompanist.

2.3. Musical information in audiovisual performance contexts

The final class of studies is one that investigated agency by using sound information embedded in realistic audiovisual performance contexts. In one example (Mitchell & MacDonald, 2014) musicians and music students first had to listen to an excerpt of around 10 s from the jazz standard 'Blue Bossa'. The music was played by one professional saxophone player, without accompaniment. After doing some filler tasks for 10 min, the participants listened to five professional saxophonists, and they were asked to identify the original saxophone player they had initially heard. The results showed that only 52% could do this task correctly, with identification rate being at chance level. In a following experiment, the audiovisual interactions in identification were explored. Professional musicians and music students were assigned in two groups: the first group watched a silent video clip of a sax player; then listened to 2 audio clips, one from the target sax player, and one from another. The second group did the process in reverse: they first listened to an audio clip of a sax player; then watched 2 silent video clips, one from the target sax player, and one from another. The task in both was to match the sound to the performer. The results revealed above chance identification for both groups, with the visual-first group being higher in accuracy than the auditory-first group, indicating that participants were more likely to recognize the performer after first having seen his performance than when having first listened to it. The experimental participants were also asked to provide verbal comments on how they made their decisions. Among the most frequently mentioned cues used for the identification were body movement/language, breathing, finger positioning, and embouchure (mouth configuration and movements).

Apart from music performance, other activities take place in synchrony with music, even by individuals with limited musical expertise. In a relevant study, Sevdalis and Keller (2009, 2010) had participants listen and then dance, walk, and clap in synchrony with dance, jazz, and folk music. In a subsequent session, the same individuals were invited back, and they were presented with audiovisual or just visual recordings of their performances. The recordings portrayed the dancing, walking, and the clapping as point-light displays either with or without the music the participants had moved with. The results showed that self-recognition depended on the complexity of the observed action, with higher recognition for the complex action (dancing vs. walking and clapping), but not on the presence of sound, as recognition was similar for visual and audiovisual displays. In a subsequent experiment, the same participants were invited back and they were shown only the clapping actions (varying in the number of point-lights), this time in a visual condition, an audiovisual condition with clapping sounds, and an audiovisual condition that included the clapping sounds along with the background music. Similar results were obtained, with higher recognition performance in movement displays containing larger number of point-lights, without any contribution from self-generated (clapping sounds) or externally-generated (music) information. In another follow-up study (Sevdalis & Keller, 2011a), the above results were confirmed by showing point-light displays of dance of parametrically decreasing duration (from 5 to 1 s): audiovisual presentation did not influence self-recognition (whereas display duration did). Empathy indices, as assessed by a self-report questionnaire, also correlated positively with self-recognition accuracy (see also Sevdalis & Raab, 2014).

Music and movement are intrinsically intertwined in another variant of the performing arts, namely, orchestral conducting. Orchestral conductors present a special case of agency, in which an individual is responsible for producing sounds, albeit indirectly, by guiding others' actions. A study that investigated the perception of agency in musical experts employed professional orchestral conductors (Wöllner, 2012). In a first session, recordings of conductors directing a string quintet were made (including body movements tracked with a motion capture system in addition to sounds). In a subsequent session, 3–4 months later, the same individuals were invited back and were presented with point-light displays. They watched visual, auditory, audiovisual, static, and walking stimuli of three individuals. For the auditory and audiovisual conditions, two musical excerpts were extracted from the recordings of the orchestra led by the conductor, of 7–11 s and 6–9 s. The task was to identify whether the self or someone else was displayed (or interpreting the music in the auditory condition). Ratings of the quality of the performance and the emotional content of the music were collected, and the Affective Communication Test (measuring individual differences in communicative expressiveness) was administered. The results showed that self-recognition was accurate in visual only and audiovisual conditions and at chance in the auditory only condition. For visual conditions, conductors rated their own performances to be of higher quality (i.e., judged their own conducting movements as having higher quality in comparison to others' movements); differences in quality ratings were not influenced by recognition accuracy. Nevertheless, conductors were better in identifying the emotional content (happiness) of the music in audiovisual presentations. Conductors with high scores in expressive communication were also better at recognition, whereas years of training and age were not related to recognition.

