



Nina Heins

**The relevance of incidental and intentional action sounds
for the evaluation and neural processing of actions**

- 2020 -

The relevance of incidental and intentional action sounds for the
evaluation and neural processing of actions

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

im Fach Psychologie

eingereicht im Fachbereich Psychologie und Sportwissenschaft

der Westfälischen Wilhelms-Universität Münster

von

Nina Heins

geboren am 22.06.1992 in Düsseldorf

Dekan:

Prof. Dr. Guido Hertel

Erste Gutachterin:

Prof. Dr. Ricarda I. Schubotz

Zweite Gutachterin:

Prof. Dr. Pienie Zwitterlood

Datum der mündlichen Prüfung:

Datum der Promotion:

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Ricarda Schubotz for giving me the opportunity to pursue my PhD, and supporting me whenever I needed it during this exciting period. I value our inspiring discussions, and she challenged me to always think critically and to thrive as a researcher. Thanks go to Prof. Markus Raab and Prof. Karen Zentgraf, who contributed valuable scientific support and were amazing co-authors.

I would like to thank all the people who supported my scientific “career” early on, even in my undergrad years, and later, during my masters. Without you, I would not have decided to pursue a PhD: Katrin Lange, Matthias Hoenen, Nora Schaal, and Erhan Genç.

An enormous thank you goes to all my wonderful colleagues (and friends) in our lab. Jennifer Pomp, for being an amazing office partner for the first two years of my PhD and bringing me back to reason more than once. Marlen Roehle, for her cheerful manner, endless funny chats at and outside of work, and the much appreciated proofreads. Lena Schliephake, for her analytical view, our critical discussions, shared appreciation for good crime series, and shared jogging routes. Ima Trempler, for providing work-related and personal guidance whenever needed, and being one of the most helpful people I know. Nadiya El-Sourani, for always being supportive and inspiring. Amelie Hübner, for therapeutic talks and her amazing parties. Sophie Siestrup, for being a fellow Ravenclaw “nerd” and shared mystery nights. Benjamin Jainta, for his optimistic attitude and for understanding all my “funny” references.

An equally large thanks goes to Monika Mertens and Jutta Linke, who helped me with carrying out my experiments and organizing my working environment. Thanks to all the student assistants who helped me during my PhD – whether it was help with the intricate filming sessions and data processing, help in conducting my experiments, or just emotional help provided as a chat in the lab’s kitchen: Theresa Eckes, Alina Eisele, Katharina Thiel, Stefan

Vinbrüx, Marie Kleinbielen, Annika Garlichs, Corinna Gietmann, Falko Mecklenbrauck, and Helena Sydlik.

In addition to all the work-related thanks, I want to thank my amazing friends and family for their never-ending support. A special thanks goes to my closest friends over the years: Ronja Rast, for so many shared memories and always reminding me of the person I want to be. Lars Rollik, for eight-hour discussions and so much coffee. Katrin Mittermeier, for sharing my sense of humor and indulging me in my love for mountains. Melinda Baranyai, for understanding all my emotional troubles and still being my friend. Thank you to my parents, for always supporting me in finding my own ways, and giving me the freedom to decide for myself. And the most thanks of all go to my wonderful siblings, Inga and Tim, who understand me like no other person in this world and whom I can never thank enough for their support.

Thanks to all the people who made this possible – there are so many of you!

Table of Contents

Summary	1
List of Original Publications	5
1 Theoretical and Empirical Background	6
1.1 What are action sounds/ auditory re-afferences?	6
1.2 Self-initiated action sounds and sensory attenuation	8
1.3 The predictive coding approach to action sounds	11
1.3.1 Predictive coding and perception	11
1.3.2 The Mismatch Negativity (MMN) as a prediction error indicator	12
1.3.3 Cortical hierarchy of action sound prediction	14
1.4 Action sounds, effects of their omission and interference, and their neural representation	16
1.4.1 Language	16
1.4.2 Music	19
1.4.3 Action sounds apart from language and music	21
2. Research Questions	23
3. Research Articles	27
3.1 Study I: Incidental or Intentional? Different Brain Responses to One's Own Action Sounds in Hurdling vs. Tap Dancing	27
3.2 Study II: Where is my sound? Omission of one's own intentional and incidental action sounds and its effect on BOLD fMRI	44
3.3 Study III: Too late! Influence of temporal delay on the neural processing of one's own incidental and intentional action sounds.	69
4. General Discussion and Future Directions	93
4.1 Summary of the presented studies	93
4.2 Limitations of our findings	95
4.3 Incidental or intentional – dichotomy or continuum?	98

4.4 Auditory, visual, multisensory – relevance of the different senses for action _____	100
4.5 The interplay of prediction and attention _____	103
4.6 Practical applications of our results _____	106
5. Conclusion _____	108
References _____	109
Abbreviations _____	122
List of Figures _____	123
Curriculum Vitae _____	124

Summary

Sounds are omnipresent in our everyday life. To navigate the rich auditory environment efficiently, our brain constantly generates predictions about upcoming sounds, and attenuates those sounds aligning with our predictions. Although this process works especially well for sounds created by our own actions, it is also applicable for action sounds generated by others. However, it has not been examined how our own intention to create a sound with an action influences these predictive processes. While sound production is our intentional goal when e.g. speaking or singing, we do not intend to generate e.g. the sounds of our footsteps on the ground when walking, which rather arise as an incidental by-product of the action. Still, the lack of those sounds would probably surprise us.

The present thesis examined the diverging relevance of incidental and intentional action sounds for the evaluation and neural processing of actions. To this end, three functional magnetic resonance imaging (fMRI) studies were conducted, investigating hurdling (incidental action sounds) and tap dancing (intentional action sounds). The first study assessed the underlying neural networks, including the primary auditory cortex, the posterior superior temporal gyrus (pSTG) and the supplementary motor area (SMA). Both attenuation in primary auditory cortex and predictive processing in pSTG and SMA were enhanced for intentional action sounds. Auditory “scrambling” additionally affected the behavioural rating scores for intentional action sounds more strongly. The second study investigated the influence of sound omission. The lack of action sounds decreased rating scores only for intentional action sounds, which could be restored to some degree by predictive input by SMA. The third study, examining effects of temporally delayed action sounds, strengthened the assumption that SMA provides additional predictive information to guarantee a smooth performance evaluation. This is especially the case when intentional action sounds are distorted.

The results of this thesis provide the foundation for a more profound understanding of the predictive mechanisms involved in action sound processing, depending on the intentionality of sound production.

Zusammenfassung

Geräusche sind in unserem Alltag allgegenwärtig. Um unsere komplexe auditorische Umgebung effizient navigieren zu können, generiert unser Gehirn fortlaufend Vorhersagen über bevorstehende Geräusche, und dämpft solche Geräusche, die mit unseren Vorhersagen übereinstimmen. Dieser Prozess verläuft besonders reibungslos für Geräusche, die wir mit unseren eigenen Bewegungen erzeugen, aber auch für von anderen generierte Bewegungsgeräusche. Es wurde allerdings bisher noch nicht untersucht, inwiefern unsere eigene Intention, Geräusche mit einer Bewegung zu erzeugen, diese Vorhersageprozesse beeinflusst. Während die Geräuschproduktion z.B. beim Sprechen oder Singen unser intendiertes Ziel ist, wollen wir z.B. die Geräusche unserer Schritte auf dem Boden beim Gehen nicht erzeugen. Diese entstehen eher als ein zufälliges Nebenprodukt der Bewegung. Ihr Ausbleiben würde uns jedoch vermutlich doch überraschen.

Die vorliegende Arbeit hat das Ziel, die divergierende Relevanz von zufälligen und intendierten Bewegungsgeräuschen für die Bewertung und neuronale Verarbeitung von Bewegungen näher zu beleuchten. Mit diesem Ziel wurden drei funktionelle Magnetomographie (fMRT)-Studien durchgeführt, die Hürdenlauf (zufällige Bewegungsgeräusche) und Stepptanz (intendierte Bewegungsgeräusche) untersuchten. Die erste Studie ermittelte die zugrundeliegenden neuronalen Netzwerke, bestehend aus dem primären auditorischen Kortex, dem posterioren superioren temporalen Gyrus (pSTG) und dem supplementär-motorischen Areal (SMA). Sowohl die gedämpfte Verarbeitung im primären auditorischen Kortex als auch die Vorhersageprozesse in pSTG und SMA waren für intendierte Bewegungsgeräusche ausgeprägter. Auditorisches „Zerhacken“ beeinträchtigte außerdem die behaviorale Bewertung von intendierten Bewegungsgeräuschen stärker. Die zweite Studie untersuchte den Einfluss des Wegfalls von Geräuschen. Das Fehlen von Bewegungsgeräuschen reduzierte die Bewertung nur für intendierte Bewegungsgeräusche, die zu einem gewissen Grad

durch Vorhersagen des SMA wiederhergestellt werden konnten. Die dritte Studie, die den Einfluss einer zeitlichen Verzögerung von Bewegungsgeräuschen untersuchte, stärkte die Annahme, dass das SMA zusätzliche Vorhersageinformationen bereitstellt, um eine reibungslose Bewertung zu ermöglichen. Dies ist besonders zutreffend, sobald intendierte Bewegungsgeräusche verzerrt sind.

Die Ergebnisse dieser Arbeit legen einen Grundstein für das bessere Verständnis der Vorhersagemechanismen, die bei der Verarbeitung von Bewegungsgeräuschen in Abhängigkeit der Intentionalität der Geräuschproduktion eine Rolle spielen.

List of Original Publications

This thesis is based on the following original research articles:

- Study I. Heins, N., Pomp, J., Kluger, D.S., Trempler, I., Zentgraf, K., Raab, M., & Schubotz, R.I. (2020). Incidental or Intentional? Different Brain Responses to One's Own Action Sounds in Hurdling vs. Tap Dancing, *Frontiers in Neuroscience*, 14, 483
- Study II. Heins, N., Pomp, J., Zentgraf, K., Raab, M., & Schubotz, R.I. (submitted to journal). Where is my sound? Omission of one's own intentional and incidental action sounds and its effect on BOLD fMRI.
- Study III. Heins, N., Trempler, I., Zentgraf, K., Raab, M., & Schubotz, R.I. (submitted to journal). Too late! Influence of temporal delay on the neural processing of one's own incidental and intentional action sounds.

1 Theoretical and Empirical Background

1.1 What are action sounds/ auditory re-aferences?

From moving your fingers over the strings of a guitar to walking on a gravel path in the park, our physical actions usually create a unique set of sounds, whether intended or not. It seems intuitive that the deprivation or alteration of these sounds would confuse us – just imagine writing an e-mail on your computer and not hearing the familiar clattering of your keyboard – you would surely do a double-take. Still, it remains unclear whether we need all action sounds that we produce to execute the corresponding action correctly, or whether we would still evaluate some actions as successful even when they do not evoke the expected sounds – or no sound at all.

Action sounds are omnipresent in our everyday life and have an overwhelming influence on how we perceive the things around us. Sound influences how stale or crisp we perceive our potato chips (Zampini & Spence, 2004) and if our electric toothbrush works (Zampini, Guest, & Spence, 2003). The sound of a car door closing does not only indicate that the door is properly shut, but also influences the perceived luxury of the car, and is artificially engineered to sound satisfying (Parizet, Guyader, & Nosulenko, 2008).

Action sounds that arise from our own actions certainly play a special role in the perception of our auditory world. The term “auditory re-aferences” has been used as a synonym for these self-initiated action sounds. It implies that afferent sensory information is reused in the motion cycle and may therefore act as important auditory feedback in motor control. Indeed, a number of studies suggest the necessity of action sounds for a smooth and optimal action execution, whether it is speaking, playing a musical instrument (for a review, see Pruitt & Pfordresher, 2015) or doing sports (for a review, see Schaffert, Janzen, Mattes, & Thaut, 2019). On the contrary, other studies do not find a positive impact of auditory feedback on motor

control for e.g. music production (Gates, Bradshaw, & Nettleton, 1974; Pfordresher, 2006) or sports (Kennel et al., 2015). Although action sounds are an integral part of our life and we seem to expect or even predict them, their overall relevance for action evaluation and execution is yet to be resolved. Especially the possible difference between the processing of action sounds that are the intentional goal of an action (goal-related action sounds, G action sounds, thereafter) and action sounds elicited as an action by-product (by-product action sounds, B action sounds) has not been examined before. Knowing whether we speak too loud or not loud enough and whether we pronounce words correctly is crucial in our everyday life, where we constantly interact with others and want to be understood. While perceiving our own action sounds seems obviously important in the case of speech (a G action sound), other action sounds, at first glance irrelevant, might still hold some relevance. Even the contact sound of a glass placed on a table (a B action sound) might provide us with the important feedback whether we used the appropriate amount of force, and whether we have to adjust our action trajectory accordingly in the future.

It is the purpose of this thesis to provide the foundation for this topic, and to help to unravel the relevance of action sound processing. The following introductory sections will illuminate the relevance of action sounds, starting with the perception of self-initiated sounds (Section 1.2), before embedding action sounds in the framework of predictive coding (Section 1.3), focusing on the mismatch negativity (MMN) as a correlate of prediction error, and the presumed hierarchy of predictions in the brain. The most examined human action sounds – language and music – and the effects of omission and altered feedback on their execution and perception are illuminated, as well as the scarce research on other human action sounds (Section 1.4). This is used to derive the main purpose of this thesis – the investigation of the distinction between action sounds created intentionally (G action sounds) and action sounds created incidentally (B action sounds). Further objectives are illustrated thereafter, outlining the main

hypotheses and research questions (Section 2) which are then answered in the three original research papers (Section 3). The results are summarized and discussed (Section 4), before reaching the conclusion of this thesis (Section 5).

1.2 Self-initiated action sounds and sensory attenuation

When we walk down a dark street at night, the information whether the footsteps that we hear are our own or someone else's is of decisive importance to us, determining whether we are alone or whether someone might be lurking in the dark behind us. Luckily, we as humans are easily able to make this distinction in the same way as we immediately recognize our own voice. Indeed, self-recognition through sounds alone is not limited to "everyday" sounds that we perceive regularly, but also applies to less familiar action sounds. This is even the case when sounds are played back at us via recordings, and not the immediate "online" results of our actions. For example, hurdlers can distinguish their own hurdling sounds from those of others (Kennel, Hohmann, & Raab, 2014), golfers are able to identify the sound of the golf club swing that they created (Murgia, Hohmann, Galmonte, Raab, & Agostini, 2012), and we are generally capable to differentiate between our own clapping sounds and those of others (Repp, 1987). We seem to be experts in identifying the sounds that we create, and have the ability to distinguish own action sounds from others.

The idea that the sensory consequences of our own actions have to be distinguished from sensory input by other sources and are therefore processed differently is not a novel account. Hermann von Helmholtz was one of the first to study the question of how we distinguish the sensory results of our own actions from externally created stimuli. In this context, he first mentions the concept of an efference copy of a motor command that is generated when a movement is executed. This efference copy, he speculated, "cancels out" the sensory feedback that is produced by the movement. He thus theorized that we have an internal

representation of the consequences arising from our own movements (von Helmholtz, 1867). This idea was revisited almost one hundred years later by Erich von Holst and Horst Mittelstaedt (von Holst & Mittelstaedt, 1950), and Roger Sperry (Sperry, 1950). They deduced that the simplest method to distinguish between self-generated and externally created sensory input was to cancel the predictable consequences of our own actions. This is accomplished by an efference copy of the motor command, which is translated into the sensory system to suppress the re-afferent sensory signal. This process helps both to increase sensitivity for environmental sounds and to detect errors in the self-generated acoustic output to help correct speech and musical output in both humans and animals (cf. Schneider & Mooney, 2018). In both humans and animals, the middle ear muscles contract during vocalization to attenuate the sound of one's own voice (Mukerji, Windsor, & Lee, 2010), with the stapedius reflex being the dominant pathway in humans (Liberman & Guinan, 1998). Indeed, our voice would otherwise be so loud that it would yield us temporally deaf (cf. Bendixen, SanMiguel, & Schröger, 2012). Self-generated sounds during locomotion are usually attenuated in animals, e.g. mice. This mechanism is still more closely coupled to survival in prey animals – without the attenuation of own locomotion sounds, animals would be temporarily unable to detect faint background environmental sounds, like the approach of a predator (Poulet & Hedwig, 2002).

The attenuation in auditory cortex is most likely frequency-specific, so that not all self-generated sounds are suppressed, but only those matching the frequency of the expected output, while tone detection of other (unexpected) sounds is enhanced (Rummell, Klee, & Sigurdsson, 2016; Schneider, Sundararajan, & Mooney, 2018). Self-generated sounds are usually perceived fainter and less salient than externally produced input on a subjective level (Sato, 2008; Weiss, Herwig, & Schütz-Bosbach, 2011). This effect is not unique to the auditory domain, but has been reported for other sensory modalities as well. The most prominent example might be that we are unable to tickle ourselves (Blakemore, Wolpert, & Frith, 2000). On a neural basis,

sensory attenuation is reflected in a dampened N1, a negative evoked potential measured by electroencephalography (EEG) and magnetoencephalography (MEG), peaking 80 to 120 msec after stimulus onset (for a review, see Näätänen & Picton, 1987). Reduced N1 responses to self-initiated vs. externally generated stimuli in EEG (Baess, Horváth, Jacobsen, & Schröger, 2011; Timm, Schönwiesner, SanMiguel, & Schröger, 2014; Timm, Schönwiesner, Schröger, & SanMiguel, 2016) and MEG (Aliu, Houde, & Nagarajan, 2009; Martikainen, Kaneko, & Hari, 2005) have mostly been examined for button presses performed by hand movements, sounds produced by the lower limbs (Van Elk, Salomon, Kannape, & Blanke, 2014) and – to a lesser extent – eye movements creating sounds artificially (Mifsud & Whitford, 2017). The attenuation, i.e. the reduced neural processing, is usually interpreted as a forward model expecting the precise sensory consequences of our own actions (Miall & Wolpert, 1996; Wolpert, Ghahramani, & Jordan, 1995). While sensory attenuation has been thought to be tied to self-initiation of sounds and the sense of agency for a long time, the precise origin remains unknown (for critical reviews, cf. Horváth, 2015; Hughes, Desantis, & Waszak, 2013). Recent studies suggest that the predictability of arising stimuli might be more relevant than the self-initiation (Kaiser & Schütz-Bosbach, 2018). This is in line with studies showing that N1 attenuation does not only occur when comparing self-generated to externally generated sounds, but also when comparing sounds resulting from the actions of another human effector to externally generated sounds (Ghio, Scharmach, & Bellebaum, 2018; Sato, 2008).

The idea of a shared representation of action execution and action perception has a long history in psychology and neuroscience, originally proposed by William James, and examined and discussed more recently in “mirror neuron” research (reviewed in Rizzolatti & Craighero, 2004). The shared representation is also postulated in the theory of event coding (Hommel, Müsseler, Aschersleben, & Prinz, 2001), stating that actions are encoded in their sensory consequences. Taking this into account, it seems likely that it is not an inherent quality of self-

production that causes the sensory attenuation of self-initiated sounds. Predictive processes related to the fact that we are able to anticipate and predict sensory consequences of actions and external events (Schubotz, 2007) seem to be more relevant.

It seems amendable to move on from motor-based forward models when evaluating action outcomes such as action sounds (for a review, see Dogge, Custers, & Aarts, 2019). Predictive coding, no longer limited to pure perceptual processes (for a review on active inference, see Adams, Shipp, & Friston, 2013), might be a more suitable framework.

1.3 The predictive coding approach to action sounds

1.3.1 Predictive coding and perception

The world around us is messy, and we are bombarded with myriads of sensory impressions every day. By generating an internal model of the world out there, our brain tries to make sense of it. We perceive our environment as relatively stable (cf. Denham & Winkler, 2020), and some go as far as calling perception a “controlled hallucination” (Keller & Mrsic-Flogel, 2018), while others assume perception is just the process of identifying which hypothesis fits our sensory percepts best (cf. Walsh, McGovern, Clark, & O’Connell, 2020). While these internal models help us deal with our environment and navigate our everyday life, it is yet unresolved how the brain succeeds in creating this internal representation of the world (Heilbron & Chait, 2018; Schröger, Kotz, & SanMiguel, 2015). Predictive coding (or predictive processing) is an appealing theory explaining how our brain deals with the outside world, going as far as postulating a “unified theory of the brain” (Friston, 2010). The framework regards the brain as a predictive machine, prone to reduce its prediction errors. Prediction errors are thought to be calculated as the difference between expected and perceived stimuli. This difference or “mismatch” is then directed forward to higher processing areas. This significantly reduces the amount of sensory data our brain has to process, rendering the process more efficient, as only

the residual error signal is regarded as new information. Neuronal responses thus reflect only prediction errors (Friston, 2005).

Predictive coding is one of the preferred frameworks for the study of perception, and was initially postulated in the visual domain (Rao & Ballard, 1999, Lee & Mumford, 2003). In the auditory domain, events usually happen in fast succession, organized in sequences rather than single isolated events. Additionally, competing stimuli overlap as different sound sources are usually active at the same time. Filtering the sensory input is crucial, as only a fraction of what we perceive can be examined in detail. In auditory scene analysis, prediction might help to reduce the rich auditory environment by subtracting predictable input (cf. Bendixen et al., 2012). This is probably one of the reasons why predictive coding has recently gained more popularity in audition (for reviews, see Carbajal & Malmierca, 2018; Denham & Winkler, 2020; Heilbron & Chait, 2018). Being able to “filter out” predictable auditory information, e.g. the familiar background noises of your workplace, renders us more capable to react to changes in the environment, e.g. the sound of a fire alarm. The neural processing of unexpected signals is explained in more depth in the following section.

1.3.2 The Mismatch Negativity (MMN) as a prediction error indicator

The numerous studies on the mismatch negativity (MMN) component in EEG and MEG supply valuable information about the processing of unexpected sensory signals, or prediction errors. MMN is usually examined in an oddball paradigm, where the same sound is presented most of the time (standard) with an occasional deviating tone (“oddball” or deviant). It is a negative component representing a difference wave between these standard and deviant stimuli, and can usually be observed between 100 to 250 msec after stimulus onset. The MMN is one of the most studied components in the predictive coding framework, because it seems an adequate representation of the prediction error signal (Garrido, Friston, Kiebel, & Stephan,

2009; Wacongne, Changeux, & Dehaene, 2012). First observed by Näätänen and colleagues (Näätänen, Gaillard, & Mäntysalo, 1978), it was initially interpreted as a memory-based process, where the actual input is compared to a memory trace of previously perceived stimuli. In this interpretation, it can be seen as an automatic auditory change detector. Different theories have been applied to explain the MMN, evolving from memory-based to prediction-based interpretations (Garrido, Kilner, Stephan, & Friston, 2009; Schröger, Kotz, et al., 2015). Most importantly, to rule out a pure adaptational explanation of this component, omission paradigms have been applied to show that it is most likely based on predictive processes. Here, instead of presenting a deviant tone in a series of standards, tones are unexpectedly omitted. If the MMN was a pure adaptational component, representing cancelled repetition suppression to a “new” sound, it should not be evoked when a sound is omitted. The omission, however, evokes an error response similar to the MMN, which might be slightly earlier than the deviant MMN (cf. Wacongne et al., 2011), but is nevertheless an indicator that a sound was indeed expected and its unexpected omission causes an error response. Importantly, this error response evoked in primary auditory cortex and superior temporal gyrus is only elicited after unexpected omissions and when an identity prediction of the omitted sound is possible (Dercksen, Widmann, Schröger, & Wetzels, 2020; SanMiguel, Saupe, & Schröger, 2013).

The results from omission studies are additionally congruent with the postulated hierarchy of prediction and prediction errors in predictive coding (for a review, see Heilbron & Chait, 2018). This is usually examined with “improved” oddball paradigms, where in different experimental blocks, various deviants/and or omissions are expected. Hence, there is a low-level prediction for each (five tone) sequence of standards and deviants, and a higher-level prediction regarding the occurrence of each “sequence type” in an experimental block. Certain sounds or their omission therefore elicit two prediction errors – one on the lower “local” level within an individual sequence, and one on the higher “global” level. The MMN is most likely

a local deviant detector (Kompus, Volehaugen, Todd, & Westerhausen, 2019) whereas the omission response reflects a more global violation (Wacongne et al., 2011).

1.3.3 Cortical hierarchy of action sound prediction

One of the key assumptions of predictive coding is that the brain works in a hierarchical fashion – represented by error units having forward connections and prediction units providing expectations about upcoming events via backward connections. While there is still no direct prove of the existence of different units – error units and prediction units - in different cortical layers (cf. Summerfield & Egner, 2009), with error units assumed to be predominantly in cortical layers II/III and prediction units in layers V/VI (Felleman & Van Essen, 1991), the hierarchical fashion in which the brain works seems to align with predictive coding assumptions, as there are asymmetrical connections with more top-down expectation input than bottom-up sensory information forwarding. This hierarchy is also reflected in distinct brain areas responsible for the different kind of prediction errors/predictive input. While there is diverging evidence about the precise areas involved, temporal auditory areas seem to elicit the prediction error signal and frontal, premotor areas seem to provide the predictive input (see Fig. 1). More precisely, the superior temporal gyrus has been found to be one of the main MMN generators (Liebenthal et al., 2003; Mathiak et al., 2002), with exact peaks varying from the primary auditory cortex in Heschl’s gyrus (Nazimek, Hunter, Hoskin, Wilkinson, & Woodruff, 2013), to more posterior areas, labelled posterior superior temporal gyrus (pSTG, Recasens, Grimm, Capilla, Nowak, & Escera, 2014). In addition, the mismatch response resulting from sound omissions is also suggested to be evoked in pSTG (Fonken et al., 2019; Raij, McEvoy, Mäkelä, & Hari, 1997). There seems to be an additional hierarchy within the auditory cortex – more anterior segments of the STG code more basic prediction errors in simple stimuli, whereas more complex prediction errors (e.g. language, music, sound sequences) are elicited in pSTG.

A primary source region for the predictive input in sound processing is the supplementary motor area (SMA). It is more activated when sounds are created actively compared to listening to a replay (Krala, van Kemenade, Straube, Kircher, & Bremmer, 2019). This is in accordance with research investigating SMA in the context of rhythm processing and both musical production and consumption (Lima, Krishnan, & Scott, 2016; Nachev, Kennard, & Husain, 2008). SMA is also activated by more global prediction errors, i.e. violations of abstract rules (Dürschmid et al., 2016). A recent study (Jo, Habel, & Schmidt, 2019) combines these three areas (primary auditory cortex, pSTG, SMA) in a dynamic causal modelling (DCM) approach to investigate the relevance of SMA for sensory attenuation, and postulates a tight interplay and reciprocal interactions between SMA and auditory cortices. The hierarchy is reflected in oscillatory activity as well. Gamma-band activity in auditory cortices increases in case of a prediction error, while beta oscillations originating in frontal areas seem to provide predictive information (Arnal & Giraud, 2012).

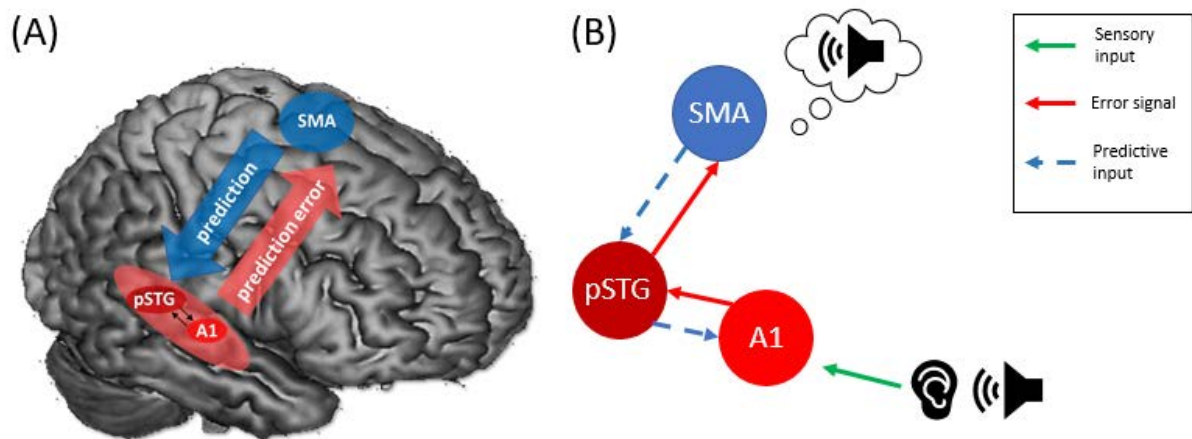


Fig. 1. Predictive hierarchy of action sound processing. (A) Temporal auditory areas, including the primary auditory cortex and the posterior superior temporal gyrus, propagate the prediction error to frontal areas, especially the supplementary motor area. SMA provides predictive information to temporal areas. (B) Sensory auditory input is evaluated based on the predicted sensory input. Predictions are transferred top-down from SMA, whereas the error signal is propagated bottom-up from auditory cortices.

Although far from conclusive, the results obtained so far support a cortical hierarchy between early auditory areas, transmitting basic prediction errors, secondary auditory areas, dealing with prediction errors in more complex stimuli, and frontal areas, examining the big picture and reaction to global rule violations (cf. Schönwiesner et al., 2007).

One major problem in the investigation of action sounds is the low ecological validity of the sounds used. Real-life auditory input is usually sequential (Bendixen et al., 2012), and therefore button presses eliciting single pure tones might not be representative of real-world circumstances. Although sequences of pure tones that are used to establish both local and global rules about stimulus occurrence at least consider the arrangement of sounds in a sequence, they lack the complexity that real-life action sounds possess.

To understand ecological action sounds better, I therefore present the research covering language production and perception, as well as music production. These two action sounds are the most thoroughly examined, considering that they result from actions that are – usually – solely executed with the intention of creating a sound. Additionally, I provide a short overview about research done on human action sounds which are a by-product of our actions but have nonetheless been found to be important in e.g. a sports context.