3. Concluding remarks

This article reviewed empirical studies that used music/auditory information as a basis for agent identification. Our overview of the literature in the field was organized into sections on 1) simple auditory information, 2) complex instrumental sound sequences, and 3) musical information embedded within audiovisual performance contexts in which action sequences are both viewed as movements and/or listened in synchrony with sounds. We conclude by briefly recapitulating the major findings, and then comment on the theoretical implications of these findings for future research and applied musical practice.

The reviewed studies revealed that agent recognition is attainable in a variety of auditory contexts, ranging from simple auditory information, to music performance, and orchestral sounds. Simple auditory information, such as clapping or sounds triggered by taps, has been shown to provide sufficient information, even if the information is impoverished with regard to its acoustic components: tempo and timing can be used as cues to identify an action's agent both in offline and online designs.

For the latter, transitions between self- and computer- controlled sequences is possible even for individuals without musical expertise or for patients with psychosis, and is improved when active (vs. passive) engagement of the individual is required. With regard to the studies that used music performances, it has been shown that agent identification emerge in the childhood years and continues to improve (although the evidence for these development trends is not robust), is possible by non-experts, but does benefit from music-related expertise (e.g., in synchronization abilities). Variations in timing play a more important role for agent identification than other cues such as global tempo, intensity, familiarity, or presence of another co-acting individual, especially when they occur in places where individual differences are more clearly articulated or sound more pronounced. Finally, studies that used concurrent audiovisual information for agent recognition converge on the conclusion that there is a general tendency for the visual modality to dominate the auditory modality, but also modulation of this tendency by social-emotional self-report indices, such as individual differences in expressive communication or empathy. Taken together, these results suggest that agent recognition is perceptually salient in most cases, but is susceptible to multiple modulating factors.

How can the process of agent identification be described in terms of its underlying mechanisms? The theoretical point made here is that the task of allocating agency to an action recruits a simulation process, where motor resonance and its neural substrates become especially active for actions that are embodied, that is, stemming from one's self (Jeannerod, 2006). The studies reviewed converge to support action simulation while perceiving auditory action effects: participants can use their own motor system to contrast both self-generated and other-generated action effects, and predict their auditory properties even when actions have not been self-generated (cf. Schubotz, 2007). The question is how simulation is induced, in other words, how to assess one's sensorimotor experience and the perception–action representations associated with it.

The current proposal is that agency judgments are largely grounded on the type of manipulation and the nature of the task. The fact that many manipulations have been effective in altering (i.e., increasing or decreasing) the accuracy of agency judgments concurs with the conclusion that separate cues to agency have to be distinguished and their contribution(s) assessed. A large amount of the studies revealed that perceptual identification alters when experimental manipulations are implemented at sensorimotor level (e.g., tempo, sound intensity, amount of information) and/or at individual differences level (e.g., trait comparisons). Both physical stimulus properties and contextual parameters are effective in influencing the agency judgments. Thus, the cues to agent identification may be distinguished in two categories: those related to sensorimotor/bottom-up cues and those related to social-cognitive/top-down cues. The ways in which cues to agency are perceived rely largely on these two sources of information.

It is suggested here that perceiving an action's agent is the outcome of the observer assessing (implicitly or explicitly) the validity of a hierarchy of cues specifying this action's agent. The cues range from low-level physical stimulus properties (e.g., intensity, tempo) to cognitive expectations (e.g., biases) and individual differences (e.g., empathy, psychopathology level, skill level). These levels (or cues), depending on their presence, absence, or co-existence (i.e., whether manipulated or not) may function in various ways: a) independently (e.g., by a cue such as timing being itself sufficient to reveal the action's agent and its manipulation influencing recognition (Repp & Knoblich, 2004)); b) in synergy or complementing each other (e.g., increased synchronization accuracy associated with increased recognition accuracy Keller et al., 2007); empathy complementing recognition accuracy (Sevdalis & Keller, 2011a, 2011b); and c) in competition or conflicting/substituting each other (e.g., visual over auditory modality dominance Wöllner, 2012); cognitive bias influencing bottom-up perceptual judgments Repp & Knoblich, 2007). In other words, the cues to agency are dynamic in nature, in the sense that they carry asymmetrical influence and thus possess different relative weights for perceptual judgments. Accuracy in agent identification may therefore not be accounted by one source of information alone, and seems to vary as a function of any (multiple) factors available.