1.4 Action sounds, effects of their omission and interference, and their neural representation

1.4.1 Language

To understand the relevance of action sounds as auditory feedback, and especially the potential differences between action sounds produced intentionally and action sounds elicited as an incidental by-product, it is important to review the existing literature. In the case of speech and music, it seems obvious that the auditory output is the goal of the action, and therefore it is

intuitive that the action sounds are important for accurate action performance and evaluation. It is thus not surprising that the investigation of action sounds is dominated by those two types of intentional action sounds. Speech is probably the most complex human motor action, involving about 100 muscles. A human speaker can produce up to 15 phonemes a second (Levelt, Roelofs, & Meyer, 1999), indicating that speech production and perception are not only extremely complex, but also fast processes. It is unlikely that they are controlled by sensory feedback alone, implying a predictive feedforward component as well. Paradigms applying disruptions and thus eliciting prediction errors are used in order to examine the predictive processes involved in language production and perception.

Sound omission, i.e. the masking or complete silencing of action sounds, and altered feedback, i.e. an artificial change of action sounds, are the most commonly used methods to examine prediction errors and the relevance of auditory feedback in speech motor control (for a review, see Parrell & Houde, 2019).

Masked feedback studies usually find differences between masked and clear speech in several aspects of speech, including pitch, voice quality and volume (Harlan Lane & Tranel, 1971; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). However, loud noise was used to mask speech feedback, which is an important limitation of these studies, as they might not represent true feedback omission (cf. Parrell & Houde, 2019). Studies with participants using a cochlear implant, where speech feedback can be completely omitted, might be a more reliable source and show similar results. Additionally, postlingually deaf participants display impaired prosody and articulatory control (Lane & Webster, 1991), and their speaking quality decreases significantly (Perkell, 2013; Waldstein, 1990), resembling the results found in songbirds, who are unable to learn new songs or adaptively modify vocal elements when specific auditory cortex neurons are genetically ablated (e.g. Roberts et al., 2017). The long-term effects of deprived auditory feedback in humans are yet unknown (Brainard & Doupe, 2000).

Altered auditory feedback studies usually use either delayed auditory feedback (DAF) or a qualitatively changed feedback, e.g. pitch or formant changes. Anyone hearing their own voice lag behind during a phone call will automatically confirm the negative effects of such kinds of feedback, inevitably leading to distress (Badian et al., 1979). DAF affects speech fluency significantly (e.g. Fairbanks, 1954; Howell, 2004), affecting a variety of speech components, e.g. inflicting a slowing of speech, stuttering, intonation and phoneme errors (cf. Sasisekaran, 2012). This disrupting effect seems to be greatest with a delay of 200 msec (Stuart, Kalinowski, Rastatter, & Lynch, 2002), which is the typical length of syllable production. Surprisingly, DAF helps participants who stutter with a more fluent speech production (Lincoln, Packman, & Onslow, 2006). Externally applied perturbations to speech (e.g. formant changes, changes in loudness) evoke a compensatory response (Bauer, Mittal, Larson, & Hain, 2006; Houde & Jordan, 2002; Purcell & Munhall, 2006b). This compensation is more complete when auditory feedback is provided, suggesting that auditory feedback is essential for maintaining accurate speech production (Jones & Munhall, 2003). However, these adaptive learning processes show great variance between individuals (Lametti, Nasir, & Ostry, 2012; Purcell & Munhall, 2006a).

The use of auditory feedback has also been examined in the brain, with studies conveying its importance in speech production. One important phenomenon is speaking induced suppression, where participants' auditory cortex activation is reduced to self-produced speech compared to speech produced by others, resembling the sensory attenuation effect (Ventura, Nagarajan, & Houde, 2009). Thus, it seems that the motor act of speaking enables the auditory cortices to anticipate its auditory consequences. This suppression vanishes when auditory feedback is altered, e.g. delayed (Christoffels, van de Ven, Waldorp, Formisano, & Schiller, 2011). Effects of altered auditory feedback yield an activation in the posterior superior

temporal gyrus (Hashimoto & Sakai, 2003; Tourville, Reilly, & Guenther, 2008; Zheng, Munhall, & Johnsrude, 2010).

All in all, results from speech production and monitoring support the intuitive importance of action sounds for proper action execution.

1.4.2 Music

The intentionality of sound production in music is similarly intuitive as in speech. Looking at it from a first-person perspective, e.g. the finger movement of a pianist to perform a piano key stroke has the sole purpose to create the resulting sound (cf. Novembre & Keller, 2014). Although auditory feedback thus seems equally important in music production as in speech, effects of silenced and altered feedback are less clear. Some studies examining omitted feedback found no significant effect (Pfordresher, 2006) and no higher error rates during performance (Gates et al., 1974). Other studies suggest that there are in fact higher error rates (Pfordresher & Beasley, 2014) and that musical expressiveness decreases (Repp, 1999).

Incongruent action sounds presented before action execution delayed the execution significantly (Drost, Rieger, Brass, Gunter, & Prinz, 2005a) and led to more errors (Drost, Rieger, Brass, Gunter, & Prinz, 2005b), but only when the musical timbre of the incongruent sound matched the participants' instrument (guitar vs. piano, Drost, Rieger, & Prinz, 2007). Auditory perception hence seems to prime action execution, resulting in a slower and more erroneous performance when the auditory percept is conflicting with the planned action. Serially shifted auditory feedback evoked the most errors (Pfordresher, 2005; Pfordresher & Palmer, 2006). These results are limited to participants with some level of musical expertise. Musical experts are generally better in suppressing interfering auditory feedback (Pfordresher, 2012). Other studies examining influences of altered auditory feedback when singing, revealed that trained/expert singers were able to compensate for errors more quickly than novices (Jones

& Keough, 2008; Keough & Jones, 2009; van der Steen, Molendijk, Altenmüller, & Furuya, 2014).

Representations of action execution can also be elicited by mere musical perception, suggesting that execution and perception are inherently coupled. This is implied by studies showing that even the presentation of unknown piano sequences evokes activity in motor planning areas in trained pianists (Bangert et al., 2006), and even the presentation of mute piano actions activates these regions (Hasegawa et al., 2004; Haslinger et al., 2005), suggesting multisensory coupled representations in musicians' brains. These results are however not limited to musically trained participants. Lahav and colleagues (Lahav, Saltzman, & Schlaug, 2007) trained non-musician participants to perform short musical sequences and then presented them with several musical sequences in an fMRI (functional magnetic resonance imaging) experiment. Motor cortex activation was greatest for those musical sequences that were previously learnt, showing that the sensorimotor coupling can occur even after a short amount of training. Having a motor representation of the perceived sounds seems to be necessary, however – visual experience in absence of motor learning is not sufficient (cf. Candidi, Sacchi, Mega, & Aglioti, 2014).

Disrupting auditory feedback in music accordingly modulates activity in both auditory and (pre-)motor areas. Pitch-altered feedback in piano performance influences the BOLD (blood-oxygen-level-dependent) response in the SMA (Pfordresher, Mantell, Brown, Zivadinov, & Cox, 2014), and musicians' superior performance in a temporal asynchrony judgement task results from a better connectivity between superior temporal sulcus (STS), the cerebellum and premotor areas, and is interpreted as a fine tuning of internal models (Lee & Noppeney, 2011).

Similar to speech, auditory “online” feedback cannot be the only adjusting mechanism. Otherwise, we would not be able to synchronize our musical production perfectly with others,

whether singing in a choir or playing an instrument in a musical ensemble. Joint musical action is probably achieved by clear predictions of the sounds to come. Predictions are especially accurate for self-induced action sounds. This is reflected in a better synchronization of pianists to recordings of their own performances compared to performances of others (Keller, Knoblich, & Repp, 2007), as well as a better synchronization to pianists with a similar preferred performance tempo and expressiveness (Loehr, Large, & Palmer, 2011).

Predictive processes are thus undeniably important in musical action execution as well.

1.4.3 Action sounds apart from language and music

Although the presented research underlines the significance of action sounds for both action perception and action execution, research on real-life action sounds apart from language and music is scarce, especially when it comes to the effects of omitted or altered auditory feedback. Studies examining audiovisual asynchrony perception, however, generally find differences between language and music (G action sounds) and sounds created by object actions (B action sounds), like a hammer hitting a peg (Dixon & Spitz, 1980), a soda can being crashed (Vatakis & Spence, 2006) or the sounds created by playing chess (Eg & Behne, 2015). A study investigating behavioural effects of acoustic hand-object contacts additionally indicated that incongruent action sounds impede accurate action execution (Castiello, Giordano, Begliomini, Ansuini, & Grassi, 2010). Similarly, delayed presentation of own action sounds when walking significantly decreased the sense of agency (Menzer et al., 2010), while masked feedback during rowing made action execution subjectively more demanding (Schaffert, Oldag, & Cesari, 2020). On the contrary, masked action sounds in hurdling did not affect the performance, and delayed action sound presentation was only disruptive upon the first trial (Kennel et al., 2015).

As it is not possible to examine the execution of any of these actions in an fMRI setting, there are, to my knowledge, no studies examining the possible surprise and adaptational effects to disrupted auditory feedback in the brain. However, presentation of familiar sport sounds activated both auditory and motor planning areas (Woods, Hernandez, Wagner, & Beilock, 2014), suggesting that action sounds activate a representation of the action necessary to create them.

The research on action sounds apart from language and music is far from conclusive, and at this point, it is unclear whether action sounds have the same relevance for action execution and action evaluation when they are not the intended outcome of the performed action. To answer the question whether intentional and incidental action sounds differ in their behavioural relevance and neural representation, we conducted three experiments, outlined in the following section.

2. Research Questions

The research presented above suggests a tight coupling of motor control and action sound perception, irrespective of the aspect whether the actions create sound as an intended goal or whether the sounds produced are just an incidental by-product. Intuitively, we would say that we expect the sounds of our own footsteps while walking to the same extent that we expect to hear the spoken word while speaking. We seem to plan our actions with their sensory consequences in mind, whether those are intended or not. Nevertheless, while the absence of the sound of our footsteps (B action sound) would most definitely surprise us, we would not deem them the goal of our action, and would probably still evaluate our walking action as successful. On the contrary, the absence of our voice when trying to speak (G action sound) would not only surprise us, but would also be regarded as a failure to achieve the goal of our speaking action. Differences between those two types of action sounds have not been examined before, and overall, research using real-life action sounds is scarce. B and G action sounds can differ both regarding their perceptual expectations and their neural processing, and it seems necessary to examine their potential differences both in regard to their psychophysiological processing and performance evaluation, and their neuronal representations. The behavioural studies by Kennel and colleagues (Kennel et al., 2015) suggest that performance can to some extent be decoupled from action sounds, i.e. is not affected by disrupted feedback, at least in B actions, which could indicate a difference in the perceptual expectations of incidental vs. intentional action sounds. However, it is unclear how the brain manages this decoupling and whether there are neurofunctional compensatory processes at work which enable a stable performance even with disruptive auditory feedback. It is possible that the brain shifts away from the actual sensory input and focusses on the predictive model instead (this is comparable to reading only what you expect to read and not what you have actually written when reading your own thesis again and again, so that you still miss errors). While interference and

compensation are usually processes affecting behavioural measures like a performance rating, this is not always the case, as the brain can somewhat compensate for the collapse of behavioural measures (Ward, 2006). The examination of the involved neuronal proceedings therefore seems absolutely crucial.

It is well established that sensorimotor representations are triggered by action observation (e.g. Sato, 2008). We predict during action observation much the same way as we do during action execution. This is especially important as it is not possible to perform complex movements and record brain activity at the same time, because fMRI scans are continuously disrupted by motion. An observational paradigm is thus the only way to observe the processing of action sounds and the related predictions in the brain. As the sensorimotor representations need a motoric basis (cf. Novembre & Keller, 2014), it is crucial for the participants to be motorically familiar with the presented actions and their action sounds. Presenting participants with their own past actions during an fMRI experiments seems the closest to actual action execution one can get.

Hurdling is a real-life whole-body action that creates incidental B action sounds. The action of hurdling and the resulting sounds have been examined in multiple regards by Kennel and colleagues, who determined that action sounds are not crucially important for a correct action execution (see Section 1.4.3). It was thus logical to use hurdling as the B action in our studies as well. Finding a similar whole-body action with the feet as a sound effector and intentional sound production, we decided to use tap dancing as the G action. The participants underwent training sessions in both actions for nine weeks, training both tap dancing and hurdling for three hours per week in two 90-minutes sessions. Each participant was equipped with point-light markers and sound recording devices and filmed and recorded a total of four times during this training period. Participants were subsequently presented with their own actions and action sounds in three separate fMRI studies (for an overview, see Fig. 2).

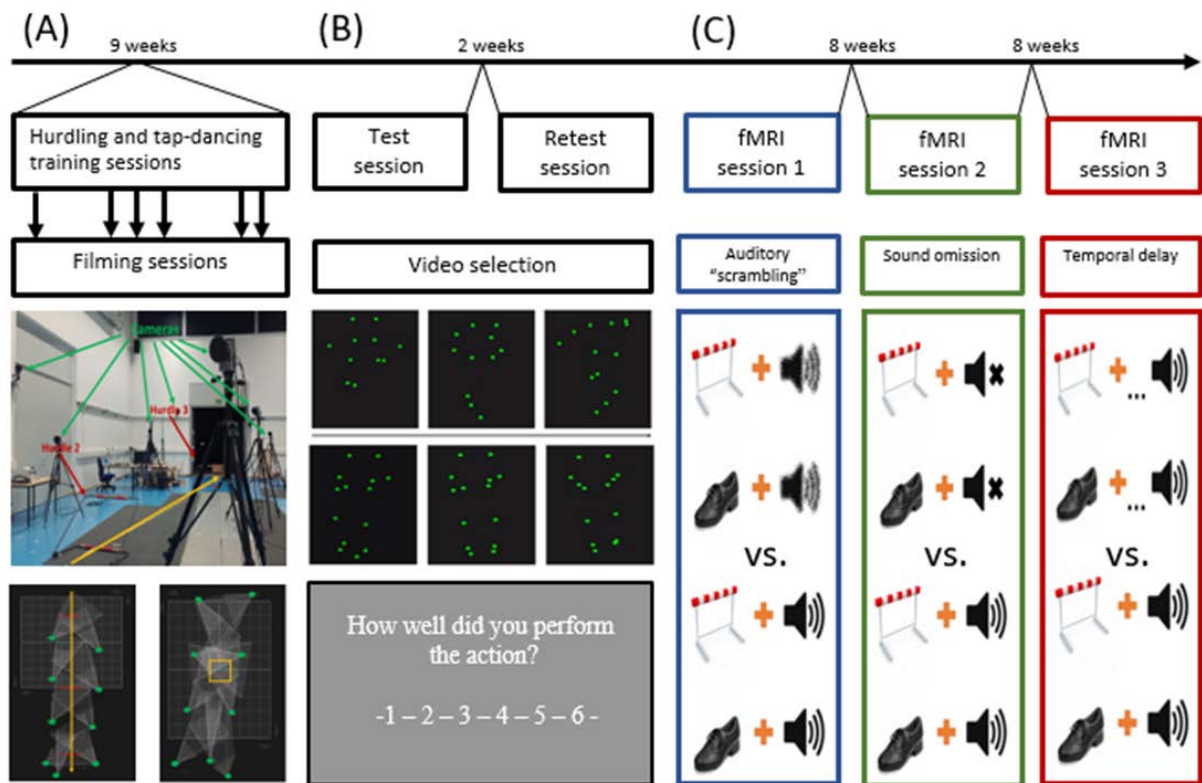


Fig. 2. Procedure of the three studies. (A) Participants engaged in hurdling and tap-dancing training sessions, spanning over a period of nine weeks. During this period, they were filmed during action execution by infra-red cameras (marked in green). They had to run a hurdling track (represented by the yellow line), including three hurdle clearances (marked in red) for the hurdling action. The tap-dancing action was a dance sequence learnt during the training sessions, and performed in a designated area (represented by the yellow square). Camera positions during filming are represented by the green dots in the bottom panel, shown from a bird's-eye view. (B) Participants took part in two behavioural sessions (two weeks apart), where they were familiarized with the point-light videos and corresponding sounds of their own actions. They had to rate how well they performed the action on a 6-Point Likert Scale (1 = low, 6 = high), and videos with the most reliable ratings were selected for the subsequent fMRI sessions. (C) Three fMRI sessions were conducted with an eight-week period between two subsequent fMRI sessions. Sessions varied in their main experimental manipulation. In the first session, influences of auditory scrambling on performance evaluation and neural representations of action sounds was examined (blue box), in the second session influences of sound omission (green box) and in the third session influences of temporally delayed sounds (red box).

All three studies examined the diverging importance of action sounds for performance evaluation and neural processing of actions producing sounds either intentionally or incidentally. While all three studies had the goal to investigate the difference between incidental and intentional action sounds, a different focus was applied in each study to provide a clear overview to this new idea of different types of action sounds.

Study I was conducted to confirm the brain regions involved in action sound processing, and to investigate whether there is more sensory attenuation and a stronger predictive component in G vs. B actions. Additionally, different types of “scrambling” were applied to the stimuli to detect whether there is a more pronounced prediction error to disrupted auditory feedback in G vs B actions. This is all based on the notion that G action sounds, being an intended action goal, are more relevant for performance evaluation.

Study II examined the effects of sound omission on the two types of sound-producing actions in order to further establish the higher relevance of action sounds for performance evaluation and more top-down predictive information in G vs. B actions.

Finally, **Study III** investigated the influence of a temporal delay on B and G action sounds, hypothesizing an initial disruption by delayed feedback for both actions, but an adaptational response only for B actions.

3. Research Articles

3.1 Study I: Incidental or Intentional? Different Brain Responses to One's Own Action Sounds in Hurdling vs. Tap Dancing

Running title: Incidental vs. intentional action sounds

Nina Heins, Jennifer Pomp, Daniel S. Kluger, Ima Trempler,
Karen Zentgraf, Markus Raab, & Ricarda I. Schubotz (2020)

Frontiers in Neuroscience, 14, 483



Incidental or Intentional? Different Brain Responses to One's Own Action Sounds in Hurdling vs. Tap Dancing

Nina Heins^{1,2*}, Jennifer Pomp^{1,2}, Daniel S. Kluger^{2,3}, Ima Trempler^{1,2}, Karen Zentgraf⁴, Markus Raab^{5,6} and Ricarda I. Schubotz^{1,2}

¹ Department of Psychology, University of Muenster, Münster, Germany; ² Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Muenster, Münster, Germany; ³ Institute for Biomagnetism and Biosignalanalysis, University of Muenster, Münster, Germany; ⁴ Department of Movement Science and Training in Sports, Institute of Sport Sciences, Goethe University Frankfurt, Frankfurt, Germany; ⁵ Department of Performance Psychology, Institute of Psychology, German Sport University Cologne, Cologne, Germany; ⁶ School of Applied Sciences, London South Bank University, London, United Kingdom

OPEN ACCESS

Edited by:

Marc Schörwiesner,
Leipzig University, Germany

Reviewed by:

Alessandro Tavano,
Max Planck Institute, Germany
Alexandra Bendixen,
Technische Universität
Chemnitz, Germany

*Correspondence:

Nina Heins
nina.heins@uni-muenster.de

Specialty section:

This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Neuroscience

Received: 17 January 2020

Accepted: 20 April 2020

Published: 13 May 2020

Citation:

Heins N, Pomp J, Kluger DS,
Trempler I, Zentgraf K, Raab M and
Schubotz RI (2020) Incidental or
Intentional? Different Brain Responses
to One's Own Action Sounds in
Hurdling vs. Tap Dancing.
Front. Neurosci. 14:483.
doi: 10.3389/fnins.2020.00483

Most human actions produce concomitant sounds. Action sounds can be either part of the action goal (GAS, goal-related action sounds), as for instance in tap dancing, or a mere by-product of the action (BAS, by-product action sounds), as for instance in hurdling. It is currently unclear whether these two types of action sounds—incidental or intentional—differ in their neural representation and whether the impact on the performance evaluation of an action diverges between the two. We here examined whether during the observation of tap dancing compared to hurdling, auditory information is a more important factor for positive action quality ratings. Moreover, we tested whether observation of tap dancing vs. hurdling led to stronger attenuation in primary auditory cortex, and a stronger mismatch signal when sounds do not match our expectations. We recorded individual point-light videos of newly trained participants performing tap dancing and hurdling. In the subsequent functional magnetic resonance imaging (fMRI) session, participants were presented with the videos that displayed their own actions, including corresponding action sounds, and were asked to rate the quality of their performance. Videos were either in their original form or scrambled regarding the visual modality, the auditory modality, or both. As hypothesized, behavioral results showed significantly lower rating scores in the GAS condition compared to the BAS condition when the auditory modality was scrambled. Functional MRI contrasts between BAS and GAS actions revealed higher activation of primary auditory cortex in the BAS condition, speaking in favor of stronger attenuation in GAS, as well as stronger activation of posterior superior temporal gyri and the supplementary motor area in GAS. Results suggest that the processing of self-generated action sounds depends on whether we have the intention to produce a sound with our action or not, and action sounds may be more prone to be used as sensory feedback when they are part of the explicit action goal. Our findings contribute to a better understanding of the function of action sounds for learning and controlling sound-producing actions.

Keywords: human action sounds, auditory action effects, action-effect association, prediction, sensory attenuation

INTRODUCTION

Most actions produce sounds. On a subjective level, we would say that some of these action sounds are the proper goal of the action (*goal-related action sounds, GAS, hereafter*), for instance in musical performance, singing, and speaking; whereas others occur rather as a by-product (*by-product action sounds, BAS*), for instance when we unlock a door or write on our laptop. Although this simple observation suggests potentially different categories of sound-producing actions, it remains to be experimentally addressed whether they indeed differ on the behavioral or neural level. Some findings point toward significant differences between GAS and BAS actions. In speech and musical performance, experimentally distorted or missing action sounds result in poorer performance (Howell, 2004; Pfordresher and Beasley, 2014). Omitted sounds disrupt GAS action performance permanently (Jones and Keough, 2008; Tourville et al., 2008). In contrast, Kennel et al. (2015) investigated the influence of masked and delayed online action sounds during hurdling performance, i.e., a BAS action. Authors found an interfering effect of delayed auditory feedback persisted only for the first trial of performance and vanished afterwards. Moreover, the complete masking of auditory feedback did not even transiently affect participants' action performance. So far, a direct and more detailed comparison of BAS and GAS actions' neural and behavioral processing is missing.

In the present fMRI study, we addressed the potential dissociation of BAS and GAS actions in the framework of predictive coding, suggesting that the brain works as a predictive device and is tuned to minimize its prediction errors (Friston, 2005). According to this model, action sounds are part of the predictive model that is engaged during action execution, or even the observation thereof (Friston, 2012). Neural responses in primary sensory cortices are attenuated for predicted self-generated sensations and the evidence is especially vast for self-initiated sounds (re-afferences, Baess et al., 2011; Kennel et al., 2015; Pizzera and Hohmann, 2015; Mifsud et al., 2016; Rummell et al., 2016; Timm et al., 2016), enabling the immediate registration of prediction-deviant sensations (prediction errors) and effective correction of sound-producing actions (Tourville et al., 2008). With regard to its neural underpinnings, prediction of self-produced sounds is considered to rely on a network consisting of the primary auditory cortex (Heschl's Gyrus) and the posterior superior temporal gyrus (pSTG; Rauschecker, 2012; Heilbron and Chait, 2018) and potentially also the supplementary motor area (SMA; Jo et al., 2019) and the cerebellum (Petrini et al., 2011; Waszak et al., 2012). The latter three structures are suggested to deliver a predictive sound model to primary auditory cortex, causing an attenuation of responses to expected sounds, and mismatch signals for unexpected sounds.

We examined potential differences between GAS and BAS as operationalized by tap dancing and hurdling, respectively. In both tap dancing and hurdling, the lower limbs are the effectors of audible sounds, and sounds produced by the lower limbs seem to show the same prediction-driven sensory attenuation as the more thoroughly examined sounds produced by our

hands (van Elk et al., 2014). We trained naïve participants and filmed them during motor execution to create point-light displays with accompanying action sounds for both tap dancing and hurdling. During a subsequent fMRI session, the same participants were presented with the point-light videos of their own actions and asked to rate the subjective quality of their actions after each video on a 6-point Likert scale. To separately investigate the impact of visual and auditory information on BAS and GAS action processing, we additionally introduced different types of "scrambling" to the action videos serving as selective baseline conditions. Scrambling was applied to either the visual modality, the auditory modality, or both. While leaving the biological motion visually and audibly perceivable, scrambling strongly reduced the information about the quality of action performance. On the behavioral level, we expected overall lower rating scores for auditory scrambled videos, and due to the presumed greater importance of auditory feedback in GAS actions, we hypothesized that this effect would be especially pronounced in the tap-dancing condition.

Correspondingly, we expected tap dancing and hurdling to differ in their BOLD activity in auditory cortices reflecting that action sounds modulate GAS action processing more than BAS action processing. In particular, we expected activity in primary auditory cortex to be more attenuated for GAS as compared to BAS, based on the notion that effective sensory attenuation results from a prediction of sensory action effects (Wolpert et al., 1995; Miall and Wolpert, 1996; Friston et al., 2010; Schröger et al., 2015b). Furthermore, we reasoned that both GAS and BAS actions entail predictions about visuospatial motion patterns, whereas predictions about action sound patterns are pronounced for GAS actions. Regions sending a top-down signal to sensory cortices, especially SMA, the pSTG (Jo et al., 2019) and the cerebellum (Petrini et al., 2011; Waszak et al., 2012) should be more active in GAS than in BAS actions. Finally, prediction errors are suggested to travel up the predictive hierarchy to enable an adaptation of the current predictive model (Phillips et al., 2015; Heilbron and Chait, 2018). Therefore, we expected auditory scrambling to induce a predictive mismatch signal that manifests as increased BOLD response in the pSTG for GAS actions (Fu et al., 2006).

MATERIALS AND METHODS

Participants

The original sample consisted of 19 participants. One participant left the study before finishing the 9-week training of hurdling and tap dancing. Therefore, video and audio data from 18 participants were processed further. Four participants dropped out of the study after the training, so that 14 participants took part in the fMRI session. One participant was excluded from the final analysis, because their reaction times recorded during the fMRI session diverged more than two standard deviations from the mean reaction time, leaving 13 participants (9 females, 4 males) for the analysis. While this is a relatively small sample size, it is comparable to other studies examining action sounds behaviorally (Menzer et al., 2010) or with fMRI

(Reznik et al., 2015). The participants' age ranged from 19 to 28 years ($M = 22.1$, $SD = 2.8$), and all of them were right-handed, as assessed by the Edinburgh Handedness Inventory (EHI; Oldfield, 1971), scores varying from +60 to +100, with a mean of +84. All participants reported to have no history of psychiatric or neurological disorders and signed an informed consent. After successful participation, participants were rewarded with both course credit and monetarily. The study was approved by the Local Ethics Committee of the University of Muenster (Department of Psychology) in accordance with the Declaration of Helsinki.

Material

The stimuli consisted of point-light displays of hurdling and tap-dancing actions with the accompanying sounds, recorded from each participant individually at different stages during training. Point-light displays were recorded using the Qualisys Motion Capture System (<https://www.qualisys.com>) with nine cameras (see **Figure 1**), while the sound was recorded by in-ear microphones (Sound-man OKM Classic II) for hurdling and by a sound recording app on a mobile phone for tap dancing.

After the acquisition, point-light displays were processed using Qualisys, ensuring visibility of all 12 recorded point-light markers during the entire recording time (for an overview of the position of the point-light markers, see **Figure 2**). Note that we selected only videos with error-free performance for our experiment, excluding BAS trials in which the hurdles were touched. Accordingly, all sounds in GAS and BAS were exclusively produced by foot-ground contacts.

Sound data were processed using Reaper v5.28 (Cockos Inc., New York, United States). In a first step, stimulus intensities of hurdling and tap-dancing recordings were normalized separately. In order to equalize the spectral distributions of both types of recordings, the frequency profiles of hurdling and tap-dancing sounds were then captured using the Reaper plugin

Ozone 5 (iZotope Inc., Cambridge, United States). Finally, the difference curve (hurdling—tap-dancing) was used by the plugin's match function to adjust the tap-dancing spectrum to the hurdling reference (see **Figure S1**; examples of the sounds are given in **Videos S1–S8**). Point-light displays and sound were synchronized, and the subsequent videos were cut using Adobe Premiere Pro CC (Adobe Systems Software, Dublin, Ireland). The final videos had a size of 640×400 pixels, 25 frames per second, and an audio rate of 44 100 Hz. A visual fade-in and fade-out of 1 s (25 frames) were added with Adobe Premiere. Video length ranged from 3 to 6 s, with an average length of 5 s.

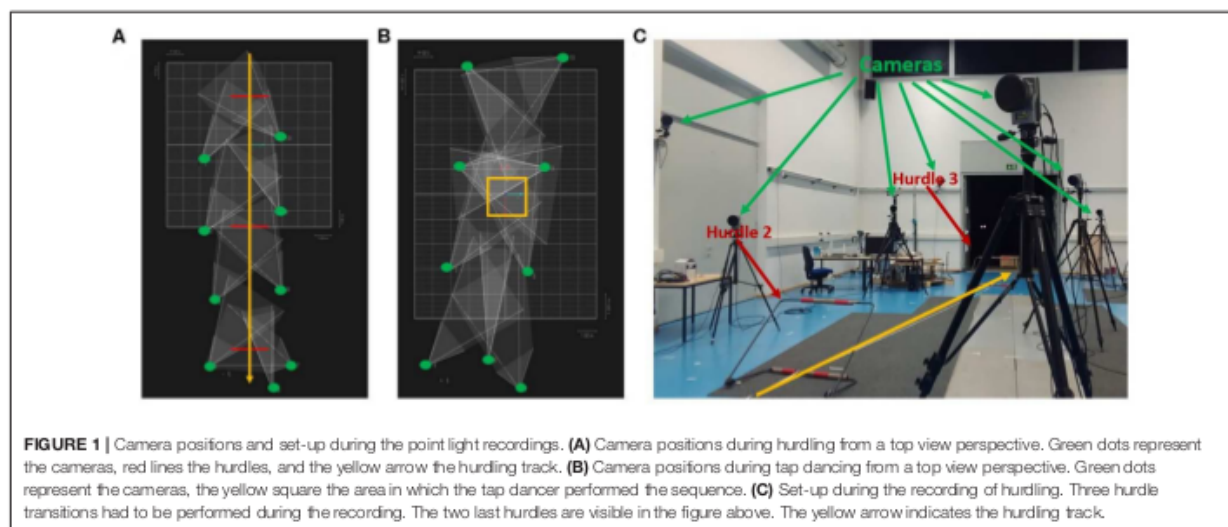
For the fMRI sessions, a subset of 27 hurdling and 27 tap dancing videos was selected for each participant, choosing the videos with the most reliable ratings from the test and retest sessions. For every selected video, additional “scrambled” versions were created using Adobe Premiere. The visual and auditory tracks of the videos were cut into 1-s segments (25 frames) and the segments were then rearranged. The same scrambling scheme was applied to all videos, that is, the segments were rearranged in a fixed order. We created three different types of “scrambling”—either the visual track, the auditory track, or both.

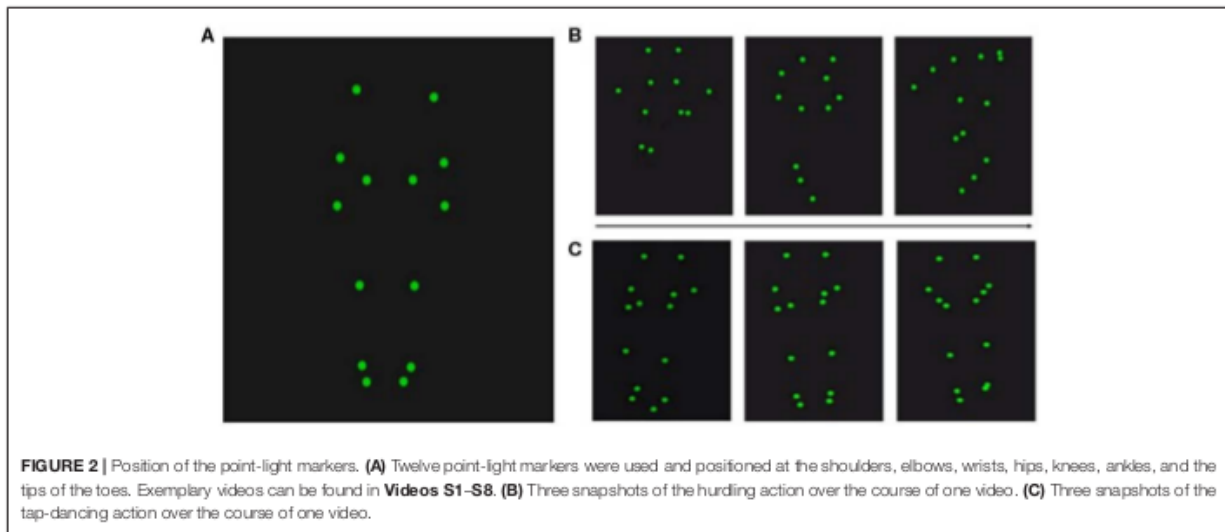
All videos were presented using the Presentation software (Version 18.1, Neurobehavioral Systems, Inc., Berkeley, CA).

Procedure

Training and Filming Sessions

Participants engaged in a 9-week training period during which they were trained in hurdling and tap dancing by professional instructors (**Figure 3**). The training in both hurdling and tap dancing was conducted two times a week, with each training session having a length of 90 min, so that participants trained both action types for 3 h a week. Before this training, none of the participants ever practiced hurdling or tap dancing. During the 9-weeks training period, participants had to take part in four filming sessions, taking place at different states of

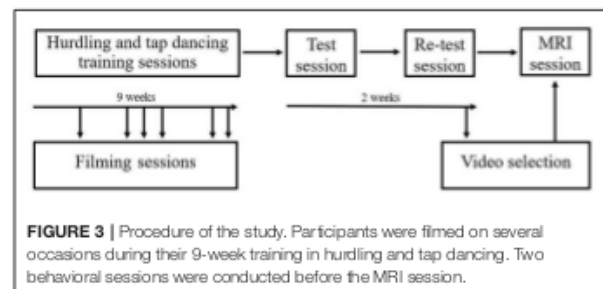




training, to observe changes in performance. The first filming sessions took place 2 weeks after the training started, with the following filming sessions taking place in weeks four, five, six, eight, and nine after training commenced. Participants could choose four sessions from the provided ones. During the filming sessions, participants were equipped with 12 point-light markers (see **Figure 2**) and filmed via infra-red cameras of the motion capturing system while performing both action types. The hurdling action consisted of three hurdle transitions (**Figure 1**), while the tap-dancing action was a movement sequence learned in the tap-dancing training sessions. Both actions increased in difficulty with the four sessions. For hurdling, the spatial distance between the three hurdles increased, requiring more speed. Whereas, for tap dancing, action elements were added to the sequence to increase difficulty.

Behavioral Test and Retest Sessions

Behavioral test-retest sessions were conducted to find the videos with the highest reliability of participants' rating. Both sessions were conducted in a computer lab in the Department of Psychology at the University of Muenster. Participants were seated in front of a computer and instructed to rate the quality of their actions on a scale from 1 ("not well at all") to 6 ("very well") based on their subjective impression. The instructions were kept intentionally liberal as to not influence participants to favor specific aspects of the action for their evaluation. The experiment consisted of two blocks with self-paced responses, both lasting between 20 and 30 min. The same videos were presented in a different order in the second block of the experiment. Videos were pseudorandomized so that not more than three videos in a row showed the same action type (hurdling vs. tap dancing). Overall durations of the test session ranged from 40 to 60 min, depending on the participants' response speed. Two weeks after the test session, participants were presented the same videos once more (in pseudo-randomized order). Twenty-seven videos for both hurdling and tap dancing were chosen per participant and

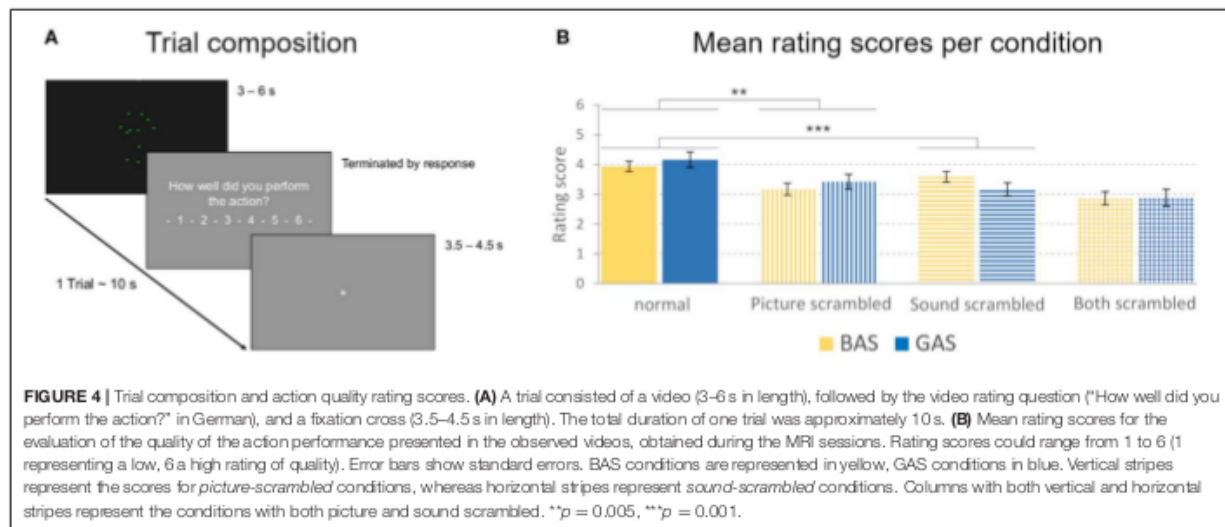


were used in the subsequent fMRI session. The videos with the highest reliability in rating were chosen. Every video was rated a total of four times (two times in the test and two times in the retest sessions). Of all chosen videos (702 videos in total, 54 per participant), 23.79% received the same rating on all four repetitions, in 69.8% ratings varied by a score of either +1 or –1 in one or two of the repetitions, and in 6.41% ratings varied by a score of ± 1 .

fMRI Session

For the fMRI session, participants were instructed to rate the quality of their actions presented in the videos. They were informed that there would be "scrambled" videos, where visual and auditory input would not match, but they should still consider both modalities in the best way possible to rate the quality of their performance. Participants were asked to regulate the volume of the sounds before the experiment started to assure that the action sounds were audible above the scanning noises. The experiment consisted of nine blocks, including 28 trials each (**Figure 4**).

Transition probabilities ensured that every condition was preceded by every condition (including the same condition) in the same number of trials over the whole experiment. The first



trial of a block was a repetition of the last trial of the preceding block, to avoid losing a transition. The remaining 27 trials consisted of 3 trials for each of the nine conditions. With the first trial after each pause discarded, 243 trials remained, 216 video trials, and 27 null events, where a fixation cross was presented (27 trials for each of the nine conditions). The duration of the null events was fixed at 5 s. Before every trial, a fixation cross was presented as an interstimulus interval, varying between 3.5 to 4.5 s in length. After every video trial, the six-point rating scale, including the rating question, was presented. The experiment continued upon the participants' button press.

After the experiment, lasting approximately 45 min, an 8-min resting-state sequence was acquired. Participants were asked to look at a fixation cross for the whole period. Throughout the entire scanning routine, participants were instructed to refrain from moving.

fMRI Recordings and Preprocessing

Participants were scanned in a 3-Tesla Siemens Magnetom Prisma MR tomograph (Siemens, Erlangen, Germany) using a 20-channel head coil. A 3D-multiplanar rapidly acquired gradient-echo (MPRAGE) sequence was used to obtain high resolution T1 weighted images ahead of functional scanning, with scanning parameters set to 192 slices, a repetition time (TR) of 2,130 ms, an echo time (TE) of 2.28 ms, slice thickness of 1 mm, a field of view (FoV) of $256 \times 256 \text{ mm}^2$, and a flip angle of 8° .

Gradient-echo echoplanar imaging (EPI) was used to measure blood-oxygen-level-dependent (BOLD) contrast for functional imaging data of the whole brain. There were 10 EPI sequences in total. One sequence for the volume adjustment and one sequence for each of the nine experimental blocks. Scanning parameters were set to a TE of 30 ms, a TR of 2,000 ms, a flip angle of 90° , 33 slices with a slice thickness of 3 mm, and a FoV of $192 \times 192 \text{ mm}^2$.

Imaging data were processed using SPM12 (Wellcome Trust, London, England). Preprocessing consisted of slice

time correction to the middle slice, realignment to the mean image, co-registration of the functional data to the individual's structural scan, normalization into the standard MNI space (Montreal Neurological Institute, Montreal, QC, Canada) based on segmentation parameters, and spatial smoothing with a Gaussian kernel of full-width at half maximum (FWHM) of 8 mm. A high-pass temporal filter equivalent to 128 s was applied to the data.

Statistical Data Analysis

Behavioral Data Analysis

Firstly, we performed the Kolmogorov-Smirnov test to ensure the normal distribution of our rating scores. A $2 \times 2 \times 2$ within-subject analysis of variance (ANOVA) was calculated to examine differences in the performance ratings between the eight experimental conditions, using SPSS (IBM, New York, United States). The first factor was ACTION, with the factor levels BAS (*hurdling*) and GAS (*tap dancing*), the second factor was PICTURE with factor levels *picture normal* and *picture scrambled* and the third factor was SOUND with factor levels *sound normal* and *sound scrambled*.

We calculated *post-hoc t*-tests for significant main effects using a Bonferroni correction for multiple comparisons, which divides the significance threshold (here we use $\alpha = 0.05$) by the number of tests (Bonferroni, 1936).

fMRI Design Specification

The design was implemented in SPM12, following a general linear model approach (GLM, Friston et al., 1994; Worsley and Friston, 1995). The modeled activation was time-locked to the onsets of the videos or null events. Epochs contained the full presentation period ranging from 3 to 6 s for the videos, and 5 s for the null events. Since tap dancing and hurdling differ with regard to their auditory event density, i.e., the number of distinguishable auditory sounds occurring per second, we controlled for this source of variance by introducing regressors

of nuisance. To this end, we used the MIR Toolbox (Lartillot et al., 2008) to calculate the action sounds' event densities. The GLM consisted of 23 regressors in total: eight regressors for the experimental conditions, eight parametric regressors modeling the event densities for each of the eight experimental conditions, one regressor for the null events, and six regressors for the motion parameters (three translations and three rotations). Activation for 27 trials was considered for the modeling of each of the experimental regressors as well as for the null event regressor. All regressors were convolved with the hemodynamic response function.

On the first level, *t*-contrasts of the experimental conditions against null were calculated (condition > rest). These contrast images were then used to set up a flexible factorial design on the second level. The flexible factorial design was chosen because it accounts best for the within-subject factor. The model consisted of 21 regressors—eight regressors for the experimental conditions, and 13 regressors for the subject effects, one for each participant.

First, *t*-contrasts for the unscrambled conditions were calculated (BAS_normal > GAS_normal) to assess basic differences between tap dancing and hurdling precluding potential effects of visual scrambling on the use of auditory information. Resulting *t*-maps were corrected using false discovery rate (FDR) correction and a threshold of $p < 0.05$. Additionally, we defined *F*-contrasts for the main effect of ACTION, thus including both the normal and the scrambled condition (BAS > GAS), and for the ACTION × SOUND interaction effect. *T*-tests were calculated to examine the direction of effects. Region of interest (ROI) analyses were performed to test our anatomically specified hypotheses, using FDR-correction with a threshold of $p < 0.001$. Structural ROIs were defined using the automated anatomical labeling (AAL) atlas and created using the WFU PickAtlas toolbox (Maldjian et al., 2003) in SPM12. Firstly, we performed a ROI analysis for the primary auditory cortex (Heschl's gyrus) for the BAS > GAS contrast, to test for the hypothesized stronger sensory attenuation for tap dancing than for hurdling. Secondly, ROI analyses for the secondary auditory cortex (pSTG), the SMA, and cerebellum for the GAS > BAS contrast were performed, to investigate a stronger activation for tap dancing due to explicit sound predictions over and above visual predictions. Thirdly, we performed a ROI analysis for the ACTION × SOUND interaction effect, using structural ROIs for the pSTG, to examine the differential involvement of secondary auditory cortices in the sound-scrambled versions of hurdling and tap dancing. We expected more activation in the tap-dancing condition due to a more pronounced mismatch whenever the sound did not fit the perceived action. We additionally extracted beta values from pSTG and the primary auditory cortex, to examine more subtle differences between the conditions.

RESULTS

Behavioral Results

The rating scores for all conditions were normally distributed, as checked by a Kolmogorov-Smirnov test. For the analysis of

the rating scores, a $2 \times 2 \times 2$ within-subject ANOVA with the factors ACTION (BAS, GAS), PICTURE (picture normal, picture scrambled), and SOUND (sound normal, sound scrambled) was calculated. Our final sample size was $n = 13$. We calculated a *post-hoc* power analysis using GPower (Erdfeiler et al., 1996), determining a critical *F*-value (defining the boundary for the rejection of the null hypothesis) of 4.75. All observed effects were well above this critical value, indicating that our findings were reliable even in this relatively small sample size.

First of all, we did not find a main effect for the factor ACTION [$F_{(1, 12)} = 0.009, p = 0.928$], indicating that participants were not biased to rate either their hurdling or tap-dancing performance as superior. Importantly, this balanced rating provides a solid basis to interpret differences between BAS and GAS without a confounding bias by preference.

As hypothesized, there was a significant main effect for the factor SOUND [$F_{(1, 12)} = 22.01, p < 0.001, \eta^2 = 0.647$], driven by lower rating for sound scrambled ($M = 3.13, SD = 0.6$, **Figure 4B**) than compared to sound normal ($M = 3.67, SD = 0.53$). Likewise, a significant main effect for the factor PICTURE [$F_{(1, 12)} = 11.86, p = 0.005, \eta^2 = 0.497$] was explained by higher rating scores in the picture normal ($M = 3.71, SD = 0.53$) vs. picture scrambled condition ($M = 3.09, SD = 0.7$, **Figure 4B**).

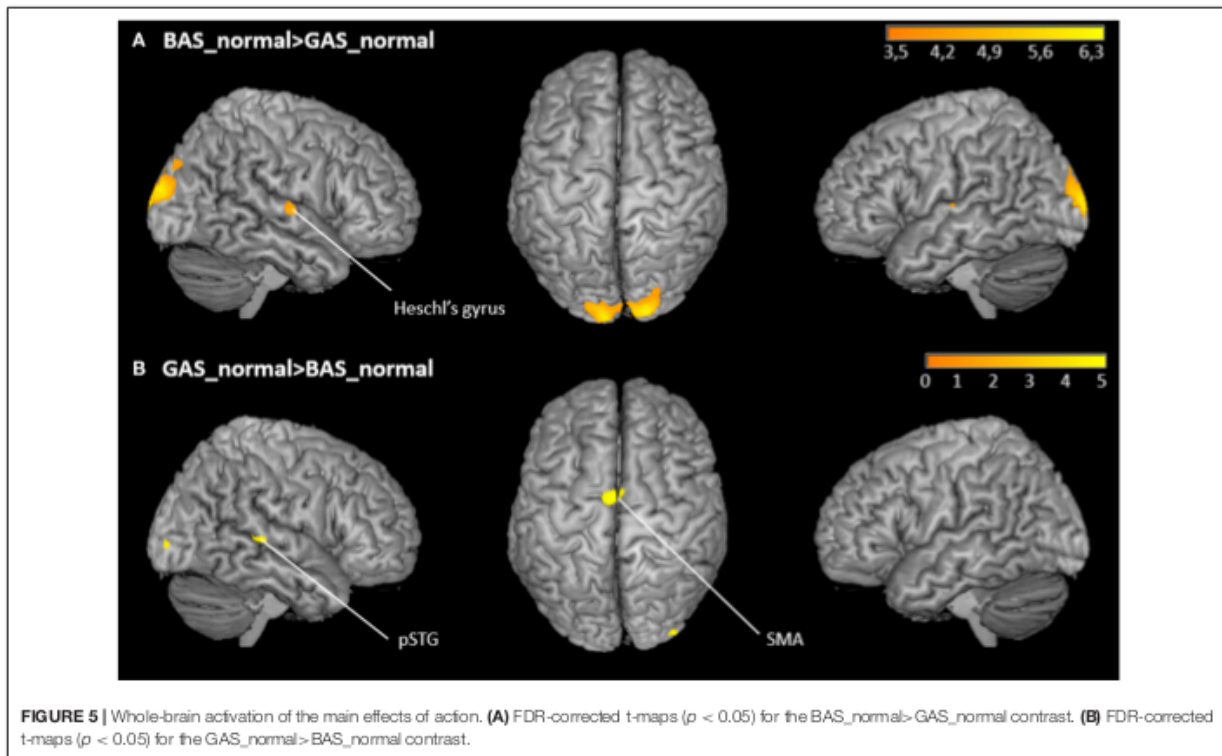
There was a significant ACTION × SOUND interaction [$F_{(1, 12)} = 11.67, p = 0.005, \eta^2 = 0.493$]. Paired *t*-tests revealed lower ratings for sound scrambled vs. normal for hurdling [$M = 3.23, SD = 0.61$ vs. $M = 3.55, SD = 0.49$; $t_{(12)} = 2.91, p = 0.013$] as well as for tap dancing [$M = 3.03, SD = 0.82$ vs. $M = 3.79, SD = 0.81$; $t_{(12)} = 5.08, p < 0.001$]. The three-way interaction between ACTION, PICTURE and SOUND reached significance [$F_{(1, 12)} = 6.66, p = 0.024, \eta^2 = 0.357$]; for *post-hoc* *t*-tests, see **Table S1** and **Table S2**], corroborating that the impact of auditory scrambling was stronger on tap dancing than on hurdling.

We performed a post-experimental survey where we asked participants to rate on a 6-point Likert scale how difficult it was for them to evaluate hurdling and tap-dancing videos. The rating difficulty did not differ between hurdling ($M = 3.77, SD = 1.17$) and tap-dancing videos [$M = 3.54, SD = 1.13, t_{(12)} = 0.507, p = 0.621$].

Finally, as for the fMRI data, we aimed to control for potential confounds of the behavioral rating by stimulus event density. To this end, we tested whether the rating scores correlated with the event density of the corresponding condition (all $p > 0.05$), implementing the same $2 \times 2 \times 2$ within-subject ANOVA for the event density values. Here, we did not find a significant ACTION × SOUND interaction [ACTION × SOUND, $F_{(1, 12)} = 0.34, p = 0.573$].

fMRI results

The whole brain contrast BAS_normal > GAS_normal yielded higher activity in right primary auditory cortex (hypothesis 1) as well as in the occipital pole. Corroborating hypothesis 2, the reverse contrast GAS_normal > BAS_normal showed significant effects in SMA and right pSTG (**Figure 5** and **Table 1**). Note that the same significant effects were observed when contrasting BAS

**TABLE 1 |** Whole-brain activation of the main effects of action.

	x	y	z	t-value	Voxels
BAS_normal>GAS_normal					
Middle occipital gyrus	18	-100	20	6.45	718
Heschl's Gyrus	54	-13	5	4.70	39
	-51	-16	8	4.19	18
GAS_normal>BAS_normal					
SMA	-6	-7	65	5.51	42
pSTG	48	-34	5	4.92	30

Regions of activation, MNI (Montreal Neurological Institute) coordinates (x, y, z), t-values for the local maxima (FDR-corrected, $p < 0.05$), activation extent in voxels (clusters larger than $k = 20$).

and GAS aggregated for all their sub-conditions, i.e., effects were independent of auditory or visual scrambling.

The ROI analysis of the right and left Heschl's gyri yielded the expected lower activation of primary auditory cortex in GAS vs. BAS. According to the hypothesis of stronger sound prediction for the GAS>BAS contrast, the ROI analyses for the SMA, the cerebellum and the pSTG revealed also significant activation increases. ROI results are summarized in **Table 2**.

The ROI analysis for the interaction effect between ACTION and SOUND, testing for a more pronounced prediction error in the auditorily scrambled GAS condition, did not show significant results in pSTG. Beta value extraction was

TABLE 2 | Region of interest (ROI) results.

	x	y	z	t-value	Voxels
BAS > GAS					
Heschl's Gyrus	54	-10	5	7.40	17
	-51	-16	8	5.99	5
GAS > BAS					
SMA	-3	-4	68	10.66	397
pSTG	54	-31	5	11.53	173
	-54	-31	23	6.72	131
	-54	5	-4	6.01	23
Cerebellum	-27	-58	-22	8.4	36
	24	-61	-19	7.95	62
	27	-64	-52	7.76	91
	-24	-64	-52	7.01	37

MNI (Montreal Neurological Institute) coordinates (x, y, z), t-values for the local maxima (FDR-corrected, $p < 0.001$), activation extent in voxels. ROI, region of interest; pSTG, posterior superior temporal gyrus; SMA, supplementary motor area.

performed to follow up on this, yielding a non-significant trend ($p = 0.061$) for the interaction effect (**Figure S2**). Additionally, we also extracted beta values from primary auditory cortex, to determine whether auditory scrambling of GAS would offset attenuation by increasing the prediction error. Indeed, we found a non-significant trend for the

GAS_sound_scrambled>GAS_sound_normal comparison ($p = 0.071$).

DISCUSSION

Most of our actions generate sounds. Intuitively, these action sounds are important for controlling some of our actions, for instance speaking and singing, but possibly much less so for other types of action. The present study employed fMRI and an action quality rating task to investigate potential differences between actions that are executed in order to generate a particular action sound (GAS actions), and actions that cause sounds rather incidentally (BAS actions). Participants were presented with point-light videos showing themselves tap dancing (a GAS action) or hurdling (a BAS action). Following the predictive coding account, we hypothesized that the impact of predicted action sounds would be positively reflected in stronger activity of higher auditory areas (more auditory prediction) and correspondingly more pronounced attenuation in primary auditory cortex (less auditory mismatch). Moreover, trials where we introduced an experimental distortion of action sounds were expected to induce a stronger auditory prediction error in the respective network, and impair the quality rating more effectively, for GAS vs. BAS actions.

To determine the impact of visual and auditory information on the evaluation of one's own performance, we manipulated the visual and the auditory information by scrambling. As expected, both visual scrambling and auditory scrambling led to a significant reduction of the rating scores for hurdling and tap-dancing performance, with scores being the lowest when both modalities were scrambled at the same time. In line with our predictions, auditory scrambling had a stronger impact on the rating of tap-dancing than hurdling performance, with rating scores decreasing in both actions, but to a stronger degree in tap dancing. This finding supports the particular relevance of action sounds as an error-monitoring tool in GAS actions (Murgia et al., 2017). While auditory scrambling had an effect on the rating of the hurdling trials as well, the effect was more pronounced in tap dancing, showing that, with the auditory output being an explicit action goal, perception of action quality was especially reduced by incoherent auditory feedback. It is however important to note that action sounds were also important for the rating of hurdling performance, although the created sounds were not explicitly intended. Together, behavioral findings suggest that the brain generates predictions of how our actions should sound like, both in case of tap-dancing (GAS actions) as well as hurdling (BAS actions). Subtle differences in the level of interference, however, point to a more prominent role of auditory expectations in the former.

In agreement with our hypotheses, we found stronger activation in Heschl's gyri for the hurdling (BAS) compared to tap dancing (GAS) trials, reflecting a more pronounced sensory attenuation when auditory action consequences are predominantly used in the predictive model (GAS). Sensory attenuation to self-initiated sounds is based on an existing association between the initiated movement and the resulting

sound (Ticini et al., 2012; Keysers and Gazzola, 2014). Self-produced sounds elicit a smaller amplitude in early EEG or MEG components, presumably due to the feeling of self-agency (Aliu et al., 2009; Baess et al., 2011; Timm et al., 2014). The precise origin of sensory attenuation of self-produced sounds has not yet been completely unraveled (Hughes et al., 2013; Horváth, 2015). A more recent study shows that sensory attenuation is not solely due to the self-generation of action sounds, but relies on the predictability of sensory input (Kaiser and Schütz-Bosbach, 2018). Self-generated sounds have high predictability, as the sensory effects are part of our motor plan when initiating and performing a movement (Shin et al., 2010). Accordingly, we suppose that tap-dancing sounds, being an intentional part of our motor plan, are more efficiently attenuated by internal predictive models, as reflected by stronger pSTG and SMA activity. In contrast, hurdling sounds may be less relevant in the predictive model, and are therefore not attenuated to the same extent in primary auditory cortices as the tap-dancing sounds.

Notably, attention was found to reverse attenuation effects by predicted stimuli, leading to enhanced rather than attenuated responses (Reznik et al., 2015; Schröger et al., 2015a,b; Wollman and Morillon, 2018). If at all, we would have expected attention to be increased for GAS vs. BAS actions. To the contrary, primary auditory cortex was attenuated in GAS compared to BAS, clearly favoring the prediction-caused attenuation over the attention-caused enhancement explanation of our findings. Corroborating this interpretation further, sound scrambling in GAS videos caused an increase of primary auditory cortex activity (non-significant trend), as would be expected for a prediction error rather than for a down-regulation of attention to the scrambled signal.

As expected, the SMA, the pSTG and the cerebellum were more active for tap-dancing than for hurdling. These and adjacent areas have been found and discussed in connection with action sound processing more generally (Herrington et al., 2011; Bischoff et al., 2014; Woods et al., 2014; Reznik et al., 2015). Reznik et al. (2015) proposed that predictive information is sent from the SMA or primary motor cortex to auditory cortices to modify activation during active sound generation. Although our participants did not actively create sounds in the scanner, they perceived sounds they actively created in the past, and perception of own past actions is thought to adequately represent the brain activity during action execution (e.g., Sato, 2008; Wutte et al., 2012). The same areas as found by Reznik et al. (2015) were more active in our tap-dancing condition compared to hurdling, indicating a similar predictive information update when action sounds are part of the intended action goal (GAS vs. BAS). This matches the stronger—less attenuated—effect in primary auditory cortex for hurdling trials as well.

While BOLD contrasts did not confirm the differential effect of auditory scrambling on tap-dancing and hurdling, beta estimates extracted from the pSTG did indicate a descriptive interaction effect for the imaging data. Thus, the beta weights for hurdling did not differ between the *sound normal* and *sound scrambled* condition, whereas there was a small difference between these two conditions for tap dancing. Both these findings and the behavioral results speak in favor of our interaction

hypothesis: While auditory scrambling has an effect on both hurdling and tap dancing, the effect on tap dancing is larger, indicating a greater relevance of the auditory domain to positively evaluate action quality and a stronger predictive mismatch in GAS actions.

The overall stronger activation in occipital visual areas for hurdling compared to tap-dancing was not hypothesized. Obviously, evaluating one's hurdling performance yielded a more extensive visual processing of the observed action, as indicated by increased BOLD activity in occipito-temporal cortices (Jastorff et al., 2010). This also matches the behavioral finding of a stronger impact of visual scrambling on the rating of hurdling quality.

Overall, our behavioral and fMRI findings speak in favor of a higher relevance of action sounds in tap dancing as compared to hurdling. These actions may be representative for two subclasses of sound-producing actions, but their distinction might reflect two manifestations on a continuum rather than a strict dichotomy. As we observed, the auditory scrambling reduced rating scores in hurdling trials as well, indicating that the auditory domain is not completely unnecessary when evaluating these actions. This aligns with several previous findings regarding the relevance of sound when performing and improving sport related actions (for a review, see Schaffert et al., 2019). Also, effects of deprivation or alteration of auditory feedback on musical performance is not completely consistent, with some studies showing no effect of at least deprived feedback (Gates et al., 1974; Finney, 1997). A continuum, reaching from language production, where the auditory output is inarguably important, to simple everyday actions producing sounds, like placing a glass on a table, seems plausible. Our chosen actions might be somewhere in between, with tap dancing being closer to language and music, and hurdling closer to simple everyday action sounds. Note that this difference is particularly remarkable given that hurdling and tap-dancing are both whole-body actions and produce sounds by feet-floor contact, ruling out confounding impact on the motor side.