Nevertheless, despite this multifariousness in cue utilization, the cues employed by individuals for agency judgments seem to work rapidly, vicariously, and efficiently. Although null effects have been sometimes reported in the reviewed studies (presumably due to a weakness of the cue used for agent identification in a given experimental condition), self–other distinction occurs with impoverished temporal information and is, thereupon, sensitive to various potentially exchangeable sources of information. Information from sensorimotor or higher cognitive pathways may be flexibly combined towards an accurate perceptual judgment. The combination of cues may be conceived of as operating across two layers: as a structural/hierarchical layer, by identifying a valid cue amongst informational sources (e.g., a more salient cue — such as visual over auditory information) and as a process/time-course layer, by identifying the timing/duration of the cue validity (e.g., parametric manipulation effects of introducing asynchronies in auditory feedback). It is useful to conceive cue utilization across structure and time because the contribution of a cue for agency (sufficiency) is not identical to dependence of agency on this cue (necessity). Flexibility implies different roles (hierarchies, probabilities) of cues in different experimental situations. Treating the topic of agent identification at multiple levels would thereby require further elucidation of components to agency (e.g., which cues play a role, when they do so, or by whom they are used).

The main implication of such a theoretical approach for research is to conceive the capacity to understand actions of the self and others as guided by a range of intertwined factors. In the music domain, this translates into assessing actions of the self or others in situations that require framing expectations about upcoming actions, anticipating outcomes of intentional actions, and responding to them in real time (i.e., using sensorimotor experiences and their auditory consequences to model or infer those of others). In a broader context, beyond music perception and performance, agency attribution may be considered a fundamental social cognition skill, related to awareness and understanding of the self and others. Mastering the ability to recognize auditory (as well as visual) action consequences may have obvious implications for mental state attribution, mind-reading, empathy, and interpersonal communication. The empirical challenge that stems from these premises is to discover the triggering factors for simulation to occur. Although some cues to agency and their contributions have been identified in the reviewed literature, the precise nature of their interaction remains an open topic for future investigations.

The multi-factorial nature of the agent identification capacity possesses additional interesting characteristics that relate to sensorimotor and cognitive functions. First, it appears to be supramodal, which means that it can occur irrespectively of which modality is stimulated. Although agent identification tasks have been commonly investigated in the visual modality (e.g., Knoblich & Prinz, 2001; Loula, Prasad, Harber, & Shiffrar, 2005), the current review indicates that self–other distinction abilities are adequate in the auditory modality alone. Indeed, within the field of action understanding, it has been shown that modality-neutral representations of actions are present and functional in many types of situations, such as in musicians or blind individuals (e.g., Münte et al., 2002). This flexible facility renders the understanding of actions possible even in terms of only the sounds they produce, such as in music-related information. Another characteristic of agent identification is that it does not presuppose a deliberate recollection of the episode of action generation: in other words, episodic memory of the action execution is not a prerequisite in order to identify an action's agent. Several of the reviewed studies used novel stimulus material, large intermissions between the action execution and the action perception session, and controlled for online feedback while the action was executed (e.g., the silent keyboard in Repp & Knoblich, 2004). The fact that agent identification was still attainable after such manipulations suggests that what drives the agency judgment process is the identification of individual differences between actions (i.e., *how* an agent would do the action) and not the identification of a remembered action (i.e., *what* an agent did in the execution session). Finally, it is worth considering agent identification as a general process of perceptual similarity estimation. This similarity can occur between different individuals (i.e., the same action performed by different agents) and within the same individual (i.e., different actions performed by the same agent). Identifying one's self does not seem to be particularly different from identifying other types of commonalities between stimuli such as music performances by comparing their physical properties (e.g., Engel & Keller, 2011; Timmers, 2005).