Future studies should avoid differences in event density as a potential source of confounding variance. While we controlled for this factor in both the fMRI and the behavioral analysis, event density could have been limited right from the beginning by choosing a tap-dancing sequence that largely matches the rhythm generated by the hurdling movement.

A limitation of our study is the small number of participants ($n = 13$), which resulted from the large extent of the investigation, including a 9-week training, several filming sessions, as well as multiple experimental sessions, both behavioral and fMRI. However, other studies examining action sound had a comparably small sample size (Menzer et al., 2010; Reznik et al., 2015). Considering that we found both robust behavioral interaction effects and FDR-corrected imaging results despite the limited number of participants, speaks in favor of a further pursue of our hypotheses with a larger sample size. An interesting approach might be to use the same stimuli with naïve participants who have not trained hurdling and tap dancing before, generalizing the results to participants who are no experts of the performed actions.

To further clarify the role of action sounds for monitoring action performance, future studies may examine the effects of deprivation and interference on the entire spectrum of sound-producing actions, as has been done for language production, musical performance, and some other actions (Howell, 2004; Keough and Jones, 2009; Pfordresher and Beasley, 2014; Kennel et al., 2015). Our study is a first step into a more systematic approach to understanding action sounds, while establishing ideas for additional research to deepen the comprehension of this relevant topic. Both musicians and athletes might benefit from a better understanding of the role of action sounds for optimizing action performance, giving them the opportunity to adequately train their skills. Research in schizophrenia might also gain from a better insight into the connections between action sounds and motor control, especially regarding the sense of agency, and the failure to attribute self-produced sensations to oneself in people with schizophrenia.

CONCLUSION

In conclusion, our study provides interesting new insights on action sounds and their relevance for evaluating executed actions. In contrast to other studies, we trained our participants in two sound-producing actions and showed them their own actions during an fMRI experiment. This is, to our knowledge, completely novel in this field of research and thus provides a unique view on how our own action sounds are processed in the brain, depending on whether sound is an intentional action goal (tap dancing) or is generated incidentally (hurdling). Our results indicate that in the former case, the brain intensifies auditory predictions, and is more surprised in case of unexpected action sounds; moreover, these are particularly harmful to quality rating on a behavioral level. Research on real-life and whole-body action sounds is still relatively sparse, although they are omnipresent in our everyday life and supposedly important for controlling, understanding, and improving at least some of our actions. Finding that goal-relevance on a subjective level modulates brain processes during sound appraisal points out that this field of research is worth further exploration.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee Department of Psychology University of Muenster. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NH, JP, DK, and RS contributed conception and design of the study. NH performed the statistical analysis and wrote the first draft of the manuscript. IT and RS wrote sections of the

manuscript. KZ, MR, and RS contributed with scientific support, supervision, and coordination. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

The project was funded by the Deutsche Forschungsgemeinschaft (SCHU1439-4-2).

ACKNOWLEDGMENTS

We would like to thank Monika Mertens for her help during data collection. Furthermore, we thank Theresa Eckes,

Alina Eisele, Marie Kleinbielen, Katharina Thiel, and Stefan Vinbrüx for their assistance during filming and with creating the stimulus material. Finally, we would like to thank Nadiya El-Sourani, Amelie Huebner, Klara Hagelweide, Laura Quante, Marlen Roehle, and Lena Schliephake for rewarding discussions, and our three reviewers for insightful input on this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2020.00483/full#supplementary-material>

REFERENCES

- Aliu, S. O., Houde, J. F., and Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *J. Cogn. Neurosci.* 21, 791–802. doi: 10.1162/jocn.2009.21055
- Baess, P., Horváth, J., Jacobsen, T., and Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: an ERP study. *Psychophysiology* 48, 1276–1283. doi: 10.1111/j.1469-8986.2011.01196.x
- Bischoff, M., Zentgraf, K., Pilgamm, S., Stark, R., Krüger, B., and Munzert, J. (2014). Anticipating action effects recruits audiovisual movement representations in the ventral premotor cortex. *Brain Cogn.* 92, 39–47. doi: 10.1016/j.bandc.2014.09.010
- Bonferroni, C. E. (1936). *Teoria Statistica Delle Classi e Calcolo Delle Probabilità*, Vol. 8. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 3–62.
- Erdfelder, E., Faul, F., and Buchner, A. (1996). GPOWER: a general power analysis program. *Behav. Res. Methods* 28, 1–11. doi: 10.3758/BF03203630
- Finney, S. A. (1997). Auditory feedback and musical keyboard performance. *Music Percept.* 15, 153–174. doi: 10.2307/40285747
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Friston, K. (2012). Prediction, perception and agency. *Int. J. Psychophysiol.* 83, 248–252. doi: 10.1016/j.ijpsycho.2011.11.014
- Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biol. Cybern.* 102, 227–260. doi: 10.1007/s00422-010-0364-z
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210. doi: 10.1002/hbm.460020402
- Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro, E., Andrew, C. M., et al. (2006). An fMRI study of verbal self-monitoring: neural correlates of auditory verbal feedback. *Cereb. Cortex* 16, 969–977. doi: 10.1093/cercor/bhj039
- Gates, A., Bradshaw, J. L., and Nettleton, N. C. (1974). Effect of different delayed auditory feedback intervals on a music performance task. *Percept. Psychophys.* 15, 21–25. doi: 10.3758/BF03205822
- Heilbron, M., and Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex? *Neuroscience* 389, 54–73. doi: 10.1016/j.neuroscience.2017.07.061
- Herrington, J. D., Nymberg, C., and Schultz, R. T. (2011). Biological motion task performance predicts superior temporal sulcus activity. *Brain Cogn.* 77, 372–381. doi: 10.1016/j.bandc.2011.09.001
- Horváth, J. (2015). Action-related auditory ERP attenuation: paradigms and hypotheses. *Brain Res.* 1626, 54–65. doi: 10.1016/j.brainres.2015.03.038
- Howell, P. (2004). Effects of delayed auditory feedback and frequency-shifted feedback on speech control and some potentials for future development of prosthetic aids for stammering. *Stammering Res.* 1, 31–46.
- Hughes, G., Desantis, A., and Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychol. Bull.* 139, 133–151. doi: 10.1037/a0028566
- Jastorff, J., Begliomini, C., Rizzolatti, G., Orban, G. A., Fabbri-Destro, M., and Jastorff, J. (2010). Coding observed motor acts: different organizational principles in the parietal and premotor cortex of humans. *J. Neurophysiol.* 104, 128–140. doi: 10.1152/jn.00254.2010
- Jo, H. G., Habel, U., and Schmidt, S. (2019). Role of the supplementary motor area in auditory sensory attenuation. *Brain Struct. Funct.* 224, 2577–2586. doi: 10.1007/s00429-019-01920-x
- Jones, J. A., and Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Exp. Brain Res.* 190, 279–287. doi: 10.1007/s00221-008-1473-y
- Kaiser, J., and Schütz-Bosbach, S. (2018). Sensory attenuation of self-produced signals does not rely on self-specific motor predictions. *Eur. J. Neurosci.* 47, 1303–1310. doi: 10.1111/ejn.13931
- Kennel, C., Strees, L., Pizzera, A., Justen, C., Hohmann, T., and Raab, M. (2015). Auditory reafferences: The influence of real-time feedback on movement control. *Front. Psychol.* 6:69. doi: 10.3389/fpsyg.2015.00069
- Keough, D., and Jones, J. A. (2009). The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing. *J. Acoust. Soc. Am.* 126, 837–846. doi: 10.1121/1.3158600
- Keysers, C., and Gazzola, V. (2014). Hebbian learning and predictive mirror neurons for actions, sensations and emotions. *Philos. Trans. R. Soc. B Biol. Sci.* 369. doi: 10.1098/rstb.2013.0175
- Lartillot, O., Toivainen, P., and Eerola, T. (2008). “A matlab toolbox for music information retrieval,” in *Data Analysis, Machine Learning and Applications*, eds C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker (Berlin; Heidelberg: Springer), 261–268. doi: 10.1007/978-3-540-78246-9_31
- Maldjian, J. A., Laurienti, P. J., Kraft, R. A., and Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19, 1233–1239. doi: 10.1016/S1053-8119(03)00169-1
- Menzer, F., Brooks, A., Halje, P., Faller, C., Vetterli, M., and Blanke, O. (2010). Feeling in control of your footsteps: conscious gait monitoring and the auditory consequences of footsteps. *Cogn. Neurosci.* 1, 184–192. doi: 10.1080/17588921003743581
- Miall, R. C., and Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Netw.* 9, 1265–1279. doi: 10.1016/S0893-6080(96)00035-4
- Mifsud, N. G., Oestreich, L. K. L., Jack, B. N., Ford, J. M., Roach, B. J., Mathalon, D. H., et al. (2016). Self-initiated actions result in suppressed auditory but amplified visual evoked components in healthy participants. *Psychophysiology* 53, 723–732. doi: 10.1111/psyp.12605
- Murgia, M., Prpic, V., Jenny, O., McCullagh, P., Santoro, I., Galmonte, A., et al. (2017). Modality and perceptual-motor experience influence the detection of temporal deviations in tap dance sequences. *Front. Psychol.* 8:1340. doi: 10.3389/fpsyg.2017.01340

- Oldfield, R. C. (1971). The assessment and analysis of handedness. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Petrini, K., Pollick, F. E., Dahl, S., McAleer, P., McKay, L., Rocchesso, D., et al. (2011). Action expertise reduces brain activity for audiovisual matching actions: an fMRI study with expert drummers. *Neuroimage* 56, 1480–1492. doi: 10.1016/j.neuroimage.2011.03.009
- Pfordresher, P. Q., and Beasley, R. T. E. (2014). Making and monitoring errors based on altered auditory feedback. *Front. Psychol.* 5:914. doi: 10.3389/fpsyg.2014.00914
- Phillips, H. N., Blenkmann, A., Hughes, L. E., Bekinschtein, T. A., and Rowe, J. B. (2015). Hierarchical organization of frontotemporal networks for the prediction of stimuli across multiple dimensions. *J. Neurosci.* 35, 9255–9264. doi: 10.1523/JNEUROSCI.5095-14.2015
- Pizzera, A., and Hohmann, T. (2015). Acoustic information during motor control and action perception: a review. *Open Psychol. J.* 8, 183–191. doi: 10.2174/1874350101508010183
- Rauschecker, J. P. (2012). "Processing streams in the auditory cortex", in *Neural Correlates of Auditory Cognition. Springer Handbook of Auditory Research*, eds Y. Cohen, R. R. Fay, and A. N. Popper (New York, NY: Springer), 7–43.
- Reznik, D., Ossmy, O., and Mukamel, R. (2015). Enhanced auditory evoked activity to self-generated sounds is mediated by primary and supplementary motor cortices. *J. Neurosci.* 35, 2173–2180. doi: 10.1523/JNEUROSCI.3723-14.2015
- Rummell, B. P., Klee, J. L., and Sigurdsson, T. (2016). Attenuation of responses to self-generated sounds in auditory cortical neurons. *J. Neurosci.* 36, 12010–12026. doi: 10.1523/JNEUROSCI.1564-16.2016
- Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. *Conscious Cogn.* 17, 1219–1227. doi: 10.1016/j.concog.2008.01.003
- Schaffert, N., Janzen, T. B., Mattes, K., and Thaut, M. H. (2019). A review on the relationship between sound and movement in sports and rehabilitation. *Front. Psychol.* 10:244. doi: 10.3389/fpsyg.2019.00244
- Schröger, E., Kotz, S. A., SanMiguel, I. (2015a). Bridging prediction and attention in current research on perception and action. *Brain Res.* 1626, 1–13. doi: 10.1016/j.brainres.2015.08.037
- Schröger, E., Marzecová, A., and Sanmiguel, I. (2015b). Attention and prediction in human audition: a lesson from cognitive psychophysiology. *Eur. J. Neurosci.* 41, 641–664. doi: 10.1111/ejn.12816
- Shin, Y. K., Proctor, R. W., and Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychol. Bull.* 136, 943–974. doi: 10.1037/a0020541
- Ticini, L. F., Schütz-Bosbach, S., Weiss, C., Casile, A., and Waszak, F. (2012). When sounds become actions: Higher-order representation of newly learned action sounds in the human motor system. *J. Cogn. Neurosci.* 24, 464–474. doi: 10.1162/jocn_a_00134
- Timm, J., Schönwiesner, M., SanMiguel, I., and Schröger, E. (2014). Sensation of agency and perception of temporal order. *Conscious. Cogn.* 23, 42–52. doi: 10.1016/j.concog.2013.11.002
- Timm, J., Schönwiesner, M., Schröger, E., and SanMiguel, I. (2016). Sensory suppression of brain responses to self-generated sounds is observed with and without the perception of agency. *Cortex* 80, 5–20. doi: 10.1016/j.cortex.2016.03.018
- Tourville, J. A., Reilly, K. J., and Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39, 1429–1443. doi: 10.1016/j.neuroimage.2007.09.054
- van Elk, M., Salomon, R., Kannape, O., and Blanke, O. (2014). Suppression of the NI auditory evoked potential for sounds generated by the upper and lower limbs. *Biol. Psychol.* 102, 108–117. doi: 10.1016/j.biopsycho.2014.06.007
- Waszak, F., Cardoso-Leite, P., and Hughes, G. (2012). Action effect anticipation: neurophysiological basis and functional consequences. *Neurosci. Biobehav. Rev.* 36, 943–959. doi: 10.1016/j.neubiorev.2011.11.004
- Wollman, I., and Morillon, B. (2018). Organizational principles of multidimensional predictions in human auditory attention. *Sci. Rep.* 8:13466. doi: 10.1038/s41598-018-31878-5
- Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882. doi: 10.1126/science.7569931
- Woods, E. A., Hernandez, A. E., Wagner, V. E., and Beilock, S. L. (2014). Expert athletes activate somatosensory and motor planning regions of the brain when passively listening to familiar sports sounds. *Brain Cogn.* 87, 122–133. doi: 10.1016/j.bandc.2014.03.007
- Worsley, K. J., and Friston, K. J. (1995). Analysis of fMRI time-series revisited — again. *Neuroimage* 2, 173–181. doi: 10.1006/nimg.1995.1023
- Wutte, M. G., Glasauer, S., Jahn, K., and Flanagan, V. L. (2012). Moving and being moved: differences in cerebral activation during recollection of whole-body motion. *Behav. Brain Res.* 227, 21–29. doi: 10.1016/j.bbr.2011.09.042

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Heins, Pomp, Kluger, Trempler, Zentgraf, Raab and Schubotz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Fig. S1. Matching of frequency spectra. After intensity normalization, we used Ozone 5's frequency matching function to capture frequency distributions of hurdling (yellow curve) and tap-dancing recordings (purple). Using hurdling as a reference, the difference curve (hurdling – tap-dancing, red) was computed to adjust the frequency distribution of tap-dancing to match those of hurdling recordings. The amount of frequency matching (see 'Amount' slider, bottom left) was set to make sure that no clipping would occur after applying the equalization. Note that the smoothing parameter was set to zero, ensuring a close approximation of both distributions.

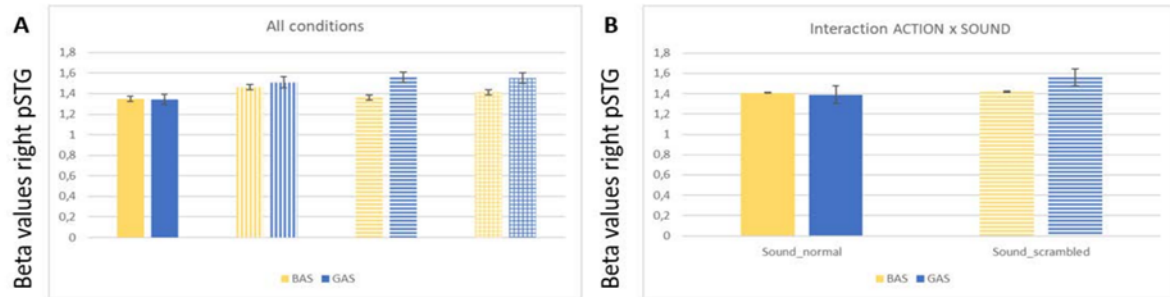


Fig. S2. Beta weights within the superior temporal gyrus. Mean beta weights extracted from the right posterior superior temporal gyrus (A) for all eight experimental conditions and (B) divided by the factors ACTION and SOUND to represent the interaction effect. Error bars represent the standard error of the mean. BAS conditions are represented in yellow, GAS conditions in blue. Vertical stripes represent the picture-scrambled conditions, whereas horizontal stripes represent sound-scrambled conditions. Columns with both vertical and horizontal stripes represent the conditions with both picture and sound scrambled.

Table S1 – Post-hoc t-tests for the two-way interactions

Interaction effect ACTION x PICTURE				
	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
BAS picture normal	3.76	0.56	3.59	.004*
BAS picture scrambled	3.02	0.71		
GAS picture normal	3.67	0.75	2.97	.012*
GAS picture scrambled	3.15	0.90		
BAS picture normal	3.76	0.56	0.43	.676
GAS picture normal	3.67	0.75		
BAS picture scrambled	3.02	0.71	0.59	.564
GAS picture scrambled	3.15	0.90		
Interaction effect ACTION x SOUND				
	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
BAS sound normal	3.55	0.49	2.91	.013*
BAS sound scrambled	3.23	0.61		
GAS sound normal	3.79	0.81	5.08	< .001*
GAS sound scrambled	3.03	0.82		
BAS sound normal	3.55	0.49	1.06	.308
GAS sound normal	3.79	0.81		
BAS sound scrambled	3.23	0.61	0.89	.392
GAS sound scrambled	3.03	0.82		

BAS = by-product action sounds, GAS = goal-related action sounds, *M* = mean, *SD* = standard deviation, significant differences are marked with asterisks (*). Values are Bonferroni-corrected for multiple comparisons.

Table S2 – Post-hoc t-tests for the three-way interactions

Interaction effect ACTION x PICTURE x SOUND			<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
BAS	Picture normal	Sound normal	3.93	0.6	2.57	.024*
		Sound scrambled	3.58	0.61		
	Picture scrambled	Sound normal	3.17	0.70	2.54	.026*
		Sound scrambled	2.87	0.78		
GAS	Picture normal	Sound normal	4.16	0.90	5.01	< .001*
		Sound scrambled	3.17	0.75		
	Picture scrambled	Sound normal	3.42	0.87	4.38	.001*
		Sound scrambled	2.88	0.99		
			<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
Picture normal	Sound normal	BAS	3.93	0.60	-0.91	.382
		GAS	4.16	0.90		
	Sound scrambled	BAS	3.58	0.61	1.92	.080
		GAS	3.17	0.75		
Picture scrambled	Sound normal	BAS	3.17	0.70	-1.14	.278
		GAS	3.42	0.87		
	Sound scrambled	BAS	2.87	0.78	-0.07	0.946
		GAS	2.88	0.99		
			<i>M</i>	<i>SD</i>	<i>t</i>	<i>P</i>
BAS	Sound normal	Picture normal	3.93	0.6	3.16	.008*
		Picture scrambled	3.17	0.70		
	Sound scrambled	Picture normal	3.58	0.61	3.86	.002*
		Picture scrambled	2.87	0.78		
GAS	Sound normal	Picture normal	4.16	0.90	3.74	.003*
		Picture scrambled	3.42	0.87		

Sound scrambled	Picture				
	normal	3.17	0.75	1.66	.123
	scrambled	2.88	0.99		

BAS = by-product action sounds, GAS = goal-related action sounds, M = mean, SD = standard deviation, significant differences are marked with asterisks (*). Values are Bonferroni-corrected for multiple comparisons.

3.2 Study II: Where is my sound? Omission of one's own intentional and incidental action sounds and its effect on BOLD fMRI

Running title: Omission of incidental vs. intentional action sounds

Nina Heins, Jennifer Pomp, Karen Zentgraf,
Markus Raab, & Ricarda I. Schubotz (2020)

Manuscript submitted for publication

Where is my sound? Omission of one's own intentional and incidental action sounds and its effect on BOLD fMRI

Nina Heins^{1,2}, Jennifer Pomp^{1,2}, Karen Zentgraf³, Markus Raab^{4,5}, Ricarda I. Schubotz^{1,2}

¹Department of Psychology, University of Münster, Germany

²Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Germany

³Institute for Sport Sciences, Goethe University Frankfurt, Germany

⁴Department of Performance Psychology, German Sport University Cologne, Germany

⁵School of Applied Sciences, London South Bank University, United Kingdom

Abstract

Most of our actions produce sounds, but some of them appear more relevant to achieve our goals than others. To test this assumption, we examined the cerebral and behavioral impact of sound omission in self-performed real-life actions, underlying either incidentally or intentionally generated sounds. To this end, we trained participants in tap dancing and hurdling. Participants were presented point-light videos of their own actions in a subsequent functional Magnetic Resonance Imaging (fMRI) study while they engaged in a performance quality rating. Behavioral results suggested a higher relevance of auditory information for tap dancing, and of visual information for hurdling. BOLD activity of the supplementary motor area (SMA) co-varied positively with rating scores for silenced tap dancing but not for silenced hurdling. This indicates a retrieval of predictive information for the evaluation of actions producing sounds intentionally, while actions producing sounds incidentally can be evaluated on visual information alone without retreating to higher-order predictive input. The present study is the first to examine the neural effects of omission of real-life action sounds and provides new insights into natural action sounds and their relevance for action performance evaluation.

1 Introduction

Sounds produced by our own actions are omnipresent in everyday life. If our action sounds were suddenly absent, we would surely be surprised. According to the predictive coding framework, the brain works as a predictive device and is tuned to minimize surprise (Friston, 2005). Masking or manipulating our action sounds would therefore lead to surprise not only on a subjective level but also in our brains (Garrido, Kilner, Stephan, & Friston, 2009; Wacongne et al., 2011). In this framework, it is assumed that predictive models are formed in higher cortical areas and then conveyed to the respective sensory cortices. Here, expected and actual sensory input is compared, and differences manifest in form of prediction errors or mismatch responses. In the auditory domain, the so-called mismatch negativity (MMN) of the EEG is considered a prediction error correlate (Näätänen, Paavilainen, Rinne, & Alho, 2007). Its magnetoencephalographic counterpart (MMNm) and combined M/EEG/fMRI studies located its source in the primary and secondary auditory cortex (Liebenthal et al., 2003; Mathiak et al., 2002; Dürschmid et al., 2016; Recasens, Grimm, Capilla, Nowak, & Escera, 2014). In the same regions, an error response does not only manifest when perceived sounds differ from our predictions but also when sounds are unexpectedly omitted (Nazimek, Hunter, Hoskin, Wilkinson, & Woodruff, 2013; Recasens & Uhlhaas, 2017).

The occurrence of prediction errors whenever action sounds are omitted or altered underlines that action sounds are an integral part of our predictive models when executing and perceiving actions (Carbajal & Malmierca, 2018; Friston, 2012; Heilbron & Chait, 2018). The same predictive processes are suggested to cause the reduced neural response to self-generated auditory stimuli, i.e. sensory attenuation (Baess, Horváth, Jacobsen, & Schröger, 2011; Timm, Schönwiesner, Schröger, & SanMiguel, 2016). While the precise origin of sensory attenuation has yet to be revealed (Horváth, 2015; Hughes, Desantis, & Waszak, 2013), a more recent study suggests the predictability of self-generated stimuli to be the cause of attenuated processing (Kaiser & Schütz-Bosbach, 2018).

The importance of sound for action control is particularly evident in language and musical performance, that is, in actions that produce sounds which we referred to as GAS (Goal-related Action Sounds) in a previous fMRI study (Heins et al., 2020). For these actions, the auditory action effect is the immediate intended goal of the respective action, from both the subjective perspective of the individual and the neural perspective of our brains (Keysers & Gazzola, 2014). Therefore, unexpected deprivation or alteration of these action sounds should definitely surprise us and possibly disrupt our usual action performance. Indeed, the influence of altered or deprived action sounds has been intensively investigated for speech (Howell, 2004;

Kalinowski, Armson, Roland-Mieszkowski, Stuart, & Gracco, 1993; Lincoln, Packman, & Onslow, 2006). Auditory feedback deprivation causes disruption in musical performance (Pfordresher & Beasley, 2014) and changes its perceived quality (Repp, 1999). Tourville, Reilly, and Guenther (2008) found that altered auditory feedback during speech production yielded compensatory motor adjustments. Similar motor compensations were found when auditory feedback was shifted in frequency during singing (Jones & Keough, 2008; Keough & Jones, 2009). Compensatory attempts were accompanied by higher activity in superior temporal cortices, where expected and actual auditory feedback are compared (Keysers & Gazzola, 2014; Tourville et al., 2008).

In contrast, there are other types of actions which create sounds as a mere by-product, like the sound produced by placing a cup back on a table or the sound of our own footsteps on the ground. We hereafter refer to this type of action sounds as BAS (By-product Action Sounds). For BAS actions, the subjective action goal does not entail an action sound. Still, the omission or distortion of such action sounds would clearly irritate the acting person and might also cause a surprise or mismatch response on the neural level. Studies examining BAS actions are sparse and mostly restricted to abstract sound-producing actions (e.g., button presses) rather than addressing real-life actions naturally producing sounds (for a review, see Hughes, Desantis, & Waszak, 2013). Interestingly, the few existing studies on real-life BAS actions showed results differing from those in speech and musical (i.e., GAS) performance. Kennel et al. (2015) indeed found that masking of online auditory feedback during hurdling (i.e., a BAS action) did not affect performance. This finding suggests that BAS are not as essential for evaluating our action performance and not as informative as GAS for adjusting the underlying predictive model. In line with this assumption, in a previous fMRI study where we presented participants with videos of their own hurdling (BAS) and tap-dancing (GAS) performance (Heins et al., 2020), we found stronger attenuation in the primary auditory cortex (A1), and correspondingly higher activity from upstream areas pSTG and SMA for GAS than for BAS actions. Both pSTG and SMA have been suggested to provide predictive top-down input to A1 in case of unexpectedly omitted sounds (Jo, Habel, & Schmidt, 2019). Specifically, pSTG shows higher activity when auditory predictions are violated (Fu et al., 2006; Keysers & Gazzola, 2014; Yamamoto et al., 2019). On a yet higher level in the predictive hierarchy, SMA rather activates action sound related motor activations and sensorimotor associations, and is not solely affected by unexpected or missing auditory input (Lima, Krishnan, & Scott, 2016). Thus, the findings of our previous fMRI study point to a higher relevance of action sounds as part of

a predictive model in GAS compared to BAS actions, which should furthermore be reflected in a greater surprise response to altered or omitted GAS vs. BAS.

Building on these findings, we here examined the impact of sound omission on both the neural processing and the performance evaluation of GAS and BAS actions. To this end, we presented participants with point-light videos of their own hurdling and tap-dancing actions with and without sound. Participants were asked to rate the quality of their own hurdling and tap-dancing performance on a trial-by-trial basis. Behaviorally, we expected that sound omission reduces rating scores to a higher degree in GAS actions than in BAS actions (Hypothesis 1).

With regard to fMRI effects, we expected SMA and pSTG to be more active in GAS vs. BAS actions reflecting a top-down predictive signal to A1, which should consequently be more attenuated in GAS than in BAS actions (Hypothesis 2, replication of (Heins et al., 2020)). Sound omission was expected to additionally lower activity in auditory cortex (Hypothesis 3) but differently for BAS vs. GAS actions. In case of GAS but not BAS actions, we expected increased BOLD response in pSTG reflecting a prediction error signal leading to an update of the model of the expected auditory pattern (Hypothesis 4).

Finally, judging mute GAS actions but not mute BAS actions was expected to resort to higher, multimodal levels of the predictive hierarchy, e.g. SMA, where motor activation and their corresponding action sounds are stored together. In particular, SMA activity should positively covary with the rating scores of mute GAS but not mute BAS actions (Hypothesis 5). To further ensure validity of this interpretation, we employed visually scrambled mute videos as additional control condition; here, the systematic relationship between SMA activity and rating scores was expected to vanish due to an ineptitude to retrieve the action sound representation from an inadequate visual presentation (Hypothesis 6).

2 Methods

2.1 Participants

Nineteen participants started the hurdling and tap-dancing training sessions. One participant dropped out during the nine-week training period, and five further participants after the training during the time in which we prepared the individual stimulus material (see limitations). Hence, 13 participants completed this fMRI session. One participant was excluded from the final analysis, because their reaction times recorded during the fMRI session diverged more than two standard deviations from the mean reaction time on a group level, leaving 12

participants (8 females, 4 males) for the analysis. This is a relatively small sample size, but we found robust results in our preceding fMRI study using the same sample. The participants' age ranged from 19 to 28 years ($M = 22.1$, $SD = 3.0$), and all of them were right-handed, as assessed by the Edinburgh Handedness Inventory (EHI; Oldfield, 1971), with scores varying from +60 to +100, with a mean of +85. All participants reported to have no history of psychiatric or neurological disorders. They signed an informed consent. After successful participation, participants were rewarded with both course credit and monetarily. The study was approved by the Local Ethics Committee in accordance with the Declaration of Helsinki.

2.2 Material

The stimuli used in the study consisted of point-light displays of hurdling and tap-dancing actions with their corresponding sounds, or mute versions, recorded from each individual participant at different stages throughout training. In the fMRI session, we presented participants with videos of their own actions only, meaning that the stimulus material had to be generated and prepared for each participant on an individual level.

Point-light displays were recorded using the Qualisys Motion Capture System (<https://www.qualisys.com>) with nine cameras (see Fig. 1), while the sound was recorded by in-ear microphones (Sound-man OKM Classic II) for hurdling and by a sound recording app on a mobile phone for tap dancing.

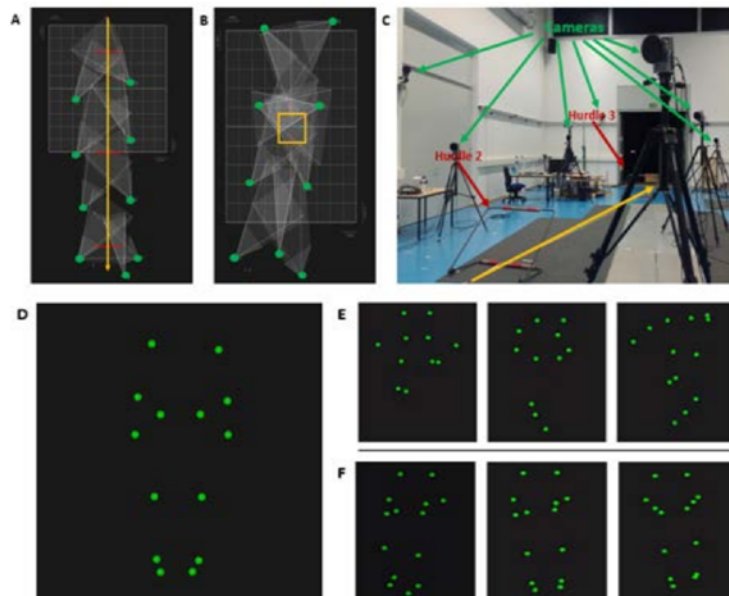


Fig. 1. Camera positions, set-up during the point light recordings and position of the point-light markers. (A) Camera positions during hurdling from a top view perspective. Green dots represent the cameras, red lines the hurdles, and the yellow arrow the hurdling track. (B) Camera positions during tap dancing from a top view perspective. Green dots represent the cameras, the yellow square the area in which the tap dancer performed the sequence (C) Set-up during the recording of hurdling. Three hurdle transitions had to be performed during the recording. The two last hurdles are visible in the figure above. The yellow arrow indicates the hurdling track. (D) Twelve point-light markers were used, positioned at the shoulders, elbows, wrists, hips, knees, ankles, and the tips of the toes. Videos can be found in the supplementary material. (E) Three snapshots of the hurdling action over the course of one video. (F) Three snapshots of the tap-dancing action over the course of one video.

After the acquisition, point-light displays were processed using Qualisys, ensuring visibility of all 14 recorded point-light markers during the entire recording time (for an overview of the position of the point-light markers, see Fig. 1). Note that we excluded videos containing movement errors, e.g. touching of a hurdle in BAS. Correspondingly, all sounds in both actions were produced by foot-ground contacts only.

Sound data were processed using Reaper v5.28 (Cockos Inc., New York, United States). First, stimulus intensities of hurdling and tap-dancing recordings were normalized separately. The frequency profiles of hurdling and tap-dancing sounds were then captured using the Reaper plugin Ozone 5 (iZotope Inc, Cambridge, United States), so that the spectral distributions of both recording types were equalized. Lastly, the difference curve (hurdling – tap dancing) was used by the plugin’s match function to adjust the tap-dancing spectrum to the hurdling reference. Point-light displays and sound were synchronized, and the subsequent videos were cut using Adobe Premiere Pro CC (Adobe Systems Software, Dublin, Ireland). The final videos had a size of 640x400 pixels, 25 frames per second, and an audio rate of 44 100 Hz. A visual fade-in and fade-out of 1 s (25 frames) were added with Adobe Premiere. Video length ranged from 3 to 6 s, with an average length of 5 s.

For the fMRI session, a subset of 35 hurdling and 35 tap-dancing videos was selected for each participant. This subset included the 27 hurdling and 27 tap-dancing videos from our previous fMRI study (Heins et al., 2020), and eight additional videos per action type, choosing the videos with the most reliable ratings from the test and retest sessions (see Section 2.3.2).

For every selected video, a mute version and additional “mute scrambled” versions were created using Adobe Premiere. For the mute version, the auditory tracks were removed from the videos. For the mute scrambled versions, the visual tracks of the mute videos were cut into one-second segments (25 frames) and were then rearranged. The same scrambling scheme was applied to all videos, that is, the segments were rearranged in a fixed order. All videos were

presented using the Presentation software (Version 18.1, Neurobehavioral Systems, Inc., Berkeley, CA).

2.3 Procedure

2.3.1 Training and filming sessions

Participants were trained in hurdling and tap dancing by professional instructors for a nine-week period (Fig.2). Ninety-minute training sessions were conducted twice a week for both hurdling and tap dancing, so that participants trained both action types for three hours a week. Before this training period, none of the participants had ever practiced hurdling or tap dancing. Participants had to take part in four out of six offered filming sessions, taking place at different states of training, to cover changes in performance. The first filming sessions took place two weeks after the training started, with the following filming sessions taking place four, five, six, eight, and nine weeks after training start (Fig. 2). During the filming sessions, participants were equipped with 12 point-light markers (see Fig. 1) and filmed via infra-red cameras of the motion capturing system while performing both action types. The hurdling action consisted of three hurdle transitions (Fig. 1), while the tap-dancing action was a movement sequence learned during the tap-dancing training sessions. Both actions increased in difficulty within the four sessions. For hurdling, the spatial distance between the three hurdles increased, requiring more speed and a smoother hurdle clearance. For tap dancing, action elements were added to the sequence to increase difficulty.

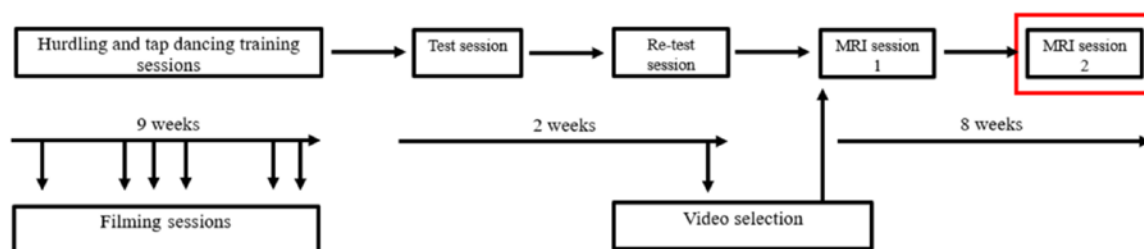


Fig. 2. Procedure of the study. Participants were filmed on several occasions during their nine-week training in hurdling and tap dancing. Two behavioral sessions and one MRI session (results reported in Heins et al., 2020) were conducted before the here presented MRI session.

2.3.2 Behavioral test and retest sessions

Behavioral test-retest sessions were conducted to find the stimuli with the highest reliability of participants' rating. Both sessions were conducted in a computer lab. Participants

were seated in front of a computer and instructed to rate the quality of their actions on a scale from 1 (“*not well at all*”) to 6 (“*very well*”) based on their subjective impression. The experiment consisted of two blocks with self-paced responses, with both blocks lasting between 20 to 30 minutes. The same videos were presented in a different order in the two blocks of the experiment. Videos were pseudorandomized so that not more than three videos in a row showed the same action type (hurdling vs. tap dancing). Overall durations of the test sessions ranged from 40 to 60 minutes depending on rating speed. Two weeks after the test session, participants were presented the same videos once more (in pseudo-randomized order). Thirty-five videos for both hurdling and tap dancing were chosen per participant and were used in the current fMRI session. The videos with the highest reliability in rating were chosen. Each video was rated a total of four times (two times in the test and two times in the retest sessions. 18.69 % of all chosen videos (840 videos in total, 70 per participant) received the same rating on all four repetitions, in 70.71 % ratings varied by a score of either +1 or -1 in one or two of the repetitions (ratings diverging in one direction), and in 10.6 % ratings varied by a score of ± 1 (ratings diverging in both directions). Twenty-seven videos for hurdling and tap dancing were previously used in the preceding fMRI study (Heins et al., 2020) .

2.3.2 fMRI session

For the fMRI session, participants completed the same task as during the test-retest sessions. Namely, they were instructed to rate the quality of the presented actions on a 6-point Likert scale. Note that participants were presented with videos of their own actions only. The participants were informed that regardless of the presence of mute and mute scrambled videos, they should still rate the quality of their performance using the given information. Participants were asked to regulate the volume of the sounds before the experiment started to assure that the action sounds were audible above the scanning noises. The experiment consisted of seven blocks, including 36 trials each (Fig. 3).

Transition probabilities ensured that each condition was preceded by every condition (including the same condition) in the same number of trials over the whole experiment. The first trial of a block was a repetition of the last trial of the preceding block, to avoid losing a transition. The remaining 35 trials consisted of five trials for each of the seven conditions (namely hurdling normal, tap dancing normal, hurdling mute, tap dancing mute, hurdling mute scrambled, tap dancing mute scrambled, and null events). Thus, after the discarding of the first trials after each pause, 245 trials remained for the analysis. These consisted of 210 video trials (35 trials per conditions) and 35 null events, where a fixation cross was presented. The duration

of the null events was 5 s. Before each trial, a fixation cross was presented as an interstimulus interval, varying between 3.5 to 4.5 secs in length. After each video trial, the rating scale from 1 to 6, including the rating question, was presented. The experiment continued upon the participants' button press.

After the experiment, lasting approximately 45 minutes, an eight-minute resting-state sequence was acquired. Participants were asked to look at a fixation cross for the whole period. Throughout the entire scanning routine, participants were instructed to refrain from moving.

2.4 fMRI recordings and preprocessing

Participants were scanned in a 3-Tesla Siemens Magnetom Prisma MR tomograph (Siemens, Erlangen, Germany) using a 20-channel head coil. A 3D-multiplanar rapidly acquired gradient-echo (MPRAGE) sequence was used to obtain high resolution T1 weighted images ahead of functional scanning, with scanning parameters set to 192 slices, a repetition time (TR) of 2130 msec, an echo time (TE) of 2.28 msec, slice thickness of 1 mm, a field of view (FoV) of 256 x 256 mm², and a flip angle of 8°.

Gradient-echo echoplanar imaging (EPI) was used to measure blood-oxygen-level-dependent (BOLD) contrast for functional imaging data of the whole brain. There were eight EPI sequences in total. One sequence for the volume adjustment and one sequence for each of the seven experimental blocks. Scanning parameters were set to a TE of 30 msec, a TR of 2000 msec, a flip angle of 90°, 33 slices with a slice thickness of 3 mm, and a FoV of 192x192 mm².

Imaging data were processed using SPM12 (Wellcome Trust, London, England). Slice time correction to the middle slice was performed, followed by realignment of all individual functional MR (EPI) images to correct for three-dimensional motion. The individual's structural scan was co-registered to the mean functional image and then segmented into the native space tissue components. Both the structural and the functional images were normalized into the standard MNI space (Montreal Neurological Institute, Montreal, QC, Canada). Spatial smoothing of the functional images was performed with a Gaussian kernel of full-width at half maximum (FWHM) of 8 mm. Finally, high-pass temporal filtering equivalent to 128 s was applied.

2.5 Statistical data analysis

2.5.1 Behavioral data analysis

We calculated Kolmogorov-Smirnov tests to assure the normal distribution of our behavioral rating score data. To examine a potential reduction (Hypothesis 1) in rating scores

by sound omission for GAS vs. BAS actions, we calculated a 2 x 2 repeated measures analysis of variance (rmANOVA) on mean rating scores using SPSS (IBM, New York, United States), and predicted a significant interaction effect.

The rmANOVA included the factor ACTION, with the factor levels *BAS (hurdling)* and *GAS (tap dancing)*, and the factor SOUND with factor levels *sound normal* and *mute*.

To further examine significant interaction effects, factors were held constant and simple main effects as well as paired *t*-test were calculated. The significance level was set to $\alpha = .05$ and Bonferroni-correction was performed for multiple comparisons.

2.5.2 fMRI design specification

The design was implemented in SPM12, following a general linear model (GLM, (Friston et al., 1994; Worsley & Friston, 1995) approach. The modeled activation was time-locked to the onsets of the videos or null events. Epochs contained the full presentation period ranging from 3 to 6 s for the videos, and 5 s for the null events. In our previous fMRI study (Heins et al., 2020), we controlled for the differing auditory event density, that is, the number of distinguishable auditory sounds occurring per second, of hurdling and tap dancing, by introducing regressors of nuisance. As these did not yield any effect, and in the current study, most videos are presented without sound, we refrained from including regressors of nuisance here. The GLM for every participant consisted thus of 13 regressors in total: six regressors for the experimental conditions (BAS normal, GAS normal, BAS mute, GAS mute, BAS mute scrambled, GAS mute scrambled), one regressor for the null events, and six regressors for the motion parameters (three translations and three rotations). Activation for 35 trials was considered for the modeling of each of the six regressors of interest and the null event regressor. All these regressors were convolved with the hemodynamic response function.

On the first level, *t*-contrasts of the experimental conditions against null events (5 s fixation cross) were calculated (condition > rest). These contrast images were then used to set up a flexible factorial design on the second level. The flexible factorial design was chosen because it accounts best for the within-subject factor. The model consisted of 18 regressors – six regressors for the experimental conditions, and 12 regressors for the subject effects, one for each participant.

Region of interest (ROI) analyses were performed to test our hypotheses and replicate the results from our previous fMRI study. Therefore, functional ROIs were defined using the peak voxels of activation from our previous study (Heins et al., 2020), including right primary auditory cortex ($x = 54, y = -13, z = 5$), right pSTG ($x = 54, y = -31, z = 5$) and SMA ($x = -3, y$

= -4, $z = 68$) The ROIs were defined as spheres of 6 mm around the peak coordinates. Firstly, we performed a ROI analysis for the primary auditory cortex for the BAS>GAS contrasts, including only the conditions with sound, to test for the hypothesized stronger sensory attenuation for tap dancing than for hurdling. Then, ROI analyses for the secondary auditory cortex (pSTG) and the SMA for the GAS>BAS contrasts were performed, including both the conditions with sound and without sound (but intact picture) to investigate a presumably stronger top-down predictive signal for tap dancing due to explicit sound predictions (Hypothesis 2). We performed a ROI analysis for the ACTION x SOUND using ROIs for the pSTG, to examine the differential involvement of secondary auditory cortices in the automatic update of the sound prediction model of action sound representation for mute versions of hurdling and tap dancing. We expected more activation in the mute tap-dancing conditions, caused by a more pronounced mismatch whenever the sound is unexpectedly omitted from the perceived action. For a more profound investigation of the interaction effects, we extracted beta values from the pSTG for all conditions (Hypothesis 4). We used false-discovery rate (FDR) correction and a threshold of $p < .05$ for all ROI analyses.

F-contrasts were used to calculate the main effect for SOUND, contrasting the normal and mute conditions. *T*-contrasts were calculated as a follow-up, to examine the direction of activation, to ensure that there was indeed a decrease of auditory cortex activation for the mute videos despite the noisy fMRI environment (Hypothesis 3). Resulting *t*-maps were corrected using false discovery rate (FDR) correction and a threshold of $p < .001$.

In addition, we extracted beta values from the SMA to investigate the predictive influence of the SMA on rating scores (Hypotheses 5 and 6).

3 Results

3.1 Behavioral Results

We tested whether sound omission reduces rating scores to a higher degree in GAS actions compared to BAS actions. The corresponding ANOVA included the factors ACTION (*BAS*, *GAS*) and SOUND (*normal*, *mute*). The interaction between ACTION and SOUND was significant (Hypothesis 1), $F(1,11) = 5.26$, $p = .005$, $\eta^2 = 0.323$, with the *mute* condition ($M = 3.56 \pm 0.85$) having lower rating scores than the *normal* condition ($M = 3.73 \pm 0.88$) in tap dancing (*GAS*), and higher rating scores ($M = 3.76 \pm 0.61$) than the *normal* condition ($M = 3.62 \pm 0.54$) in hurdling (*BAS*, Fig. 3). Post-hoc paired *t*-tests did not yield any significant results (all $p > .05$).

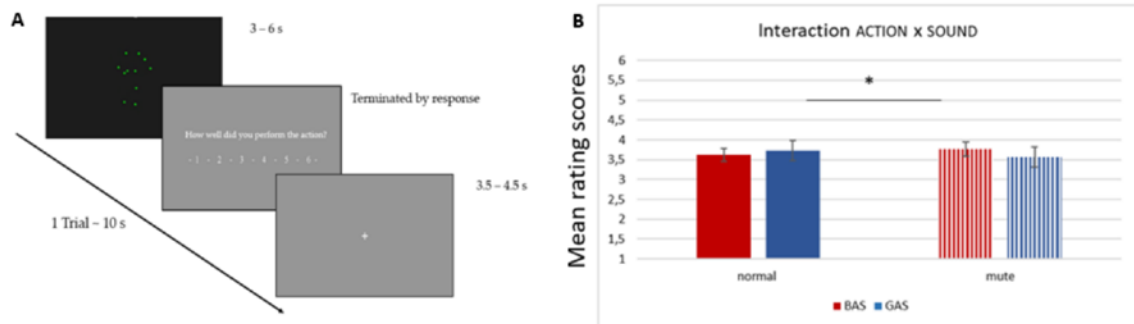


Fig. 3. Schematic representation of a trial and behavioral interaction effect. (A) A trial consists of a video (3 – 6 s in length), followed by the video rating question (“How well did you perform the action?” in German), and a fixation cross (3.5 – 4.5 s in length). The total duration of one trial is approximately 10 s. (B) Interaction effect of the factors ACTION and SOUND for the behavioral rating scores. Mean rating scores for BAS are presented in red, mean rating scores for GAS in blue. Behavioral rating scores ranged from 1 – 6. Error bars represent the standard error of the mean.

3.2 fMRI results

The whole-brain contrast GAS>BAS revealed activation in the SMA and pSTG, corroborating Hypothesis 2. As in our previous study, the reverse contrast revealed higher activation in an occipital region posterior to the ascending portion of the inferior temporal sulcus, particularly in the right hemisphere (Fig. 4). Most probably, this area corresponded to the occipital place area (OPA), which is suggested to encode egocentric distance information for navigation as one traverses the environment (Persichetti & Dilks, 2016). More specifically, OPA has been proposed to contribute to navigating the local visual environment rather than orienting the individual to the broad environment (Kamps, Julian, Kubilius, Kanwisher, & Dilks, 2015). It fits very nicely to this functional interpretation that we here found OPA to be significantly more active in observing point light displays of one’s own hurdling performance, entailing a travelling distance of about 20 m, as compared to stationary tap dancing.

Contrary to our expectations, a stronger attenuation effect in A1 for GAS>BAS was neither found for the whole-brain contrast nor for the planned ROI analysis based on coordinates of Heins et al. (2020). To test for a subtler effect in A1, we extracted beta values from A1 and calculated a paired sample *t*-test to determine potential differences between BAS

and GAS with sound. This analysis revealed significant lower scores for GAS ($M = 1.12$, $SD = 0.33$) compared to BAS ($M = 1.46$, $SD = 0.5$, $t(11) = -3.05$, $p = .011$, Fig. 4).

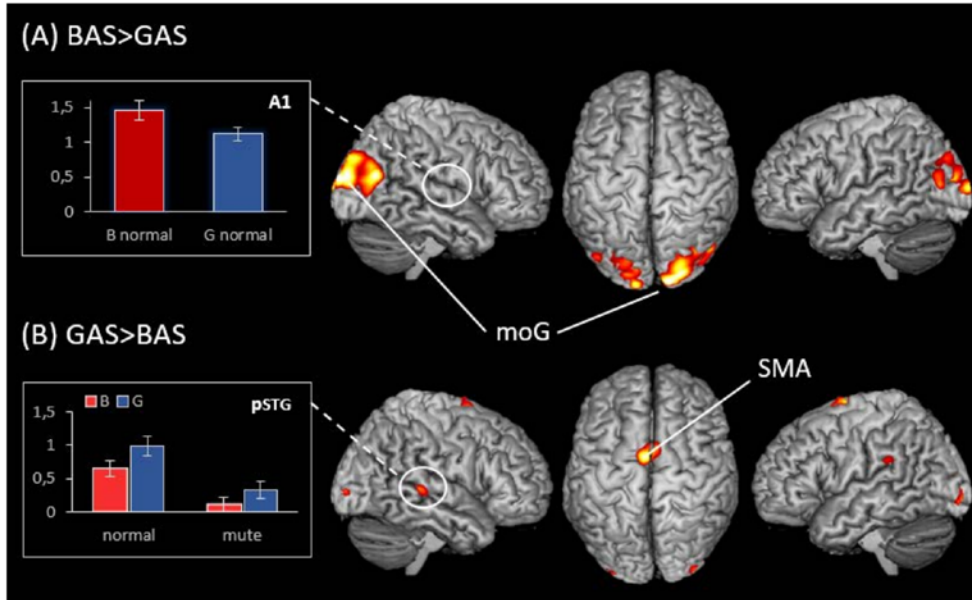


Fig. 4. Main effect of ACTION, including the normal and mute conditions. (A) FDR-corrected t-maps ($p < .001$) for the BAS>GAS contrast and A1 beta values for the BAS normal and GAS normal conditions. (B) FDR-corrected t-maps ($p < .001$) for the GAS>BAS contrast and pSTG beta values for the BAS normal, GAS normal, BAS mute and GAS mute conditions.

Regarding Hypothesis 3, sound omission resulted in lower activity in auditory cortex, more specifically pSTG (peak voxel $x = -51$, $y = -22$, $z = 8$, $x = 57$, $y = -25$, $z = 11$, Fig. 5). Contrary to Hypothesis 4, the omission effect was not less pronounced in GAS vs. BAS actions (as expected by a stronger prediction error signal); i.e., neither the ROI analysis nor beta values extracted from pSTG differentiated between GAS and BAS actions (Fig. 4). The contrast mute>sound showed an effect in the inferior occipital gyrus (peak at $x = -45$, $y = -70$, $z = 2$, $x = 45$, $y = -67$, $z = -4$), corresponding to hMT+ (Kolster, Peeters, & Orban, 2010, Fig. 5).

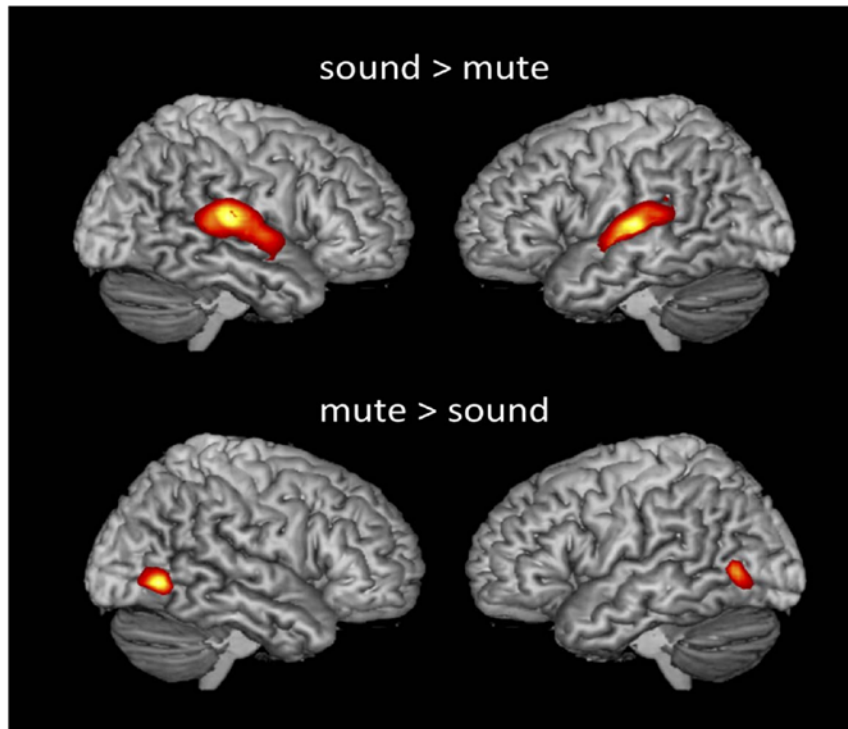


Fig. 5. Effects of muting the sound in hurdling and tap dancing. FDR-corrected t-maps ($p < .001$) are shown for sound > mute, resulting in a bilateral activity of the STG (A1), and for mute > sound, yielding increased activity in the inferior occipital gyrus, corresponding to area hMT+.

To further examine this activation, we performed a post-hoc analysis of the beta values extracted from hMT+ (peak voxels taken from Kolster et al., 2010) by correlating them with the SMA beta values and the rating scores for the mute BAS and mute GAS conditions. For the mute BAS condition, we found correlations between the rating scores and left ($r = .620, p = .032$) and right ($r = .864, p < .001$) hMT+ beta values. Neither the left ($r = .279, p = .379$) nor the right ($r = .531, p = .075$) hMT+ beta values correlated significantly with SMA beta values. For the mute GAS conditions, left hMT+ beta values correlated with both the rating scores ($r = .742, p = .006$) and SMA beta values ($r = .752, p = .005$).

According to Hypothesis 5 and Hypothesis 6, we expected that judging the quality of GAS action is more dependent on sound than judging BAS actions, and hence, judging mute GAS actions but not mute BAS actions was expected to resort to higher, multimodal levels of the predictive hierarchy, especially SMA. Here, we found a significant correlation between the rating scores and the beta values of the SMA ($r = .518, p = .042$, one-sided testing) for the mute

GAS condition, whereas there was no significant correlation in the mute BAS condition ($r = .214, p = .253$, Fig. 6). SMA beta values did not correlate with rating scores in the normal BAS ($r = .094, p = .386$) or normal GAS condition ($r = .349, p = .133$, both one-sided testing). As expected, our control condition revealed that this correlation was not significant for mute scrambled GAS videos ($r = .338, p = .141$).

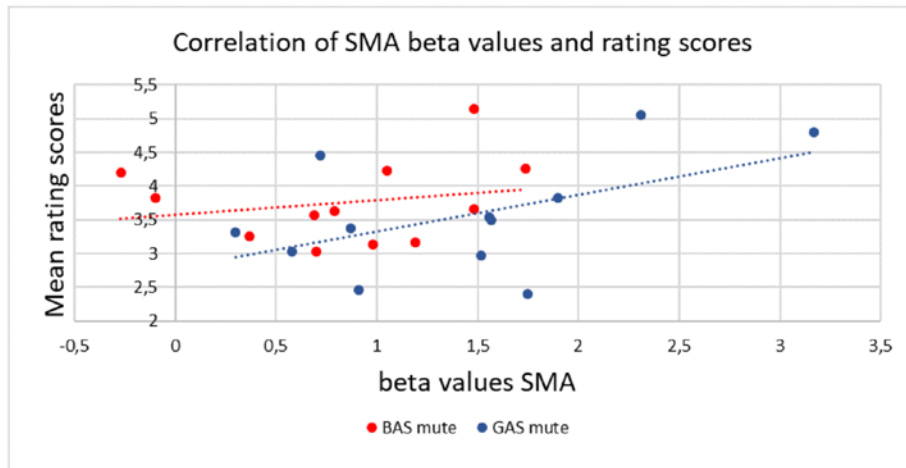


Fig. 6. Correlation of SMA beta values and behavioral rating scores. Correlation between the mean behavioral rating scores and beta values extracted from the SMA ($x = -3, y = -4, z = 68$) for the BAS mute condition (presented in red) and the GAS mute condition (presented in blue). Behavioral rating scores ranged from 1 – 6.

4 Discussion

Some of our action sounds are intended, others are produced rather incidentally. While this difference is intuitive on a subjective level, it remains unclear whether it is maintained on the neural level. The concept of predictive coding suggests that in either case, artificially omitting sounds of actions results in a prediction error signal and subsequent adaptation of the generative predictive model. The present fMRI study tested whether these processes characteristically differentiate between sounds being action goals (GAS) and sounds being action by-products (BAS). With regard to a condition in which point-light videos were presented with their original action sounds, we replicated previous findings suggesting stronger auditory prediction in GAS than in BAS (Hypothesis 2, Heins et al., 2020). Thus, SMA and pSTG were more activated for observing GAS as compared to BAS actions, and correspondingly, beta scores in the primary auditory cortex were significantly lowered. With regard to the new manipulation, we found condition-specific effects in both brain activity and

behavior pointing to significant differences in the perception of masked actions sounds that were initially produced either intentionally or incidentally.

As hypothesized (Hypothesis 1), mute GAS actions generally obtained lower quality rating scores by the participants than GAS actions with sound. This finding extends previous findings, where scrambled (i.e. less informative) auditory input was associated with decreased rating scores in GAS actions (Heins et al., 2020) , and other studies observing that the deprivation of auditory feedback has a destabilizing effect on GAS actions (Repp, 1999; Waldstein, 1990). Moreover, while BAS actions were less affected by auditory scrambling than GAS actions in our previous study, rating scores for mute BAS actions even increased relative to BAS actions with sound in the current study. This fits to the results of Kennel and colleagues (2015) who did not find interfering effects of sound deprivation on BAS actions either. While action sounds have been deemed beneficial for motor learning in BAS actions (Aglioti & Pazzaglia, 2010; Schaffert, Janzen, Mattes, & Thaut, 2019), our results suggest that they are not mandatory, at least not at the level of expertise which our participants achieved after nine weeks of training. BAS might get more important with increasing levels of expertise, as is suggested when comparing the neural response to sport sounds in novices and experts (Woods, Hernandez, Wagner, & Beilock, 2014). For moderately trained novices like our participants, a pure visually-based performance evaluation was presumably favorable for BAS, while auditory input generally improved performance ratings in GAS. This interpretation is also corroborated by the finding that the contrast between BAS vs. GAS trials with sound yielded enhanced activity in the middle occipital gyri which are specialized for visuo-spatial information processing.

These differential effects of BAS vs. GAS sound omission were subtly reflected in our brain activation results. Hardly surprising, the omission of action sounds generally resulted in decreased activity in A1 and pSTG (Hypothesis 3). We reasoned that this effect should be less pronounced for GAS vs. BAS actions if auditory predictions generate activity in these downstream auditory areas partially compensating omitted sounds in GAS (Hypothesis 4); however, this was not the case. Instead, we found stronger activation in motion-sensitive area hMT+ (V5) for both mute conditions implying that, as soon as the sound was missing as a source of information, participants more strongly resort to the processing of the visual motion. This points to a strategic switch to the body movement as the only available basis of qualitative evaluation. Compensatory switches to a visual evaluation when auditory cues are removed have been reported before (Repp & Keller, 2004; van Kemenade, Arikan, Kircher, & Straube, 2016).