How can these characteristics of the process of agent identification be used in future applied practice contexts of learning to perform music? A number of suggestions can be implemented in order to corroborate and extend the current findings. First, one should consider designs that employ different instrumental sounds, given that the dominant instrument so far has been the piano. More precisely, certain instruments (e.g., string and wind instruments cf. Mitchell & MacDonald, 2014) may be more sensitive to nuances in the performer's actions: this would enable a better understanding of how acoustic sound properties affect agent identification. Alongside musical sounds, a more ecologically valid approach could implement designs where simple auditory information co-occurs with instrumental sounds: in many musical traditions worldwide, it is common for music performances to be accompanied by clapping sounds or simple (nonverbal) vocalizations. Further advances can be made by exploring the performances of jazz musicians, the 'par excellence' producers of distinctive music playing styles through the art of improvisation. It is common knowledge amongst music aficionados that famous performers can be identified within fragments of seconds of their playing, but there is little empirical research exploiting this potential, especially with regards to agent identification.

Another possibility is to implement psychophysical designs in which a musician's performance is edited in order to parametrically match/blend with the performance of another individual. Longitudinal designs can be also of assistance since they would help explore changes in one's production and perception abilities, and provide insights about brain plasticity and behavioral learning, especially with regards to an individual's awareness of his/her style and the process of deliberatively articulating it. Such designs could be beneficial for pedagogical practices, with the aim of developing musical expertise: presumably, learning to monitor and guide one's own actions and their effects could lead to improvements in music perception and production abilities in challenging contexts, such as ensemble performance. Finally, the investigation of agent recognition in music-related tasks may advance through the application of neuroimaging methods, with a view to unraveling potential neural substrates involved in self-recognition via the perception of sounds. For instance, Novembre, Ticini, Schütz-Bosbach, and Keller (2012) showed that corticospinal excitability was modulated (lowered) by the belief that a recorded musical piece was associated with the self (vs. the co-performer). Outside of the music domain, valuable insights have been gained with regards to the differing degrees of activation of the motor system when observing actions (Justen, Herbert, Werner, & Raab, 2014; Ticini, Schütz-Bosbach, Weiss, Casile, & Waszak, 2011): transferring such designs to the music domain could potentially advance knowledge of the use of cues to agency. Such advances could offer further insights on the interplay of musical sounds and the motor system for understanding the specific contribution of self-generated actions to compliment previous accounts on music-motor brain mappings (cf. Zatorre, Chen, & Penhune, 2007).

Finally, in line with some recent trends in the cognitive sciences and neurosciences that invite research that investigates cognition through action and interaction (Engel, Maye, Kurthen, & König, 2013; Schilbach et al., 2013), this review makes a similar case for music. We hold that music perception and production are participatory and interactive activities for individuals of all ages, naturally present in one's life, and do not require extensive expertise: as such, they provide ideal conditions to test some of the research and applied practice proposals made above. By exploring the natural instances where music is present (e.g., certain sports, or performing arts contexts, cf. Kennel, Hohmann, & Raab, 2014; Sevdalis & Raab, 2014) one can benefit by understanding both the functions of the music itself, and the behavior in the context which music occurs.

Acknowledgments

The research was supported by the German Sport University Cologne. The authors would like to thank Bruno H. Repp for earlier discussions on the topic of this paper.

References

Aglioti, S. M., & Pazzaglia, M. (2010). Representing actions through their sound. *Experimental Brain Research*, *206*, 141–151.

Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision*, *6*, 260–268.

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., et al. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*, *30*, 917–926.

Butterworth, G. (1992). Origins of self-perception in infancy. *Psychological Inquiry*, *3*, 103–111.

Byrne, R. W., & Bates, L. A. (2010). Primate social cognition: Uniquely primate, uniquely social, or just unique? *Neuron*, *65*, 815–830.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*, 409–412.

Couchman, J. J., Beasley, R., & Pfordresher, P. Q. (2012). The experience of agency in sequence production with altered auditory feedback. *Consciousness and Cognition*, *21*, 186–203.

D'Ausilio, A., Altenmuller, E., Olivetti Belardinelli, M., & Lotze, M. (2006). Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *European Journal of Neuroscience*, *24*, 955–958.

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: A social cognitive neuroscience view. *Trends in Cognitive Sciences*, *17*, 527–533.

Delogu, F., & Olivetti Belardinelli, M. (2003/2004). Children's recognition of their musical performance. *Musicae Scientiae*(Special Issue), 31–48.

Engel, A., & Keller, P. E. (2011). The perception of musical spontaneity in improvised and imitated jazz performances. *Frontiers in Psychology*, *2*, 83.

Engel, A. K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences*, *17*, 202–209.

Flach, R., Knoblich, G., & Prinz, W. (2004). Recognizing one's own clapping: The role of temporal cues. *Psychological Research*, *11*, 147–156.

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology*, *16*, 1824–1829.

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., & McAdams, S. (2011). Perceiving musical individuality: Performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception*, *40*, 1206–1220.

Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Annals of the New York Academy of Sciences*, *1156*, 97–117.

Hauser, M., Knoblich, G., Repp, B. H., Lautenschlager, M., Gallinat, J., Heinz, A., et al. (2011). Altered sense of agency in schizophrenia and putative psychotic syndrome. *Psychiatry Research*, *186*, 170–176.

Herwig, A., Beisert, M., & Prinz, W. (2013). Action science emerging: Introduction and leitmotifs. In W. Prinz, M. Beisert, & A. Herwig (Eds.), *Action science: Foundations of an emerging discipline* (pp. 1–33). Cambridge, MA: MIT Press.

Hirsch, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America*, *31*, 759–767.

Hirsh, I. J., & Watson, C. S. (1996). Auditory psychophysics and perception. *Annual Review of Psychology*, *47*, 461–484.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding: A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*, 869–878.

Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin*, *139*, 133–151.

Jeannerod, M. (2006). *Motor cognition: What actions tell the self.* New York: Oxford University Press.

Justen, C., Herbert, C., Werner, K., & Raab, M. (2014). Self vs. other: Neural correlates underlying agent identification based on unimodal auditory information as revealed by electrotomography (sLORETA). *Neuroscience*, *259*, 25–34.

Kannape, O. A., & Blanke, O. (2012). Agency, gait and self-consciousness. *International Journal of Psychophysiology*, *83*, 191–199.

Keller, P. E. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective to the study of interactions* (pp. 205–221). Amsterdam: IOS Press.

Keller, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures* (pp. 260–282). Oxford: Oxford University Press.