Based on the assumption of a hierarchical structure for action sound prediction, we expected the effect of sound omission to clearly differ from the effect of sound distortion that was applied in a previous study. While distorting action sounds triggered increased activity on the level of auditory processing (pSTG) in the precursor fMRI study (Heins et al., 2020), the full omission of action sounds applied in the present study was expected to violate prediction not merely at the level of auditory quality but at a yet higher, multimodal level of the sensorimotor model, that is, SMA (Jo et al., 2019; Lima et al., 2016). We reasoned that if sounds are a fundamental component in the predictive model of GAS actions, their omission would lead to a substantial adaptation of this multimodal model. Mute BAS actions in contrast, presumably relying much less on the auditory modality, should not trigger as strong an adaptation at the SMA level.

In support of these hypotheses (Hypothesis 5 and 6), rating scores for *mute* videos were positively correlated with SMA beta values for GAS, but not for BAS. Importantly, this correlation for GAS vanished when the visual channel was scrambled. Thus, the SMA appears to provide sufficient predictive input about the quality of the action in mute GAS, as long as enough action information is attainable through visual input. As to the specific role of SMA in this process, this area may provide the specific auditory predictive input missing in the current stimulus, which might align with sensorimotor areas being more active during the mental imagery of action sounds (Gu et al., 2019). Alternatively, SMA may rather enhance proprioceptive or visual portions of the generative model to enable the subject to engage in a quality rating of the observed movement (Langner et al., 2011). As the systematic relation between SMA activation and rating scores vanishes when scrambling the visual input, and sound omission yields stronger activation in motion-sensitive area hMT+, our results seem to favor the explanation of an enhancement of other than auditory modalities through SMA activity. In this context, it is interesting that muting either GAS or BAS actions increased activity in hMT+. A post-hoc analysis of hMT+ beta values showed them to correlate with the rating scores for mute BAS, while they correlated with both the rating scores and the beta values extracted from SMA for mute GAS. How can this pattern of results be explained?

First, it is important to note that the increase of activity in hMT+ for the mute conditions coincided with a more positive evaluation of BAS actions, and a more negative evaluation of GAS actions. The evaluation, solely based on visual input, thus led to a positive bias for the rating in BAS. Here, the auditory information does not seem important enough to cause a model retrieval at a level as high as the SMA to adjust for the missing sound. For GAS, in contrast, the auditory input seems to be essential for action evaluation: in contrast to the mute BAS

condition, the stronger activation in hMT+ in the mute GAS condition did not co-occur with higher rating scores. We therefore assume that higher ratings for the mute GAS were driven by processes in the SMA: In this condition, predictive input had to be obtained from the highest level in the model hierarchy to enhance visual processing and to guarantee a visually motion-based evaluation as in BAS. Speculatively, this pattern of results may reflect a driving input from SMA to hMT+ in mute GAS videos, conveying a multimodal action model required for the evaluation task in this condition.

Taken together, findings suggest that especially the SMA is important for processing actions whose sounds are part of the *intended* action goal rather than an incidental by-product. Here, SMA seems to provide predictive information about the performed action, a function that particularly manifests when the actual auditory output is lacking. Notably, the chosen actions, tap dancing (GAS) and hurdling (BAS), might still be actions on a continuum of sound-generating actions, ranking from more intended to more incidental action sounds. However, our results further emphasize the differences between the two, most essentially the importance of top-down prediction provided by the SMA, when evaluating GAS vs BAS.

The extent to which sounds of one's actions are efficiently used to (implicitly or explicitly) evaluate and improve our performance may strongly depend on proficiency. The relevance of BAS and GAS for action execution is not a strict dichotomy, with GAS being crucially important and BAS being irrelevant. A description on a continuum might be more accurate. Thus, the sound of a hurdling action may classify as neglectable – a true BAS - for a novice but as rather relevant - closer to GAS - for a professional hurdling athlete. Notably, natural movement sounds can be strategically exploited to improve athletes' movement execution (for review, see Schaffert et al., 2019). So, whereas one and the same action may classify as BAS or GAS action depending on previous training, it is important to note that in our sample, participants were untrained – and hence equally familiar with – hurdling and tap dancing before they participated in our study. Additionally, they consistently judged their progress in both disciplines as being equal during the course of the training protocol, as reflected in their own performance rating scores as well, which did not differ between the two action types. Therefore, we consider differences in BAS and GAS producing actions to be representative for incidentally and intentionally produced action sounds, with hurdling and tap dancing providing appropriate operationalizations in (untrained or moderately trained) novices.

As a final caveat, the final sample size of our study was comparably small since more participants than expected dropped out in the course of the demanding nine-week training and videotaping protocol due to various reasons. This drop-out rate further increased in the time

when we generated the final stimulus material. Note that the stimulus material was tailored for each participant on an individual level, including elaborate checking, editing and correction of the original videotapes (about 240 videos per participant) from the training phase. This procedure took about 70 hours per participant, including those which dropped out before the fMRI session. Replicating the results of a previous study (Heins et al., 2020), Hypothesis 2 therefore considerably increased confidence in our findings. Moreover, correcting for multiple comparisons using FDR at 0.05 satisfies the standards for false discovery control in fMRI studies. Still, future studies need to further validate our findings, including also correlation between beta values and rating scores, in different and larger samples.

5 Conclusion

Action sounds are ubiquitous in our everyday life, but research on real-life action sounds other than speech and music is still very scarce. The present study is the first to demonstrate the neural effects of omission of real-life action sounds and provides new insights into natural action sounds and their relevance for action performance evaluation. Differential BOLD and behavioral effects suggest sound omission to trigger adaptive processes on the highest level of the predictive hierarchy in tap dancing but not hurdling. Our findings empirically underline the intuitive difference between actions with intentional and incidental sound generation.

Acknowledgements

We would like to thank Monika Mertens for her help during data collection. Furthermore, we thank Theresa Eckes, Alina Eisele, Marie Kleinbielen, Katharina Thiel and Stefan Vinbr x for their assistance during filming and with creating the stimulus material, Daniel Kluger for helping with the processing of the auditory stimuli, and Ima Trempler for helpful comments on an earlier version of the manuscript. Finally, we would like to thank Amelie Huebner, Marlen Roehle and Lena Schliephake for rewarding discussions. The project was funded by the Deutsche Forschungsgemeinschaft (SCHU 1439-4-2).

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Aglioti, S. M., & Pazzaglia, M. (2010). Representing actions through their sound. *Experimental Brain Research*, 206(2), 141–151. <https://doi.org/10.1007/s00221-010-2344-x>
- Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology*, 48(9), 1276–1283. <https://doi.org/10.1111/j.1469-8986.2011.01196.x>
- Carbajal, G. V., & Malmierca, M. S. (2018). The Neuronal Basis of Predictive Coding Along the Auditory Pathway: From the Subcortical Roots to Cortical Deviance Detection. *Trends in Hearing*, 22, 1–33. <https://doi.org/10.1177/2331216518784822>
- Dürschmid, S., Edwards, E., Reichert, C., Dewar, C., Hinrichs, H., Heinze, H. J., ... Knight, R. T. (2016). Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 113(24), 6755–6760. <https://doi.org/10.1073/pnas.1525030113>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Friston, K. (2012). Prediction, perception and agency. *International Journal of Psychophysiology*. <https://doi.org/10.1016/j.ijpsycho.2011.11.014>
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. -P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4), 189–210. <https://doi.org/10.1002/hbm.460020402>
- Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro, E., Andrew, C. M., ... McGuire, P. K. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex*, 16(7), 969–977. <https://doi.org/10.1093/cercor/bhj039>
- Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, 120(3), 453–463. <https://doi.org/10.1016/j.clinph.2008.11.029>
- Gu, J., Zhang, H., Liu, B., Li, X., Wang, P., & Wang, B. (2019). An investigation of the

- neural association between auditory imagery and perception of complex sounds. *Brain Structure and Function*, 224(8), 2925–2937. <https://doi.org/10.1007/s00429-019-01948-z>
- Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, 389, 54–73. <https://doi.org/10.1016/j.neuroscience.2017.07.061>
- Heins, N., Pomp, J., Kluger, D. S., Trempler, I., Zentgraf, K., Raab, M., & Schubotz, R. I. (2020). Incidental or Intentional? Different Brain Responses to One's Own Action Sounds in Hurdling vs. Tap Dancing. *14*(May), 1–11. <https://doi.org/10.3389/fnins.2020.00483>
- Horváth, J. (2015). Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain Research*. <https://doi.org/10.1016/j.brainres.2015.03.038>
- Howell, P. (2004). Effects of delayed auditory feedback and frequency-shifted feedback on speech control and some potentials for future development of prosthetic aids for stuttering. *Stammering Research: An on-Line Journal Published by the British Stammering Association*, 1(1), 31–46. <https://doi.org/10.1126/science.1182395>
- 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(1), 133–151. <https://doi.org/10.1037/a0028566>
- Jo, H. G., Habel, U., & Schmidt, S. (2019). Role of the supplementary motor area in auditory sensory attenuation. *Brain Structure and Function*, (0123456789). <https://doi.org/10.1007/s00429-019-01920-x>
- Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research*, 190(3), 279–287. <https://doi.org/10.1007/s00221-008-1473-y>
- Kaiser, J., & Schütz-Bosbach, S. (2018). Sensory attenuation of self-produced signals does not rely on self-specific motor predictions. *European Journal of Neuroscience*, 47(11), 1303–1310. <https://doi.org/10.1111/ejn.13931>
- Kalinowski, J., Armson, J., Roland-Mieszkowski, M., Stuart, A., & Gracco, V. L. (1993). Effects of alteration in auditory feedback and speech rate on stuttering frequency. *Language and Speech*, 36(1), 1–16.
- Kamps, F., Julian, J., Kubilius, J., Kanwisher, N., & Dilks, D. (2015). The occipital place area represents the local elements of scenes. *Journal of Vision*, 15(12), 514. <https://doi.org/10.1167/15.12.514>

- Keough, D., & Jones, J. A. (2009). The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing. *The Journal of the Acoustical Society of America*, *126*(2), 837–846. <https://doi.org/10.1121/1.3158600>
- Keysers, C., & Gazzola, V. (2014). Hebbian learning and predictive mirror neurons for actions, sensations and emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1644). <https://doi.org/10.1098/rstb.2013.0175>
- Kolster, H., Peeters, R., & Orban, G. A. (2010). The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. *Journal of Neuroscience*, *30*(29), 9801–9820. <https://doi.org/10.1523/JNEUROSCI.2069-10.2010>
- Langner, R., Kellermann, T., Boers, F., Sturm, W., Willmes, K., & Eickhoff, S. B. (2011). Modality-specific perceptual expectations selectively modulate baseline activity in auditory, somatosensory, and visual cortices. *Cerebral Cortex*, *21*(12), 2850–2862. <https://doi.org/10.1093/cercor/bhr083>
- Liebenthal, E., Ellingson, M. L., Spanaki, M. V., Prieto, T. E., Ropella, K. M., & Binder, J. R. (2003). Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *NeuroImage*, *19*(4), 1395–1404. [https://doi.org/10.1016/S1053-8119\(03\)00228-3](https://doi.org/10.1016/S1053-8119(03)00228-3)
- Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of Supplementary Motor Areas in Auditory Processing and Auditory Imagery. *Trends in Neurosciences*, *39*(8), 527–542. <https://doi.org/10.1016/j.tins.2016.06.003>
- Lincoln, M., Packman, A., & Onslow, M. (2006). Altered auditory feedback and the treatment of stuttering: A review. *Journal of Fluency Disorders*. <https://doi.org/10.1016/j.jfludis.2006.04.001>
- Mathiak, K., Rapp, A., Kircher, T. T. J., Grodd, W., Hertrich, I., Weiskopf, N., ... Ackermann, H. (2002). Mismatch responses to randomized gradient switching noise as reflected by fMRI and whole-head magnetoencephalography. *Human Brain Mapping*, *16*(3), 190–195. <https://doi.org/10.1002/hbm.10041>
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. <https://doi.org/10.1016/j.clinph.2007.04.026>
- Nazimek, J. M., Hunter, M. D., Hoskin, R., Wilkinson, I., & Woodruff, P. W. (2013). Neural basis of auditory expectation within temporal cortex. *Neuropsychologia*, *51*(11), 2245–2250. <https://doi.org/10.1016/j.neuropsychologia.2013.07.019>
- Oldfield, R. C. (1971). The assessment and analysis of handedness. *Neuropsychologia*, *9*(1),

97–113.

- Persichetti, A. S., & Dilks, D. D. (2016). Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *77*, 155–163. <https://doi.org/10.1016/j.cortex.2016.02.006>
- Pfordresher, P. Q., & Beasley, R. T. E. (2014). Making and monitoring errors based on altered auditory feedback. *Frontiers in Psychology*, *5*(AUG), 1–13. <https://doi.org/10.3389/fpsyg.2014.00914>
- Recasens, M., Grimm, S., Capilla, A., Nowak, R., & Escera, C. (2014). Two sequential processes of change detection in hierarchically ordered areas of the human auditory cortex. *Cerebral Cortex*, *24*(1), 143–153. <https://doi.org/10.1093/cercor/bhs295>
- Recasens, M., & Uhlhaas, P. J. (2017). Test–retest reliability of the magnetic mismatch negativity response to sound duration and omission deviants. *NeuroImage*, *157*, 184–195. <https://doi.org/10.1016/j.neuroimage.2017.05.064>
- Repp, B. H. (1999). Effects of auditory feedback deprivation on expressive piano performance. *Music Perception*, *16*(4), 409–438. <https://doi.org/10.2307/40285802>
- Repp, B. H., & Keller, P. E. (2004). Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*. Repp, Bruno H.: Haskins Laboratories, 270 Crown Street, New Haven, CT, US, 06511-6695, repp@haskins.yale.edu: Taylor & Francis. <https://doi.org/10.1080/02724980343000369>
- Schaffert, N., Janzen, T. B., Mattes, K., & Thaut, M. H. (2019). A Review on the Relationship Between Sound and Movement in Sports and Rehabilitation. *Frontiers in Psychology*, *10*(February), 244. <https://doi.org/10.3389/fpsyg.2019.00244>
- Timm, J., Schönwiesner, M., Schröger, E., & SanMiguel, I. (2016). Sensory suppression of brain responses to self-generated sounds is observed with and without the perception of agency. *Cortex*, *80*, 5–20. <https://doi.org/10.1016/j.cortex.2016.03.018>
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443. <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- van Kemenade, B. M., Arikan, B. E., Kircher, T., & Straube, B. (2016). Predicting the sensory consequences of one’s own action: First evidence for multisensory facilitation. *Attention, Perception, and Psychophysics*, *78*(8), 2515–2526. <https://doi.org/10.3758/s13414-016-1189-1>
- Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S.

- (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, *108*(51), 20754–20759.
<https://doi.org/10.1073/pnas.1117807108>
- Waldstein, R. S. (1990). Effects of postlingual deafness on speech production: Implications for the role of auditory feedback. *Journal of the Acoustical Society of America*. US: Acoustical Society of American. <https://doi.org/10.1121/1.400107>
- Woods, E. A., Hernandez, A. E., Wagner, V. E., & Beilock, S. L. (2014). Expert athletes activate somatosensory and motor planning regions of the brain when passively listening to familiar sports sounds. *Brain and Cognition*, *87*(1), 122–133.
<https://doi.org/10.1016/j.bandc.2014.03.007>
- Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited — Again. *NeuroImage*. <https://doi.org/10.1006/nimg.1995.1023>
- Yamamoto, A. K., Parker Jones, O., Hope, T. M. H., Prejawa, S., Oberhuber, M., Ludersdorfer, P., ... Price, C. J. (2019). A special role for the right posterior superior temporal sulcus during speech production. *NeuroImage*.
<https://doi.org/10.1016/j.neuroimage.2019.116184>

3.3 Study III: Too late! Influence of temporal delay on the neural processing of one's own incidental and intentional action sounds.

Running title: Delay of incidental vs. intentional action sounds

Nina Heins, Ima Trempler, Karen Zentgraf,
Markus Raab, & Ricarda I. Schubotz (2020)

Manuscript submitted for publication

Too late! Influence of temporal delay on the neural processing of one's own incidental and intentional action sounds

Nina Heins^{1,2}, Ima Trempler^{1,2}, Karen Zentgraf³, Markus Raab^{4,5}, Ricarda I. Schubotz^{1,2}

¹Department of Psychology, University of Münster, Germany

²Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Germany

³Institute for Sport Sciences, Goethe University Frankfurt, Germany

⁴Institute of Psychology, German Sport University Cologne, Germany

⁵School of Applied Sciences, London South Bank University, United Kingdom

Abstract

The influence of delayed auditory feedback on action evaluation and execution of real-life action sounds apart from language and music is still poorly understood. Here, we examined how a temporal delay impacted the behavioral evaluation and neural representation of hurdling and tap-dancing actions in an fMRI experiment, postulating that effects of delay diverge between the two, as we create action sounds intentionally in tap dancing, but incidentally in hurdling. Delayed auditory feedback did not negatively influence behavioral rating scores in general. BOLD response transiently increased and then adapted to repeated presentation of point-light videos with delayed sound in the posterior superior temporal gyrus. This region also showed a significantly stronger correlation with the supplementary motor area (SMA) under delayed feedback. Notably, SMA activation increased more for delayed feedback in the tap-dancing condition, covarying with higher rating scores. Findings suggest that action evaluation is more strongly based on top-down predictions from SMA when sounds of intentional action are distorted.

1 Introduction

Most human actions produce sounds, and these are either the goal of the performed action (goal-related action sounds, G sounds hereafter) or a mere by-product (by-product action sounds, B sounds). This apparent difference suggests diverging neural implementation as well, especially with regard to areas that serve the selection and execution of action goals. For instance, we would expect that a tap-dancing sound (a G sound) is part of the brain's action goal representation, whereas the sound produced by a hurdling action (a B sound) is rather not. However, when performing a B action, we would be starkly surprised if the corresponding sound would not ensue, suggesting that B sounds are part of the brain's expectations during action execution.

Physiological evidence for this view is provided by sensory attenuation to self-initiated sounds, which manifests as reduction of amplitudes in early electroencephalography (EEG) or magnetoencephalography (MEG) components (Aliu, Houde, & Nagarajan, 2009; Baess, Horváth, Jacobsen, & Schröger, 2011; Timm, Schönwiesner, SanMiguel, & Schröger, 2014). Functional Magnetic Resonance Imaging (fMRI) studies indicate that this attenuation reflects decreased activity in the primary auditory cortex (A1) (Straube et al., 2017). Top-down modulations causing this suppression are conceived of as predictive models which are formed in higher cortical areas and conveyed to the respective sensory cortices to minimize prediction errors (Friston, 2005). Especially premotor areas are associated with a forward model which is important for the precise predictions about anticipated action effects, whether those are represented independently of their modality (Schubotz & Von Cramon, 2003) or specifically as auditory effects (Waszak, Cardoso-Leite, & Hughes, 2012; Woods, Hernandez, Wagner, & Beilock, 2014). With regard to self-produced sounds, A1 is suggested to be modulated by predictive models from supplementary motor area (SMA; Jo, Habel, & Schmidt, 2019) and the posterior superior temporal gyrus (pSTG; (Heilbron & Chait, 2018; Rauschecker, 2012), damping responses to expected sounds.

The functional role of self-induced action sounds has been investigated by modification of action sounds evoking prediction errors (Heilbron & Chait, 2018). Such auditory prediction errors have been studied mostly in language and music, where the action sound is the immediate goal of the action (G sounds). Delayed action sounds impair the process of speaking (Howell, 2004; Sasisekaran, 2012) and musical production (Finney, 1997; Pfordresher, 2006; Pfordresher & Beasley, 2014), although professional musicians seem to be less affected (van der Steen, Molendijk, Altenmüller, & Furuya, 2014). Qualitative manipulations of action sounds, e.g. manipulation of loudness or formant manipulation, evoke compensatory

articulation while speaking (Bauer et al., 2006; Purcell & Munhall, 2006; Tourville et al., 2008) and singing (Jones & Keough, 2008; Keough & Jones, 2009). This interfering influence is suggested to be either caused by a distorted feedback signal in higher cortical areas or by the automatic activation of competing forward models (Pfordresher, 2006).

In contrast to language and music, other types of action sounds are often a mere by-product of our actions (B sounds), like the sound produced by placing a cup back on a table or the sound of our own footsteps on the ground. While it is not our subjective goal to produce an audible sound by these actions, we might still be irritated if the sound differs from our expectations, and hence, B sounds may be also part of the predictive model on a neuronal level. As for G sounds, delaying B sounds has an interfering effect on action performance. For instance, delaying the sound of walking interferes with our sense of agency (Menzer et al., 2010). Moreover, Kennel et al. (2015) found that delayed auditory feedback during the performance of hurdling interferes with performance, but only for the first trial. The authors suggest a dynamic and very fast adaptation of the predictive forward-loop, comparable to the adaptation to temporal asynchrony in judging audiovisual stimuli (Vroomen, Keetels, De Gelder, & Bertelson, 2004). Also, sensory attenuation seems to be less pronounced for B sounds compared to G sounds, and higher cortical areas, especially the SMA, are more strongly involved in the processing of G as compared to B sounds (Heins et al., 2020). Thus, B actions may entail less predictive activity in the auditory system, and B sounds may be less relevant for adjusting forward models compared to G sounds. Further highlighting the importance of SMA in the predictive hierarchy of action sounds, this region was found to activate sensorimotor associations regarding action sounds (Lima, Krishnan, & Scott, 2016).

Building on these findings, we here examine the impact of delaying self-produced G and B sounds on their neural processing and the performance evaluation of G and B actions. To this end, we trained our participants in hurdling and tap dancing, to guarantee a motoric representation of these actions and their sensory consequences, including action sounds (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005; Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, 2006). In the subsequent fMRI experiment, participants were then presented with point-light videos of their own hurdling and tap-dancing performances (factor ACTION) either with their original sound or with their sound 200 ms delayed (factor DELAY), comparable to similar delays used in other studies (Kennel et al., 2015; Menzer et al., 2010). A number of four to six videos of the same delay type (delayed, undelayed) was presented consecutively to examine adaptation effects. Participants were instructed to rate the quality of their own hurdling and tap-dancing performance on a trial-by-trial basis.

Behaviorally, we expected the delayed presentation of sounds to lower the rating scores, for both B and G actions (Hypothesis 1). In addition, this decrease in rating scores was expected to be larger and to persist for repeated presentation of delayed G compared to delayed B sounds (Hypothesis 2).

With regard to the BOLD response effects, we hypothesized that A1 should be more activated, that is, less attenuated, for the B actions with undelayed sound compared to G actions with undelayed sound, whereas regions providing the predictive forward model, namely pSTG and SMA, should be more active for G actions compared to B actions in general (Heins et al., 2020, Hypothesis 3). We expected that the presentation of the delayed conditions would generate a prediction error, especially earlier compared to later presentations. The pSTG and also the anterior cingulate cortex (ACC) (Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004) should reflect this error signal in form of an activation increase (Hypothesis 4). Additionally, SMA activation should increase especially for the delayed G actions. Here, at the apex of the hierarchy, the predictive model is adapted to cope with the changed auditory input. This is essential when sounds are integral for action evaluation as are G sounds, whereas delayed B sounds can be ignored and should therefore lead to a less pronounced SMA response (Hypothesis 5). In both delayed conditions, activity in SMA and pSTG should be positively correlated, as SMA is suggested to regulate the error signal in pSTG via top-down predictions (Hypothesis 6).

Finally, SMA activation should positively correlate with behavioral rating scores in G, but not in B actions, due to the predictive input provided by this region whenever the produced sound is an integral part of the action, guaranteeing a positive evaluation of observed actions even when the sound differs from what we expect (Hypothesis 7).

2 Methods

2.1 Participants

Nineteen participants started the hurdling and tap-dancing training sessions. 18 of them finished the nine-week training period, and therefore, their video and audio data were processed. Four participants dropped out of the study after the training and one additional participant after the first fMRI session (Heins et al., 2020). Hence, 13 participants completed this fMRI session. One participant was excluded from the final analysis, because their reaction times recorded during the fMRI session diverged more than two standard deviations from the mean reaction time on a group level, leaving twelve participants (8 females, 4 males) for the analysis. This is a relatively small sample size, but we found robust results in our preceding fMRI study using

the same sample size. The participants' age ranged from 19 to 28 years ($M = 22.1$, $SD = 3.0$), and all of them were right-handed, as assessed by the Edinburgh Handedness Inventory (EHI; Oldfield, 1971), with scores varying from +60 to +100, with a mean of +85. All participants reported to have no history of psychiatric or neurological disorders. They signed an informed consent. After successful participation, participants were rewarded with both course credit and monetarily. The study was approved by the Local Ethics Committee in accordance with the Declaration of Helsinki.

2.2 Material

The stimuli used in the study consisted of point-light displays of hurdling and tap-dancing actions with their corresponding sounds, or sound delayed by 200 ms, recorded from each individual participant at different stages throughout training. Point-light displays were recorded using the Qualisys Motion Capture System (<https://www.qualisys.com>) with nine cameras (see Fig. 1), while the sound was recorded by in-ear microphones (Sound-man OKM Classic II) for hurdling and by a sound recording app on a mobile phone for tap dancing.

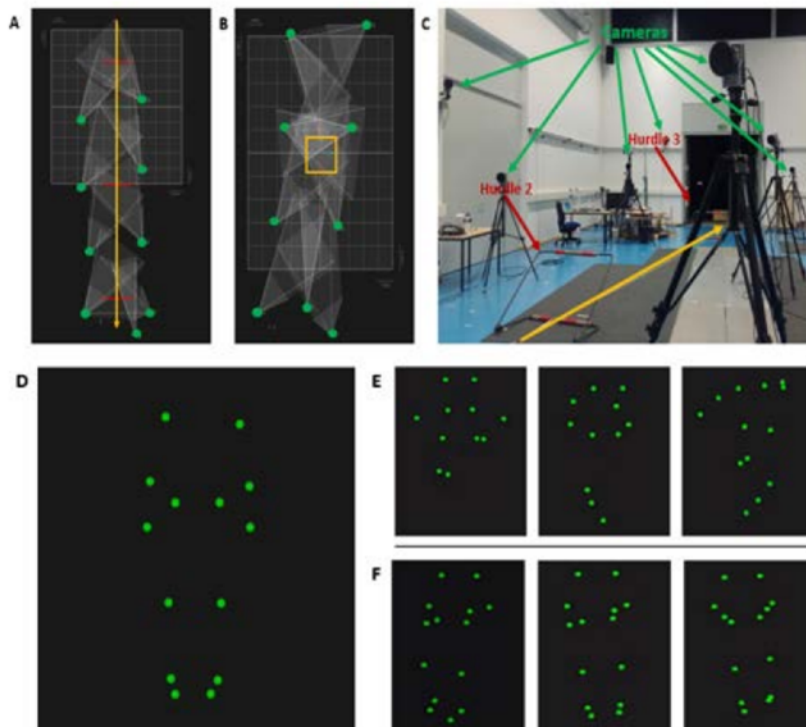


Fig. 1. Camera positions, set-up during the point light recordings and position of the point-light markers. (A) Camera positions during hurdling from a top view perspective. Green dots represent the cameras, red lines the hurdles, and the yellow arrow the hurdling track. (B) Camera positions during tap dancing from a top view perspective. Green dots represent the cameras, the yellow square the area in which the tap dancer performed the sequence (C) Set-up during the recording of hurdling. Three hurdle transitions had to be performed during the recording. The two last hurdles are visible in the figure above. The yellow arrow indicates the hurdling track. (D) Fourteen point-light markers were used, positioned at the shoulders, elbows, wrists, hips, knees, ankles, and the tips of the toes. (E) Three snapshots of the hurdling action over the course of one video. (F) Three snapshots of the tap-dancing action over the course of one video.

After data acquisition, point-light displays were processed using Qualisys, ensuring visibility of all 14 recorded point-light markers during the entire recording time (for an overview of the position of the point-light markers, see Fig. 1). Note that we excluded videos containing movement errors, e.g. touching of a hurdle. Correspondingly, all sounds in both actions were produced by foot-ground contacts only.

Sound data were processed using Reaper v5.28 (Cockos Inc., New York, United States). Stimulus intensities, i.e. loudness of the stimuli, of hurdling and tap-dancing recordings were normalized separately. The spectral distributions of both recording types were then equalized by capturing the frequency profiles of hurdling and tap-dancing sounds using the Reaper plugin Ozone 5 (iZotope Inc, Cambridge, United States). The plugin's match function used the difference curve (hurdling – tap dancing) to adjust the tap-dancing spectrum to the hurdling reference. Point-light displays and sound were synchronized, and the subsequent videos were cut using Adobe Premiere Pro CC (Adobe Systems Software, Dublin, Ireland). The final videos had a size of 640x400 pixels, 25 frames per second, and an audio rate of 44 100 Hz. A visual fade-in and a fade-out of 1 s (25 frames) were added with Adobe Premiere. Video length ranged from 3 to 6 s, with an average length of 5 s.

For the fMRI session, a subset of 48 individual hurdling and 48 tap-dancing videos was selected for each participant. This subset included the 35 hurdling and 35 tap-dancing videos from our previous fMRI study (Heins et al. 2020b, submitted to journal), and 13 additional videos per action type, choosing the videos with the most reliable ratings from test and retest sessions (see Section 2.3.2).

For every selected video, a delayed sound version was created using Adobe Premiere. The sound was adjusted to start with a 200 ms (5 frames) delay in reference to the beginning of the video. All videos were presented using the Presentation software (Version 18.1, Neurobehavioral Systems, Inc., Berkeley, CA).

2.3 Procedure

2.3.1 Training and filming sessions

Participants were trained in hurdling and tap dancing by professional instructors for a nine-week period. They trained both action types for three hours a week, participating in two ninety-minute training sessions a week for both hurdling and tap dancing. None of the participants had any experience in hurdling and tap dancing before starting the training. Participants had to take part in four out of six offered filming sessions. The first filming sessions took place two weeks after the training started, with the following sessions taking place four, five, six, eight, and nine weeks after training start. During these sessions, participants were equipped with 14 point-light markers (see Fig. 1), which were tracked via infra-red cameras of the motion capturing system while performing both action types. The hurdling action consisted of three hurdle clearances (Fig. 1), while the tap-dancing action was a movement sequence learned during the tap-dancing training sessions. Both actions increased in difficulty within the four sessions. For hurdling, the spatial distance between the three hurdles increased, requiring more speed and a smoother hurdle clearance. For tap dancing, action elements were added to the sequence to increase difficulty.