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.
Keller, P. E., Novembre, G., & Hove, M. J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B* (in press).
Kennel, C., Hohmann, T., & Raab, M. (2014). Action perception via auditory information: Agent identification and discrimination with complex movement sounds. *Journal of Cognitive Psychology*, *26*, 157–165.
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. H. (2009). Inferring agency from sound. *Cognition*, *111*, 248–262.
Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, *27*, 308–314.
Larsson, M. (2014). Self-generated sounds of locomotion and ventilation and the evolution of human rhythmic abilities. *Animal Cognition*, *17*, 1–14.
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.
McGuiness, A., & Overy, K. (2011). Music, consciousness and the brain: Music as shared experience of an embodied present. In E. F. Clarke, & D. Clarke (Eds.), *Music and consciousness: Philosophical, psychological and cultural perspectives* (pp. 245–262). Oxford: Oxford University Press.
Menzer, F., Brooks, A., Halje, P., Faller, C., Vetterli, M., & Blanke, O. (2010). Feeling in control of your footsteps: Conscious gait monitoring and the auditory consequences of footsteps. *Cognitive Neuroscience*, *1*, 184–192.
Mitchell, H. F., & MacDonald, R. A. R. (2014). Listeners as spectators? Audio-visual integration improves music performer identification. *Psychology of Music*, *42*, 112–127.
Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, *3*, 473–478.
Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2012). Distinguishing self and other in joint action. Evidence from a musical paradigm. *Cerebral Cortex*, *22*, 2894–2903.
Pacherie, E. (2012). The phenomenology of joint action: Self-agency vs. joint-agency. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 343–389). Cambridge, U.S.A.: MIT Press.
Pfordresher, P. Q., Keller, P. E., Koch, I., Palmer, C., & Yildirim, E. (2011). Activation of learned action sequences by auditory feedback. *Psychonomic Bulletin & Review*, *18*, 544–549.
Platek, S. M., Thomson, J. W., & Gallup, G. G. (2004). Cross-modal self-recognition: The role of visual, auditory, and olfactory primes. *Consciousness and cognition*, *13*, 197–210.
Posner, M. I., Nissen, M. J., & Klein, M. (1976). Visual dominance: An information processing account of its origins and significance. *Psychological Review*, *83*, 157–171.
Prather, J. F., & Mooney, R. (2004). Neural correlates of learned song in the avian forebrain: Simultaneous representation of self and others. *Current Opinion in Neurobiology*, *14*, 496–502.
Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann, & W. Prinz (Eds.), *Relationships between perception and action: Current approaches* (pp. 167–201). Berlin: Springer.
Repp, B. H. (1987). The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America*, *81*, 1100–1109.
Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's '*Träumerei*'. *Journal of the Acoustical Society of America*, *92*, 2546–2568.
Repp, B. H. (1995). Expressive timing in Schumann's '*Träumerei*': An analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America*, *98*, 2413–2427.
Repp, B. H. (1997). The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception*, *14*, 419–444.
Repp, B. H., & Keller, P. E. (2010). Self versus other in piano performance: Detectability of timing perturbations depends on personal playing style. *Experimental Brain Research*, *202*, 101–110.
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.
Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 1085–1099.
Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research*, *68*, 252–270.
Ricciardi, E., Bonino, D., Sani, L., Vecchi, T., Guazzelli, M., Haxby, J. V., et al. (2009). Do we really need vision? How blind people 'see' the actions of others. *The Journal of Neuroscience*, *29*, 9719–9724.
Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development*, *75*, 1387–1401.
Rochat, P. (1998). Self-perception and action in infancy. *Experimental Brain Research*, *123*, 102–109.
Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*, 393–414.
Schubert, T. W., & Semin, G. R. (2009). Embodiment as a unifying perspective for psychology. *European Journal of Social Psychology*, *39*, 1135–1141.
Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences*, *11*, 211–218.
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.
Sevdalis, V., & Keller, P. E. (2010). Cues for self-recognition in point-light displays of actions performed in synchrony with music. *Consciousness and Cognition*, *19*, 617–626.
Sevdalis, V., & Keller, P. E. (2011a). Perceiving performer identity and intended expression intensity in point-light displays of dance. *Psychological Research*, *75*, 423–434.
Sevdalis, V., & Keller, P. E. (2011b). Captured by motion: Dance, action understanding, and social cognition. *Brain and Cognition*, *77*, 231–236.
Sevdalis, V., & Keller, P. E. (2012). Perceiving bodies in motion: Expression intensity, empathy, and experience. *Experimental Brain Research*, *222*, 447–453.
Sevdalis, V., & Raab, M. (2014). Empathy in sports, exercise, and the performing arts. *Psychology of Sport and Exercise*, *15*, 173–179.
Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, *7*, 269–284.
Sidtis, D., & Kreiman, J. (2012). In the beginning was the familiar voice: Personally familiar voices in the evolutionary and contemporary biology of communication. *Integrative Psychological and Behavioral Science*, *46*, 146–159.
Suddendorf, T., & Butler, D. L. (2013). The nature of visual self-recognition. *Trends in Cognitive Sciences*, *17*, 121–127.
Ticini, L. F., Schütz-Bosbach, S., Weiss, C., Casile, A., & Waszak, F. (2011). When sounds become actions: Higher-order representation of newly learnt action sounds in the human motor system. *Journal of Cognitive Neuroscience*, *24*, 464–474.
Timmers, R. (2005). Predicting the similarity between expressive performance of music from measurements of tempo and dynamics. *Journal of the Acoustical Society of America*, *117*, 391–399.
Van Vugt, F. T., Jabusch, H. -C., & Altenmüller, E. (2013). Individuality that is unheard of: Systematic temporal deviations in scale playing leave an inaudible pianistic fingerprint. *Frontiers in Psychology*, *4*, 134.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625–636.
Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, *131*, 460–473.
Wöllner, C. (2012). Self-recognition of highly skilled actions: A study of orchestral conductors. *Consciousness and Cognition*, *21*, 1311–1321.
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.
Zera, J., & Green, D. M. (1993). Detecting temporal onset and offset asynchrony in multicomponent complexes. *Journal of the Acoustical Society of America*, *93*, 1038–1052.