2.3.2 Behavioral test and retest sessions

Behavioral test-retest sessions were conducted to determine the stimuli with the highest reliability of participants' rating. Sessions were conducted in a computer lab. Participants were seated in front of a computer and instructed to rate the quality of their actions on a scale from 1 (*"not well at all"*) to 6 (*"very well"*) based on their subjective impression. The experiment consisted of two blocks with self-paced responses, with both blocks lasting between 20 to 30 minutes. The same videos were presented in a different order in the two blocks of the experiment. Videos were pseudorandomized so that not more than three videos in a row showed the same action type (hurdling vs. tap dancing). Overall durations of all test sessions ranged from 40 to 60 minutes, depending on rating speed. Two weeks after the first test session, participants were presented the same videos again (in pseudo-randomized order). Forty-eight videos for both hurdling and tap dancing were chosen per participant and were used in the current fMRI session. The videos with the highest reliability in ratings were chosen. Each video was rated a total of four times (two times in the test and two times in the retest sessions). 16.52 %

of all chosen videos (672 videos in total, 96 per participant) received the same rating on all four repetitions, in 61.76 % ratings varied by a score of either +1 or -1 in one or two of the repetitions (ratings diverging in one direction), and in 21.72 % ratings varied by a score of ± 1 (ratings diverging in both directions). A subset of the videos was used in a preceding fMRI study (Heins et al., 2020).

2.3.2 fMRI session

For the fMRI session, participants completed the same task as during the test-retest sessions; namely, to rate the quality of the presented actions on a 6-point Likert scale. Since we did not intend to draw participants' attention to the delayed sounds, they were not informed about the presence of videos with delayed sound. Participants were asked to regulate the volume of the sounds before the experiment started to assure that the action sounds were audible above the scanning noises.

The experiment consisted of five blocks, including 11 trial-blocks each. Trial-blocks consisted of four to six trials of the same condition. Transition probabilities ensured that each condition was preceded by every condition (including the same condition) in the same number of trial-blocks over the whole experiment. The first trial-block of a block was a repetition of the last trial-block of the preceding block, to avoid losing a transition. The remaining ten trial-blocks consisted of two trial-blocks for each of the five conditions (namely hurdling normal, tap dancing normal, hurdling delayed, tap dancing delayed, - and null events, see supplementary figure S1 for an overview). Thus, after the discarding of the first trial-block after each pause, 50 trial-blocks, containing 240 trials, remained for the analysis. These consisted of 192 video trials (48 trials per conditions) and 48 null events, where a fixation cross was presented. The duration of the null events was 5 s. Before each trial, a fixation cross was presented as an inter-stimulus interval, varying between 3.5 to 4.5 s in length. After each video trial, the rating scale from 1 to 6, including the rating question, was presented. The experiment continued upon the participants' button press.

Throughout the entire scanning routine, participants were instructed to refrain from moving.

2.4 fMRI recordings and preprocessing

Participants were scanned in a 3-Tesla Siemens Magnetom Prisma MR tomograph (Siemens, Erlangen, Germany) using a 20-channel head coil. A 3D-multiplanar rapidly acquired gradient-echo (MPRAGE) sequence was used to obtain high resolution T1 weighted images

ahead of functional scanning, with scanning parameters set to 192 slices, a repetition time (TR) of 2130 ms, an echo time (TE) of 2.28 ms, slice thickness of 1 mm, a field of view (FoV) of 256 x 256 mm², and a flip angle of 8°.

Gradient-echo echoplanar imaging (EPI) was used to measure blood-oxygen-level-dependent (BOLD) contrast for functional imaging data of the whole brain. There were six EPI sequences in total. One sequence for the volume adjustment and one sequence for each of the five experimental blocks. Scanning parameters were set to a TE of 30 ms, a TR of 2000 ms, a flip angle of 90°, 33 slices were acquired interleaved with a slice thickness of 3 mm, and a FoV of 192x192 mm².

Imaging data were processed using SPM12 (Wellcome Trust, London, England). Slice time correction to the middle slice was performed, followed by realignment of all individual functional MR (EPI) images to correct for three-dimensional motion. The individual's structural scan was co-registered to the mean functional image and then segmented into the native space tissue components. Both the structural and the functional images were normalized into the standard MNI space (Montreal Neurological Institute, Montreal, QC, Canada). Spatial smoothing of the functional images was performed with a Gaussian kernel of full-width at half maximum (FWHM) of 8 mm. To additionally reduce effects of motion, we performed a denoising procedure on the EPI data using the default settings of the CONN toolbox in MATLAB (Whitfield-Gabrieli & Nieto-Castanon, 2012), which implements the anatomical component-based noise correction method (aCompCor). Denoising included regressing out the first five principal components associated with white matter and cerebrospinal fluid as well as the motion parameters and their temporal derivatives from the BOLD signal. A high-pass temporal filter equivalent to 128 s was applied to the data.

2.5 Statistical data analysis

2.5.1 Behavioral data analysis

We calculated Kolmogorov-Smirnov tests to assure the normal distribution of our behavioral rating score data. To examine a potential reduction in rating scores by initial sound delay in general (Hypothesis 1), we calculated 2 x 2 x 3 repeated measures analyses of variance (rmANOVA) on mean rating scores using SPSS (IBM, New York, United States). To examine adaptational effects which are reflected in restored rating scores and happen very fast (cf. Kennel et al., 2015 for performance restoration after the first delayed trial), we only included the first three trials of each trial-block in the analysis, so that the rmANOVA included the factor ACTION, with the factor levels *B* (*hurdling*) and *G* (*tap dancing*), the factor DELAY with factor

levels *undelayed* and *delayed* and the factor REPETITION with factor levels *first*, *second*, and *third presentation*. We expected a significant main effect for DELAY (Hypothesis 1). Testing for the diverging impact of sound delay on B and G actions, i.e. the stronger decrease of rating scores for delayed G sounds, and the persisting effect of this delay over several consecutive trials (Hypothesis 2), we additionally hypothesized an interaction effect ACTION x DELAY, and an interaction effect ACTION x DELAY x REPETITION.

To further examine significant interaction effects, factors were held constant and simple main effects as well as paired *t*-test were calculated. The significance level was set to $\alpha = .05$ and Bonferroni-correction was performed for multiple comparisons.

2.5.2 fMRI design specification

The design was implemented in SPM12, following a general linear model (GLM, Friston et al., 1994; Worsley & Friston, 1995) approach. The modeled activation was time-locked to the onsets of the videos for each of the experimental conditions or the null events. Epochs contained the full presentation period ranging from 3 to 6 s for the videos, and 5 s for the null events. For the delayed conditions, we added the number of repetitions within a trial-block, indicating the first, second, third ... up to sixth delayed sound video presented in a row, as parametric modulators to the respective regressor in order to examine the initial interference and adaptational processes. The GLM for every participant thus consisted of 16 regressors in total: four regressors for the experimental conditions (B normal, G normal, B delayed, G delayed), two parametric regressors modelling the number of repetitions for the two delayed conditions (repetition B delayed, repetition G delayed), one regressor for the null events, and six regressors for the motion parameters (three translations and three rotations). Activation for 48 trials was considered for the modeling of each of the four regressors for the experimental conditions and for the null event regressor. All these regressors were convolved with the hemodynamic response function.

On the first level, *t*-contrasts of the experimental conditions against null events were calculated (condition > rest). These contrast images were then used to set up a flexible factorial design on the second level. The flexible factorial design was chosen because it accounts best for the within-subject factors. The model consisted of 16 regressors – four regressors for the experimental conditions, and twelve regressors for the subject effects, one for each participant. Additionally, *t*-contrasts for the parametric regressors were calculated on the first level, and one-sample *t*-tests on the second level.

Region of interest (ROI) analyses were performed to test our hypotheses and to replicate results from our previous fMRI study (Heins et al., 2020) from which we derived functional ROIs using peak voxels of the right primary auditory cortex ($x = 54, y = -13, z = 5$), right pSTG ($x = 54, y = -31, z = 5$) and SMA ($x = -3, y = -4, z = 68$). ROIs were defined as spheres of 6 mm around these peak coordinates. To test for the hypothesized stronger sensory attenuation for tap dancing (G action) than for hurdling (B action), we performed a ROI analysis focusing on the primary auditory cortex for the B>G contrast including only trials with normal sound as sensory attenuation should only arise when sounds are undelayed and thus predictable. Then, ROI analyses for the secondary auditory cortex (pSTG) and the SMA for the G>B contrasts were performed, including both delayed and undelayed sound conditions to test a stronger top-down predictive signal for tap dancing than for hurdling (Hypothesis 3).

To test for effects of the prediction error elicited in ACC and pSTG by delayed sound videos (Hypothesis 4), we first calculated the contrast delayed>undelayed. Considering the possibility that delayed sound does not evoke a persisting prediction error, but that the error signal is strongest for the first presentation of delayed sound (cf. Kennel et al., 2015), we additionally calculated a one-sample *t*-test for the effects of parametric regressors (repetition B delayed, repetition G delayed) on the second level. This regressor was suggested to capture the transient increase and subsequent decline of activation caused by the first delayed sound trial within a trial-block.

To test Hypothesis 5, we extracted beta values from SMA to test for the adaptational shift to the predictive model evoked by the delayed sound videos, especially in the delayed G condition, and examined the expected correlation between pSTG and SMA beta values for both delayed sound conditions (Hypothesis 6). Finally, assuming that SMA provides predictive input to ensure proper performance evaluation even when sound is delayed, we tested for a positive correlation between SMA beta values and rating scores for G sounds but not B sounds (Hypothesis 7). Significance tests of the correlational analyses were performed at $\alpha = 0.05$, one-sided, based on directional hypotheses.

3 Results

3.1 Behavioral Results

A 2 x 2 x 3 rmANOVA with the factors ACTION (*B, G*), DELAY (*undelayed, delayed*) and repetition (*first, second, third presentation*) yielded no main effect for DELAY ($F(1,11) = 3.77$,

$p = .078$). Contrary to our hypothesis (Hypothesis 1), rating scores increased slightly for the delayed conditions ($M = 3.59$, $SD = 0.66$) compared to the normal sound conditions ($M = 3.45$, $SD = 0.59$). Neither the interaction ACTION \times DELAY ($F(1,11) = 0.04$, $p = .847$) nor the interaction ACTION \times DELAY \times REPETITION ($F(1,11) = 1.22$, $p = .335$) were significant. Descriptively, ratings decreased for the delayed B actions from the first ($M = 3.78$, $SD = 0.58$) to the second ($M = 3.67$, $SD = 0.61$) to the third presentation ($M = 3.52$, $SD = 0.66$), while the opposite was observed for the delayed G actions (first presentation $M = 3.41$, $SD = 1.02$, second presentation $M = 3.51$, $SD = 0.95$, third presentation $M = 3.65$, $SD = 1.03$).

3.2 fMRI results

With regard to our first functional hypothesis (Hypothesis 3), the whole-brain contrast G>B revealed activation in the SMA ($x = 0$, $y = -4$, $z = 68$) and pSTG ($x = 48$, $y = -34$, $z = 5$), as expected. Expected stronger A1 activation for B vs. G action sounds was not revealed after FDR-correction at $p < .05$ when restricted to normal action sounds, but became significant when contrasting all B vs. G actions (Fig. 2). Moreover, beta values extracted from A1 differed significantly between the B ($M = 2.32$, $SD = 1.18$) and the G condition ($M = 1.11$, $SD = 1.01$), $t(11) = 2.79$, $p = .018$), determined by a paired sample t -test between B and G with normal sound.

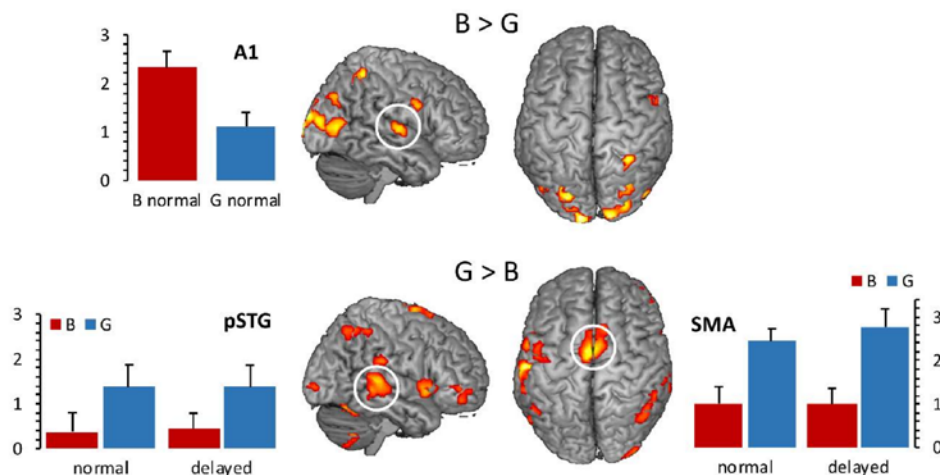


Fig. 2. Main effect of ACTION. FDR-corrected t -maps ($p < .05$) for the B>G contrast are presented in the top row. FDR-corrected t -maps ($p < .05$) for the G>B contrast are presented in the bottom row. Additional beta values for the primary auditory cortex, the pSTG and the SMA are presented, with mean beta values for the B conditions presented in red, and mean beta values for the G conditions presented in blue.

To test for the expected error response by the presentation of conditions with delayed sounds (Hypothesis 4), the whole-brain contrast delayed>undelayed, did not reveal significant effects. In contrast, the conjoined effect of the parametric regressors for delayed B and G sounds, emphasizing the response to the first delayed sound video in a row, revealed activation in pSTG, the intraparietal sulcus and the posterior cingulate cortex (pCC, uncorrected, $p < .001$, Fig. 3).

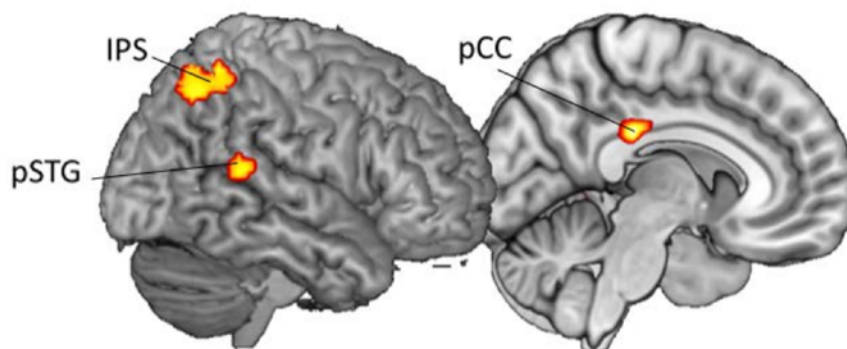


Fig. 3. Initial effect of delayed sound. Uncorrected t-maps ($p < .001$) for the parametric effect of repetition for the delayed conditions. Activations are evoked by the initial vs. repeated presentation of videos with delayed sound (both for B and G actions).

Beta values extracted from SMA showed a trend for the interaction effect ACTION x SOUND ($F(1,11) = 3.78$, $p = .078$). Descriptively, the adaptational effect in SMA was slightly stronger for G than for B actions, as expected (Hypothesis 5), albeit post-hoc *t*-tests were only approaching significance ($p > .10$).

Testing for a positive correlation of beta scores in SMA and pSTG for the delayed G and B conditions (Hypothesis 6), we found a non-significant trending correlation for B sounds ($r = .467$, $p = .063$) and for G sounds ($r = .390$, $p = .105$), but no correlations for undelayed B sounds ($r = .221$, $p = .245$) or undelayed G sounds ($r = .164$, $p = .306$). Finally, we calculated correlations between the SMA beta values and rating scores for the delayed B and G conditions (Hypothesis 7). Here, we did not find a significant correlation for delayed B sounds ($r = -.363$, $p = .123$) as expected, but a non-significant trend for delayed G sounds ($r = .412$, $p = .092$).

4 Discussion

It strikes us when a sound differs from what we expected – especially when it is generated by our own actions. Starting from the assumption that this surprise is larger when we actively intend to create a sound, compared to when the sound is an incidental by-product of our performed action, we here used fMRI to examine how the performance rating and the neuronal processing of one's own action videos are affected when the sound playback is systematically delayed. Videos were recorded from our participants during a nine-week training schedule of two whole body actions – hurdling, operationalizing incidental action sounds (B sounds), and tap dancing, operationalizing intentional action sounds (G sounds). While we found an error response to the delayed playback of both types of action sounds, higher cortical areas, especially the SMA, were more strongly involved in the processing of delayed G sounds and their behavioral performance evaluation, as discussed below.

Comparing rating scores for actions presented with correctly timed or 200 ms delayed sound, we had two hypotheses: first, that rating scores would decrease, i.e. action performance is perceived as worse, for the delayed sound conditions (Hypothesis 1) and second, that this decrease would be especially prominent and persistent for G actions, as action sounds were expected to be more important for the rating of G actions (Hypothesis 2). Surprisingly, we observed a trend in the other direction – rating scores for the delayed sound conditions tended to be higher than for the undelayed sound conditions. Examining the delayed conditions only, we additionally observed an interaction effect that was reflected in increasing rating scores from the first to the third presentation of delayed stimuli for G sounds, while rating scores decreases for B sounds. This finding contrasts with others, as the same delay length caused at least an initial performance disruption in hurdling performance (Kennel et al., 2015) and a decrease in sense of agency during walking (Menzer et al., 2010). It is possible that delayed auditory feedback has a less pronounced effect when presented offline in a video presentation setting rather than during real performance, as in these two studies. Studies examining audiovisual asynchrony showed that a small audio-lag results in higher perceived synchrony ratings than synchronous presentation (Eg & Behne, 2015; Vatakis & Spence, 2006). Thus, the delayed conditions were possibly perceived as more synchronous and therefore rated slightly more positive. The employment of a delayed sound condition was not mentioned during participants' instruction, and when asked if they noticed anything in a post-fMRI survey, only two of the twelve participants reported to have perceived a delay in some of the videos. Although the slight increase of rating scores for G actions over repeated presentation of delayed sounds opposed to the decrease for B actions was not expected, it could reflect a shift away from the actual auditory

feedback to a reliance on top-down predictive information, which is supported by our neuronal findings. In contrast, assuming that auditory feedback is not as relevant for B actions as for G actions, the time lag may remain rather unnoticed for a sequence of delayed B sound videos before rating scores begin to decrease.

Replicating a previous study (Heins et al., 2020), we observed stronger activation of SMA and pSTG in the tap-dancing condition and higher activation – i.e. less attenuation – in primary auditory cortex in the hurdling condition (Hypothesis 3). We take this result to strengthen the assumption that the brain does not engage in prediction of B sounds as much as G sounds. This interpretation is also supported by the observation that SMA and pSTG provide stronger top-down modulations of primary auditory cortex in case of action sounds produced by voluntary actions (Reznik, Simon, & Mukamel, 2018). An alternative explanation is that B sounds may be simply less *predictable* (not less predicted per se) than G sounds, as a sound's predictability has been found to co-vary with attenuation in primary auditory cortex (Straube et al., 2017). However, we take this alternative account to be less convincing as our participants had intensively trained both B and G actions, they had rated all of the videos two times in the test re-test sessions, and therefore could be considered being quite familiar with both types of stimuli.

As also hypothesized, we found pSTG to provide a significant error signal to the first presentations of the delayed conditions that adapted over the course of repeated presentations in the trial-blocks (Hypothesis 4). This is in line with studies regarding pSTG as an auditory error detector responding to altered feedback (Fu et al., 2006; Johnson, Belyk, Schwartz, Pinheiro, & Kotz, 2019; Zheng, Munhall, & Johnsrude, 2010). While we had expected the same response pattern in ACC, reflecting an error signal in action monitoring (Van Veen & Carter, 2002), it was actually present in the *posterior* cingulate cortex (pCC). While the entire cingulate cortex is engaged in the control of action, relating emotion, action, and memory, posterior and anterior cingulate differ with regard to their specific contribution. In contrast to ACC which is more related to limbic aspects such as reward and action outcome, pCC seems more related to memory and learning (Rolls, 2019). The pCC receives visuospatial and somatosensory action-related information from the dorsal stream, including parietal cortex and hippocampus, that probably serve action–outcome learning (Rolls, 2019). Moreover, pCC is also engaged in tasks that require self-imagery (Johnson et al., 2002; Kircher et al., 2002, 2000; Sugiura et al., 2005) and self-referential processing (Northoff et al., 2006). Interestingly, pCC was found to be particularly engaged when self-generated sensory consequences are temporally deviant (cf. Straube et al., 2017). Against this backdrop, we speculate that pCC provides a specific error

response to the delayed conditions in our data, reflecting deviation from the learned auditory effects of participants' hurdling and tap dancing.

We found SMA beta values to be specifically higher for the delayed G sounds, according to Hypothesis 5. SMA hence seems not immediately involved in detecting altered feedback, in contrast to pSTG and pCC, but rather provides additional top-down information about the intended action outcome whenever perceived outcome differs from the expected. Being critical for audio-motor associations (Johnson et al., 2019; Lima et al., 2016), the SMA might get engaged when auditory delays persist over a longer time to modify audio-motor models or to amplify other modalities for motor control (compare Heins et al. 2020b, submitted to journal). This also fits to the increasing of the rating scores over the course of repeated presentations of sound delayed videos in the trial-blocks. Accordingly, pSTG and SMA beta values also tended to be correlated in the delayed conditions, but not in the undelayed conditions (Hypothesis 6). Implementing a hierarchical network for action sound processing and model adjustment (cf. Heilbron & Chait, 2018), pSTG provides an error signal caused by altered auditory feedback that is conveyed to the SMA to adjust audio-motor models. In the present study, this process only manifested in case of G sounds, where rating scores positively correlated with SMA beta values in the delayed condition, whereas no such correlation was found for delayed B sounds (Hypothesis 7). This mirrors the results of our previous study, finding that SMA beta values correlated with rating scores when sounds were deprived (Heins et al., 2020b, submitted to journal). The SMA thus seems to intervene especially when the intended action sounds are somewhat disrupted, and accurate action sound information has to be restored from the model to perform the action evaluation task reliably.

Taken together, our study emphasizes the role of SMA in audio-motor processing. SMA seemingly is the apex of the action sound prediction hierarchy, intervening to resolve surprise elicited by altered auditory feedback that pSTG on a hierarchical midlevel cannot resolve. This seems to be especially true for intentional action sounds, which are particularly important for a proper action evaluation.

5 Conclusion

Sounds created by our own actions, though omnipresent in our everyday life, have been scarcely examined in an ecologically valid fashion. Providing participants with delayed auditory feedback when watching their own tap dancing and hurdling performance, we found behavioral

and fMRI evidence for the intuitive difference between action sounds which are the intended goal of our actions and action sounds created as a by-product of performed actions. In contrast to the latter, brain responses revealed increased predictive engagement for evaluating actions with intended action sounds to cope with disrupted auditory feedback. Future research may focus on a better understanding of effects of delayed auditory feedback by systematically testing when and how time matters in brain function and behavior creating a pathway for unsolved problems such as schizophrenia, stuttering or imitating others – it is never too late.

Acknowledgements

We would like to thank Monika Mertens for her help during data collection. Furthermore, we thank Theresa Eckes, Alina Eisele, Marie Kleinbielen, Katharina Thiel and Stefan Vinbrüx for their assistance during filming and with creating the stimulus material, and Daniel Kluger for helping with the processing of the auditory stimuli. Finally, we would like to thank Amelie Huebner, Jennifer Pomp, Marlen Roehe and Lena Schliephake for rewarding discussions. The project was funded by the Deutsche Forschungsgemeinschaft (SCHU 1439-4-2).

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience*, *21*(4), 791–802.
<https://doi.org/10.1162/jocn.2009.21055>
- Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology*, *48*(9), 1276–1283. <https://doi.org/10.1111/j.1469-8986.2011.01196.x>
- Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *J Acoust Soc Am*, *119*(4), 2363–2371. https://doi.org/10.1007/978-1-4939-3274-0_14
- Eg, R., & Behne, D. M. (2015). Perceived synchrony for realistic and dynamic audiovisual events. *Frontiers in Psychology*, *6*(JUN). <https://doi.org/10.3389/fpsyg.2015.00736>
- Finney, S. A. (1997). Auditory Feedback and Musical Keyboard Performance. *Music Perception: An Interdisciplinary Journal*, *15*(2), 153–174.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836.
<https://doi.org/10.1098/rstb.2005.1622>
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. -P, Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*(4), 189–210. <https://doi.org/10.1002/hbm.460020402>
- Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro, E., Andrew, C. M., ... McGuire, P. K. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex*, *16*(7), 969–977.
<https://doi.org/10.1093/cercor/bhj039>
- Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, *389*, 54–73.
<https://doi.org/10.1016/j.neuroscience.2017.07.061>
- Heins, N., Pomp, J., Kluger, D. S., Trempler, I., Zentgraf, K., Raab, M., & Schubotz, R. I. (2020). Incidental or Intentional? Different Brain Responses to One's Own Action Sounds in Hurdling vs. Tap Dancing. *14*(May), 1–11.

- <https://doi.org/10.3389/fnins.2020.00483>
- Howell, P. (2004). Effects of delayed auditory feedback and frequency-shifted feedback on speech control and some potentials for future development of prosthetic aids for stammering. *Stammering Research : An on-Line Journal Published by the British Stammering Association*, 1(1), 31–46. <https://doi.org/10.1126/science.1182395>. Evolution
- Jo, H. G., Habel, U., & Schmidt, S. (2019). Role of the supplementary motor area in auditory sensory attenuation. *Brain Structure and Function*, (0123456789). <https://doi.org/10.1007/s00429-019-01920-x>
- Johnson, J. F., Belyk, M., Schwartze, M., Pinheiro, A. P., & Kotz, S. A. (2019). The role of the cerebellum in adaptation: ALE meta-analyses on sensory feedback error. *Human Brain Mapping*, 40(13), 3966–3981. <https://doi.org/10.1002/hbm.24681>
- Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., & Prigatano, G. P. (2002). Neural correlates of self-reflection. *Brain*, 125(8), 1808–1814. <https://doi.org/10.1093/brain/awf181>
- Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research*, 190(3), 279–287. <https://doi.org/10.1007/s00221-008-1473-y>
- Kennel, C., Streese, L., Pizzera, A., Justen, C., Hohmann, T., & Raab, M. (2015). Auditory reafferences: The influence of real-time feedback on movement control. *Frontiers in Psychology*, 6(JAN), 1–6. <https://doi.org/10.3389/fpsyg.2015.00069>
- Keough, D., & Jones, J. A. (2009). The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing. *The Journal of the Acoustical Society of America*, 126(2), 837–846. <https://doi.org/10.1121/1.3158600>
- Kircher, T. T. J., Brammer, M., Bullmore, E., Simmons, A., Bartels, M., & David, A. S. (2002). The neural correlates of intentional and incidental self processing. *Neuropsychologia*, 40(6), 683–692. [https://doi.org/10.1016/S0028-3932\(01\)00138-5](https://doi.org/10.1016/S0028-3932(01)00138-5)
- Kircher, T. T. J., Senior, C., Phillips, M. L., Benson, P. J., Bullmore, E. T., Brammer, M., ... David, A. S. (2000). Towards a functional neuroanatomy of self processing: effects of faces and words. *Cognition*, 10, 133–144. Retrieved from <papers3://publication/uuid/D04A7C9E-4543-43F3-8EB6-97E3070FE564>
- Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of Supplementary Motor Areas in Auditory Processing and Auditory Imagery. *Trends in Neurosciences*, 39(8), 527–542. <https://doi.org/10.1016/j.tins.2016.06.003>
- Martikainen, M. H., Kaneko, K. I., & Hari, R. (2005). Suppressed responses to self-triggered

- sounds in the human auditory cortex. *Cerebral Cortex*, 15(3), 299–302.
<https://doi.org/10.1093/cercor/bhh131>
- 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(3), 184–192.
<https://doi.org/10.1080/17588921003743581>
- Northoff, G., Heinzel, A., de Greck, M., Bormpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain-A meta-analysis of imaging studies on the self. *NeuroImage*, 31(1), 440–457. <https://doi.org/10.1016/j.neuroimage.2005.12.002>
- Oldfield, R. C. (1971). The assessment and analysis of handedness. *Neuropsychologia*, 9(1), 97–113.
- Pfordresher, P. Q. (2006). Coordination of perception and action in music performance. *Advances in Cognitive Psychology*, 2(2), 183–198. <https://doi.org/10.2478/v10053-008-0054-8>
- Pfordresher, P. Q., & Beasley, R. T. E. (2014). Making and monitoring errors based on altered auditory feedback. *Frontiers in Psychology*, 5(AUG), 1–13.
<https://doi.org/10.3389/fpsyg.2014.00914>
- Purcell, D. W., & Munhall, K. G. (2006). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, 119(4), 2288–2297. <https://doi.org/10.1121/1.2173514>
- Rauschecker, J. P. (2012). *Neural Correlates of Auditory Cognition* (Vol. 45).
<https://doi.org/10.1007/978-1-4614-2350-8>
- Reznik, D., Simon, S., & Mukamel, R. (2018). Predicted sensory consequences of voluntary actions modulate amplitude of preceding readiness potentials. *Neuropsychologia*, 119.
<https://doi.org/10.1016/j.neuropsychologia.2018.08.028>
- Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science (New York, N.Y.)*, 306(5695), 443–447. <https://doi.org/10.1126/science.1100301>
- Rolls, E. T. (2019). The cingulate cortex and limbic systems for emotion, action, and memory. *Brain Structure and Function*, 224(9), 3001–3018. <https://doi.org/10.1007/s00429-019-01945-2>
- Sasisekaran, J. (2012). Effects of delayed auditory feedback on speech kinematics in fluent speakers. *Perceptual and Motor Skills*, 115(3), 845–864.
<https://doi.org/10.2466/15.22.PMS.115.6.845-864>

- Schubotz, R. I., & Von Cramon, D. Y. (2003). Functional-anatomical concepts of human premotor cortex: Evidence from fMRI and PET studies. *NeuroImage*, *20*(SUPPL. 1), 120–131. <https://doi.org/10.1016/j.neuroimage.2003.09.014>
- Straube, B., Van Kemenade, B. M., Arikani, B. E., Fiehler, K., Leube, D. T., Harris, L. R., & Kircher, T. (2017). Predicting the multisensory consequences of one's own action: Bold suppression in auditory and visual cortices. *PLoS ONE*, *12*(1), 1–25. <https://doi.org/10.1371/journal.pone.0169131>
- Sugiura, M., Watanabe, J., Maeda, Y., Matsue, Y., Fukuda, H., & Kawashima, R. (2005). Cortical mechanisms of visual self-recognition. *NeuroImage*, *24*(1), 143–149. <https://doi.org/10.1016/j.neuroimage.2004.07.063>
- Timm, J., Schönwiesner, M., SanMiguel, I., & Schröger, E. (2014). Sensation of agency and perception of temporal order. *Consciousness and Cognition*, *23*(1), 42–52. <https://doi.org/10.1016/j.concog.2013.11.002>
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443. <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- van der Steen, M. C., Molendijk, E. B. D., Altenmüller, E., & Furuya, S. (2014). Expert pianists do not listen: The expertise-dependent influence of temporal perturbation on the production of sequential movements. *Neuroscience*. <https://doi.org/10.1016/j.neuroscience.2014.03.058>
- Van Veen, V., & Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience*, *14*(4), 593–602. <https://doi.org/10.1162/08989290260045837>
- Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neuroscience Letters*, *393*(1), 40–44. <https://doi.org/10.1016/j.neulet.2005.09.032>
- Vroomen, J., Keetels, M., De Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, *22*(1), 32–35. <https://doi.org/10.1016/j.cogbrainres.2004.07.003>
- Waszak, F., Cardoso-Leite, P., & Hughes, G. (2012). Action effect anticipation: Neurophysiological basis and functional consequences. *Neuroscience and Biobehavioral Reviews*, *36*(2), 943–959. <https://doi.org/10.1016/j.neubiorev.2011.11.004>
- Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn: A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks. *Brain Connectivity*, *2*(3),

125–141. <https://doi.org/10.1089/brain.2012.0073>

Woods, E. A., Hernandez, A. E., Wagner, V. E., & Beilock, S. L. (2014). Expert athletes activate somatosensory and motor planning regions of the brain when passively listening to familiar sports sounds. *Brain and Cognition*, *87*(1), 122–133.

<https://doi.org/10.1016/j.bandc.2014.03.007>

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited — Again.

NeuroImage. <https://doi.org/10.1006/nimg.1995.1023>

Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, *22*(8), 1770–1781.

<https://doi.org/10.1162/jocn.2009.21324>

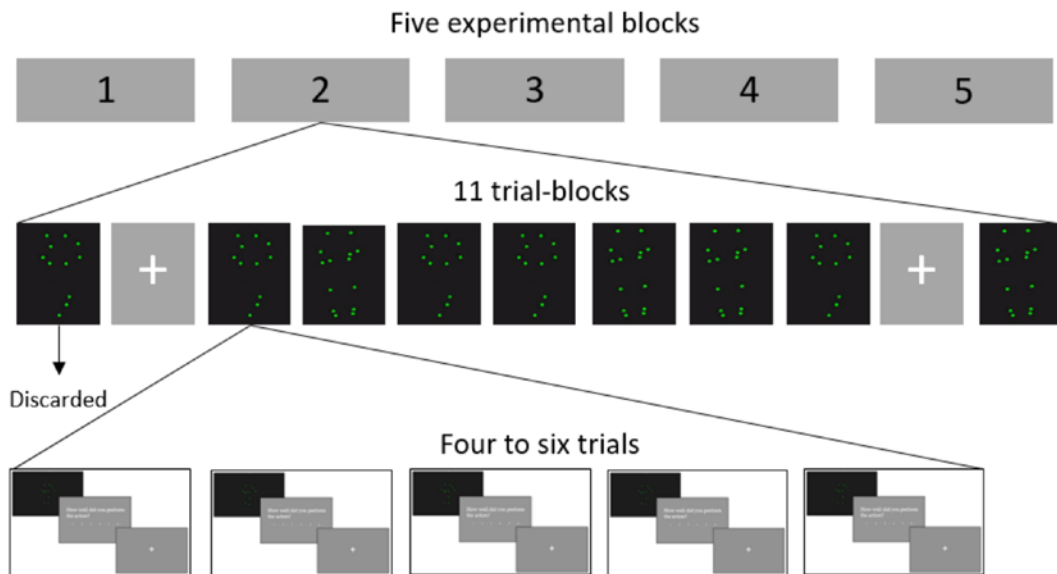


Fig. S1. Experimental procedure. The experiment consisted of five blocks in total. Each block itself consisted of 11 trial-blocks. The first trial-block was discarded for the analysis. The remaining ten trial-blocks were two trial-blocks of the five conditions (B normal, G normal, B delayed, G delayed, null events) each. Trial-blocks contained four to six trials of the same condition, with each trial consisting of the video, the rating question and a fixation-cross as interstimulus interval.

4. General Discussion and Future Directions

4.1 Summary of the presented studies

The presented studies were conducted with the goal to shed light into the scarcely studied field of real-life action sounds, their relevance for action performance and evaluation, and the neuronal representation thereof. Most importantly, a new distinction between different types of action sounds is postulated and examined systematically for the first time. Namely, the difference between action sounds that are the intentional goal of our actions (G action sounds), and action sounds that are produced as a mere incidental by-product (B action sounds).

We trained participants in two sound producing actions – hurdling, a B action, where sound is created incidentally, and tap dancing, a G action, where sound is an intended action goal. To examine both behavioural and neuronal measures, we chose an action observation design, where we presented participants with videos of their own actions and corresponding action sounds, to reactivate neuronal networks usually involved during action execution. In our three experiments we applied different perturbations to the stimuli, to investigate the influence of (1) less informative (scrambled) action sounds, (2) missing action sounds, and (3) delayed action sounds on behavioural rating scores and the underlying neural networks.

In **Study I**, we were able to find indicators that there are indeed differences between the two action sounds subclasses, both in regards to their behavioural relevance and their neuronal underpinnings. For instance, auditory “scrambling” solidified in a stronger reduction of behavioural rating scores in G vs. B actions. G action sounds additionally yielded more pronounced anticipation, as reflected by enhanced sensory attenuation (i.e. less activity in the primary auditory cortex) and stronger activity higher up in the action sound processing hierarchy, i.e. in pSTG and SMA. Furthermore, pSTG revealed a small, albeit not significant

difference in activity for scrambled G sounds compared to scrambled B sounds, which might be an indicator for a more pronounced prediction error when G sounds are less informative.

Building on these results, **Study II** examined the influence of the complete omission of both B and G sounds, by presenting mute videos in between videos with sounds. While we were able to replicate the stronger attenuation for G sounds in primary auditory cortex, as well as the overall higher activation of pSTG and SMA, results of Study II especially reinforced the assumption that SMA seems to be the apex of the hierarchy of action sounds. It provides important predictive information when sound is missing in G actions. This information is seemingly used to restore behavioural rating scores, which are reduced when G sounds are omitted, whereas no such effect was found for B sounds.

Results of **Study III** strengthened the indication that SMA intervenes actively in action evaluation by supplying predictive input whenever the actual sensory input is disrupted, i.e. in this case, when sounds are temporally delayed. The behavioural rating task is influenced mainly by the predictive information and moved away from actual sensory sources especially in G sounds. Study III additionally indicates pSTG as an error detector, being activated mostly by first presentations of delayed action sounds. PSTG seems to transmit this error signal to SMA for both B and G sounds, but only for G sounds is SMA able to resolve the problem and restore rating scores.

The studies presented in this thesis are some of the few studies examining the neural representation of complex human action sounds apart from language and music, and the first to postulate a behavioural and functional distinction between action sounds created intentionally and incidentally. While our results heavily indicate that there is indeed a difference in the relevance of action sounds tied to their intentionality, it is important to evaluate the limitations of our study to motivate continuing research (Section 4.2) and to embed our findings in a bigger context. First, I discuss the possibility that a continuum might be a more suitable account for

explaining the relevance of human action sounds, in contrast to the strict dichotomy of B and G sounds that we postulated in our studies (Section 4.3). Additionally, I discuss the interactions of sound with other sensory modalities, to understand the multisensory interplay in action sound perception and the possible compensatory effects of visual and/or proprioceptive information (Section 4.4). Then, I critically examine the influence that attention might have on the predictive processes that we investigated (Section 4.5). Some practical applications of our results are laid out (Section 4.6), and further research is motivated at the end of each section, before a conclusion is reached (Section 5).

4.2 Limitations of our findings

Studying more ecological valid stimuli is always challenging, as they are highly complex and less controllable. Nevertheless, moving on from research that uses button presses and simple tones in laboratory settings is crucial to broaden our understanding. This can only be accomplished by examining action sounds that we actually create in real life.

In our studies, we did our best to make the presented action sounds – hurdling and tap-dancing sounds – as similar as possible, so that we truly examined the difference between the intentionality of sound production. At the same time, we still wanted to maintain their ecological validity. Both tap dancing and hurdling are whole-body movements with the feet as the sound effectors, and we additionally adjusted their sound spectra to make them more similar. However, stimuli still differed to some extent in their event density (i.e. the number of elements presented per second) and their rhythmical structure. We controlled for the event density in our first experiment. Finding no effects, we excluded this parameter from our following analyses. Still, event density might be one of the factors why audio-visual asynchrony detection collapses (Fujisaki & Nishida, 2005), and should be controlled for in future studies. In terms of controllability, it would have been better to mirror the rhythm and event density of hurdling in

the learned tap-dancing sequences, to eliminate any chance of a confound. This might however reduce ecological validity in the tap-dancing stimuli.

In order to examine real-life action sounds in an fMRI setting, we had to refrain to an observation paradigm, as participants were not able to perform hurdling and tap dancing in an MRI (magnetic resonance imaging) scanner. It is well established that similar predictive processes are at work during observation and execution of action (e.g. Sato, 2008), and we further enhanced this connection by presenting participants with their own actions and corresponding action sounds only. Still, we cannot rule out that results would be different if actions were actually executed. Note that during the execution of G actions, a temporal delay led to a disrupted performance (e.g. Christoffels et al., 2011), while this was not the case during the execution of B actions (e.g. Kennel et al., 2015). Contrary to that, audio-visual asynchrony is detected later, i.e. at larger temporal delays, in G actions compared to B actions (Eg & Behne, 2015; Vatakis & Spence, 2006). This could indicate that delays are immediately disruptive for action execution of G actions, but that the focus is shifted away from the actual sensory input and more weight is given to the predictive component during their observation. This matches the results of Study III, where rating scores for G actions increased slightly with repeated delay presentation, accompanied by enhanced SMA activity. SMA seemingly performs the shift to a more prediction-based evaluation approach. This is not only the case for evaluation of G actions with delayed action sounds (Study III), but also for evaluation of G actions with omitted sounds (Study II). Still, the possible divergence of effects during execution and observation should definitely be kept in mind for future studies.

In our experiments, we used the same number of trials for all conditions, both the “standard” stimuli (normal sound and picture) and the “deviant” stimuli, where a disruption was applied (“scrambling” in Study I, omission in Study II, delay in Study III). While this is optimal from a data analysis point of view, it led to participants seeing and hearing the “deviant”

stimuli the same number of times (Study III) or variants of the deviants even more frequently than the standard stimulus (Study I & Study II). It is unclear whether the participants expected the “deviant” stimuli to occur, which might in turn have led to a hampered error response, resulting from an expected change (cf. Bendixen et al., 2012, who suggest to compare unexpected repetitions with expected changes to optimally examine error responses). Transitional probabilities were balanced within the experiments, so that every condition was equally likely to occur after every other condition. This means participants were not able to make exact identity predictions about the next stimulus on a trial-by-trial basis, but the overall number of occurring “deviant” trials might still have dampened the error response. Note however that although the occurrence of “deviant” trials was known, participants’ rating scores for G actions showed an evident reduction for those trials in Study I & II, while this was the case to a lesser extent (Study I) or not at all (Study II) for B action “deviant” trials. This indicates that even expected manipulations significantly impeded performance evaluation. However, the exact number of “standard” and “deviant” stimuli and their predictability should be considered when planning future studies.

Furthermore, we only used correlational measures to investigate the proposed hierarchy of action sounds, involving the primary auditory cortex, pSTG and SMA. To date there is only one study examining the interactions of these brain regions during action sound production using a dynamic causal modelling approach (Jo et al., 2019). While the results of this study are in accordance with our own results, deeming a prediction signal in SMA necessary to explain the sensory attenuation of self-generated sounds, it once again used simple tones elicited by button presses as action sounds. It is thus still necessary to examine the interplay between (pre-)motor and auditory regions during real-life action sound production and perception, using a network rather than a correlational approach.

The main limitation of our studies is surely the small number of participants ($n = 13$ in Study I, $n = 12$ in Study II and III). This is due to the intricate and time-consuming procedure of the studies. The study involved five instructors for hurdling and tap-dancing training spanning over nine weeks. Ten people helped with the filming protocol (conducted in 12-hour shifts on seven days, recording about 240 videos from each of the originally 19 participants), and more than fifteen people helped to process the stimulus material, tailored specifically for every participant, which took about 70 hours per participant. Including the test-retest sessions and the three fMRI sessions, the study spanned over a time frame of approximately two years, in which, unfortunately, a considerable number of participants dropped out. While some studies examining action sounds had comparably few participants (cf. Menzer et al., 2010; Reznik, Ossmy, & Mukamel, 2015), and we were able to replicate our main findings in all three studies, this is undeniably a constraint to our results. To further illuminate this important research question, replications with larger groups are mandatory.

4.3 Incidental or intentional – dichotomy or continuum?

Although it seems to be a human desire to sort things into clear, distinct categories, almost everything in real life is distributed on a continuum rather than categorical. This holds true even for gender/sex (Ainsworth, 2015), although this has been one of the top examples for a dichotomous distribution. This is to say, dividing action sounds into two distinct subclasses has its shortcomings when aiming for an optimal representation of the underlying processes. Our used action sounds, hurdling and tap-dancing sounds, might be more accurately described as two points on a continuum, ranging from language, where the conducted action has the sole purpose to generate the outgoing sounds, to simple everyday action sounds, e.g. the rustling of our clothes when we unintendedly brush against something. To fully understand the relevance of different action sounds, the whole spectrum has to be studied. Our results

contribute to that, showing that action sounds which are more inclined to the intentional side are more attenuated, overall more involved in predictive processes, and more relevant for action evaluation. However, even our more incidental action sounds were rated more negatively when auditory feedback was disrupted (see Study I), indicating that they have at least some relevance for performance evaluation. All sport sounds seem to fall on that point on the continuum (for a review, see Schaffert et al., 2019). While subjectively, we still would not deem them our action goal, studies show that action sounds can be helpful for action improvement and experienced athletes are able to extract highly specific information from action sounds. For example, hammer throwing performances improved when auditory feedback was used (Agostini, Righi, Galmonte, & Bruno, 2004), as well as the performance in hurdling (Pizzera, Hohmann, Streese, Habbig, & Raab, 2017). Athletes are able to anticipate relevant sports-specific information using action sounds in the sports they are experts in. This is the case in several different types of sports, including football (Sors et al., 2018a), volleyball (Sors et al., 2018b), basketball (Camponogara, Rodger, Craig, & Cesari, 2017), fencing (Allerdissen, Guldenpenning, Schack, & Bläsing, 2017), skateboarding (Cesari, Camponogara, Papetti, Rocchesso, & Fontana, 2014) and table tennis (Bischoff et al., 2014). While studies regarding rowing (Schaffert et al., 2020) and tennis (Takeuchi, 1993) have shown that missing auditory feedback can negatively affect performance, most sports studies imply that action sounds can have facilitating effects when available, but are not relevant for action monitoring in the same way as in e.g. language. This is also indicated in a study examining the effects of auditory feedback on rhythmic tapping – accurate auditory feedback helps to improve performance, but delayed or masked auditory feedback does not impair it (Van Vugt & Tillmann, 2015). In line with this is the rich research on the benefits of movement sonification, i.e. the translation of movement parameters into sounds, which are then provided as a real-time auditory feedback (for an overview, see e.g. Effenberg, Fehse, Schmitz, Krueger, & Mechling, 2016; Pizzera & Hohmann, 2015). Artificial

auditory feedback can help improve performance but is obviously not necessary to perform an action correctly.

Overall, this shows that action sounds that we deem incidental can still help us in both perception and execution but are more a “nice to have” feature than an absolute necessity. This quality definitely ranges on a continuum – while we still might get some valuable information from purely incidental action sounds, like information about the applied force when placing a glass on a table, there are other action sounds that we could easily discard completely, like the sound of chewing or the aforementioned rustling of clothes during locomotion.

The necessity of intentional action sounds, like language and music, varies as well. Disruptive feedback influences language production exceedingly negatively, while effects for music production have been rendered less decisive (cf. Section 1.4.1 and Section 1.4.2). Nevertheless, we would refer to both of these action sounds as intentional.

To evolve a more distinguished understanding of the relevance of action sounds and their neural underpinning, it is mandatory to keep this continuum in mind, and to plan future studies accordingly.

4.4 Auditory, visual, multisensory – relevance of the different senses for action

Sensory outcomes of our actions in real life are rarely constricted to one modality only. Even as simple an act as placing a glass on a table is accompanied by multiple percepts. Considering we do not apply the visual sense, because we are able to perform the action without looking where exactly we place the glass (which could either work or end in a disaster), we still receive the following sensory inputs: 1) proprioceptive information about our limb position, 2) haptic feedback about the contact with the surface and 3) auditory feedback that the glass

creates upon contact. Contemplating this, we have to consider which effects other modalities might have had on our results, and whether rather than predicting action sounds alone, we always predict all multisensory outcomes of our actions (Krala et al., 2019; Straube et al., 2017; van Kemenade, Arikan, Kircher, & Straube, 2016). Although self-recognition is possible based on sounds alone (see e.g. Kennel, Hohmann, & Raab, 2014), our participants were always provided with the visual presentation of their actions to maximize self-recognition, and to observe effects of omission and delay of action sounds relative to the action onset. Additionally, all our participants conducted the presented actions multiple times during the course of the training and the filming sessions, so that they had proprioceptive knowledge of the actions. Especially the visual modality may have had a significant impact on our results, because visual information of some degree (visual “scrambling” was applied on some trials in Study I & II) was available on every trial. While all actions have multisensory consequences, I argue here that for some of them, the visual modality is more important, whereas the auditory modality is more relevant in others, and that this might fall in line to some degree with the difference between incidental vs. intentional action sounds.

In all of our studies, we observed a stronger activation of visual cortices for B compared to G actions. This included the middle occipital gyrus, which has been found to be involved in spatial vs. non-spatial activities (cf. Renier et al., 2010), and an area most likely the occipital place area, activated when navigating through an environment (cf. Persichetti & Dilks, 2016). Both of these areas could be specifically tied to our chosen B action – hurdling – as participants had to run the hurdling track (approximately 20 m) opposed to the locally stationary tap-dancing action. However, integrating these results with our behavioural findings, it is likely that the visual domain is indeed dominant for actions that produce incidental action sounds. Rating scores decreased more in B vs. G actions when the visual modality was uninformative, i.e. scrambled (non-significant trend, Study I) and rating scores even increased when B actions

were evaluated on visual input alone (Study II). This fits the results that masked or delayed auditory feedback did not influence hurdling performance (Kennel et al., 2015). For G actions, the effects revealed a different pattern. Rating scores decreased more when the auditory modality was less informative (Study I), and evaluation on the visual domain alone resulted in lower rating scores, only somewhat restored when SMA was more active (Study II). Correspondingly, another study using tap-dancing stimuli found the auditory modality to be more informative than the visual modality (Murgia et al., 2017). In addition, a study examining effects of delayed auditory and delayed visual feedback on speech production found that delayed visual feedback was only disruptive when auditory feedback was already delayed (Chesters, Baghai-Ravary, & Möttönen, 2015). These results could be indicative of differential relevance of visual and auditory information for B and G actions respectively. Note, however, that a study on sequence production on a keyboard did not find differential effects of delayed auditory and delayed visual feedback (Kulpa & Pfordresher, 2013), although it should classify as a G action and the auditory domain should be more relevant when in accordance with the aforementioned results.

In Study II, a switch to a more visual evaluation occurred in both B and G actions as soon as the sound was missing, reflected in stronger activity of area hMT+. This was concurrent with an increase in rating scores only for B actions. Rating scores for G actions presented with no sound were somewhat restored by SMA activity. This relationship vanished when the visual modality was “scrambled”, indicating that SMA is only able to retrieve predictive input for an action when some valid information is obtainable. This is in line with SMA storing information about action outcomes independent of modality (Schubotz & Von Cramon, 2003), and being activated for visual, haptic and auditory beat perception (Araneda, Renier, Ebner-Karestinios, Dricot, & De Volder, 2017).

In our current design, we are unable to determine whether SMA provides specific auditory predictions in case of a disruption of action sounds, or whether it compensates for the lack of reliable auditory feedback through other modalities. Based on the notion that visual knowledge about what an action should look like might not be enough to activate these sensorimotor representations, but that a motoric representation is required (cf. Candidi et al., 2014), it would be especially interesting to implement our paradigm for participants who have not had any previous training in hurdling and tap dancing.

Overall, we can conclude that action sounds are not usually a unimodal occurrence, but part of an interplay between different senses. It would certainly be experimentally interesting to separate these modalities and examine the effects of action sounds alone. Nonetheless, it is the more ecologically valid approach to include other modalities in the study, and especially proprioceptive and motor representation cannot be excluded at least in trained participants. As mentioned above, using untrained participants might be a solution to that. Studying everyday action sounds, that we all create numerously, will however always include multiple senses as well as a motoric component, as we all are “experts” for them.

From our results, and supported by other research, I can cautiously conclude that the auditory modality plays a special role for those actions that create sounds rather intentionally, opposed to actions creating sounds incidentally, which can be evaluated more visually-based. A stronger attentional focus on action sounds for G actions is another possible explanation and is discussed in the following section.

4.5 The interplay of prediction and attention

It is hardly possible to investigate prediction without stumbling on the question which role attention plays regarding the examined processes. The terms “attention” and “expectation” have even been used interchangeably by some (Corbetta & Shulman, 2002). The interplay between prediction and attention is highly debated (for recent reviews, see e.g. Schröger, Kotz,

et al., 2015; Schröger, Marzecová, & Sanmiguel, 2015), and although an extensive dive into the existing literature would lead too far, it is still mandatory to discuss the possible attentional effects in our studies.

Overall, prediction and attention both aid perceptual processing and can both be implemented in a single framework of predictive coding (cf. Clark, 2013). While attention is mostly thought to magnify specific predictive processes (cf. Schröger, Marzecová, et al., 2015), there are instances where prediction and attention have opposing effects. Contrary to the overall supported claim that processing of self-initiated stimuli is attenuated, as reflected e.g. by a lower N1 amplitude (see Section 1.2), there are studies showing an enhancement of self-generated sound processing (Reznik et al., 2015; Reznik, Henkin, Levy, & Mukamel, 2015) and predictable stimuli in general (Kok, Rahnev, Jehee, Lau, & De Lange, 2012). It is not clearly understood when self-initiated sounds might lead to an enhancement opposed to an attenuation, but studies in the visual system suggest that the task-relevance of the used stimuli might play a role (cf. Summerfield & Egnér, 2009), with predicted stimuli actually being enhanced when they are relevant to perform the task at hand. In our studies, G action sounds, being presumably more task-relevant than B action sounds, are attenuated, at least at the level of the primary auditory cortex. It is however important to note that both B and G action sounds are both somewhat task-relevant, as a rating had to be performed on both, contrary to comparing active sound generation with a passive listening task (Reznik et al., 2015).

Considering the hierarchical predictive structure, it is possible that G action sounds are attenuated compared to B action sounds on the level of primary cortex, and enhanced on the level of pSTG, which would match our results of consistently higher pSTG activity for G action sounds perfectly. It is thus possible that automatic attentional processes are drawn to B action sounds, which are less predictable in a sense, while intentional attention is applied to monitoring G action sounds, and that these processes are reflected in the hierarchy. Again, research from

the visual domain shows similar results, with attenuation being restricted to the primary visual cortex, whereas attention enhances the forward drive of information along the cortical hierarchy and is thus reflected in activation of later cortical areas (Buffalo, Fries, Landman, Liang, & Desimone, 2010; Kok et al., 2012). Attention additionally seems to affect later cortical components more than earlier ones (Chennu et al., 2016). This would explain why we consistently, in all three studies, found higher activation of pSTG and SMA for G action sounds. Attention might be used as a precision-weighting process within a predictive model (Feldman & Friston, 2010; Hohwy, 2012), allocating attention to those stimuli where a prediction error signal would yield the most gain (Smout, Tang, Garrido, & Mattingley, 2019). This approach has to be tested more precisely, as results are still somewhat inconclusive (for a discussion, cf. Heilbron & Chait, 2018). Note however that the prediction error signals, like the MMN, are still elicited when no attention is involved at all, i.e. in a sleeping (Sculthorpe, Ouellet, & Campbell, 2009) or even comatose state (Fischer, Morlet, & Giard, 2000).

Although attention and prediction are most likely two separate processes that are both “in service of perception and action” (Schröger, Kotz, et al., 2015) and aid each other, it is important to sharpen awareness to their distinctions in future research. Applying different experimental instructions to yield different stimuli consciously task-relevant and task-irrelevant may be a way to accomplish this. The experimental modulation of attention could be another approach. Dichotic listening tasks, where sounds are presented to both ears but input to one ear only has to be attended (cf. Foldal et al., 2020 for a study examining auditory rhythm predictability using a dichotic listening task), might be a valuable method for that

4.6 Practical applications of our results

Reviewing our findings, we know that intentional action sounds are crucial for proper action evaluation and execution, and that incidental action sounds have at least facilitating effects and serve as error detectors (see results of Study I & Study II). The most obvious applications can be found for musicians and athletes. Musicians (instrumentalists and singers) produce sounds intentionally, and would therefore benefit the most from training using auditory feedback. Athletes (as presented in Section 4.2) also benefit from auditory training, although our results suggest that visual training or a multisensory combination might be of more use for them.

Moving away from the obvious applications, there might be more extravagant and modern applications as well. One of the first examples for action sounds I used was the clicking sound of our keyboard when we type on our laptop. However, advances in technology lead to a replacement of actual, physical buttons that we have to press and that create sounds (and haptic feedback) with touchscreens that provide no such feedback. Technologically advanced products seem to be generally more silent. This is the case in both a private context, e.g. when we use our smartphones, and in a more professional context, in e.g. airplane cockpits. Does the lack of auditory (and other feedback) irritate us? Studies suggest that applications artificially including auditory and haptic feedback generally reduce error rates (Altinsoy & Merchel, 2009; Wan, Prinet, & Sarter, 2017). Additionally, studies with a force sensor device showed that participants used auditory stimuli to optimize the execution of tone-eliciting actions (Neszmélyi & Horváth, 2017) and action optimization was more pronounced when both visual and auditory feedback were available (Neszmélyi & Horváth, 2019). These results indicate that we can benefit from auditory feedback while operating devices which usually do not create sounds.

The relevance of action sounds for ergonomic purposes and object design is a very interesting field, which will become more relevant in the future.

A similarly novel field are virtual realities, which might benefit in the same way by the inclusion of action sounds, as they help to increase our sense of agency. Considering virtual training exercises, it might be useful to include additional auditory feedback to the provided visual feedback for a smoother error detection and an improved performance.

The ubiquity of action sounds in our everyday life makes it simple to find multiple applications for our findings, even if they are not immediately obvious. By examining these practical applications experimentally, we might get a clearer picture of the underlying necessity of the myriads of action sounds engulfing us and shadowing our every movement – even the ones that seem unnecessary at first glance.

5. Conclusion

We are the first to study the possible distinction between different kinds of action sounds, namely, intentional action sounds, where the sound production is the (sole) goal of the action, and incidental action sounds, where sounds are produced as a by-product. Additionally, we add to the sparse amount of research on the relevance of real-life action sounds by using two whole-body sound producing actions – tap dancing and hurdling. Our three studies aim to strengthen the concepts of intentional and incidental action sounds as two subclasses of human action sounds, that differ from each other both in their behavioural relevance and their neural representation. We were able to provide evidence for a hierarchical processing of action sounds, with the primary auditory cortex on the lowest level, followed by pSTG one level up, and SMA as the apex of this hierarchical framework. Intentional G action sounds were more attenuated in primary auditory cortex, and yielded consistently higher activation of pSTG and SMA. This is in concordance with more predictive processing in intentional action sounds, as those are the goal of the action and crucial for performance evaluation. Although we faced a few limitations in our studies, and moved away from the idea of a clear distinction between the two subtypes by postulating a continuum instead, we are confident that our results contribute to the understanding of action sound processing and encourage this line of research to be continued. Multisensory representations of action sounds and the interplay of prediction and attention are especially noteworthy aspects that need further investigation.

References

- Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. <https://doi.org/10.1007/s00429-012-0475-5>
- Agostini, T., Righi, G., Galmonte, A., & Bruno, P. (2004). The Relevance of Auditory Information in Optimizing Hammer Throwers Performance. In P. B. Pascolo (Ed.), *Biomechanics and Sports* (pp. 67–74). Vienna: Springer Vienna.
- Ainsworth, C. (2015). Sex redefined. *Nature*, *518*(7539), 288–291. <https://doi.org/10.1038/518288a>
- Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience*, *21*(4), 791–802. <https://doi.org/10.1162/jocn.2009.21055>
- Allerdissen, M., Güldenpenning, I., Schack, T., & Bläsing, B. (2017). Recognizing fencing attacks from auditory and visual information: A comparison between expert fencers and novices. *Psychology of Sport and Exercise*, *31*(April), 123–130. <https://doi.org/10.1016/j.psychsport.2017.04.009>
- Altinsoy, M. E., & Merchel, S. (2009). Audiotactile feedback design for touch screens. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5763 LNCS*, 136–144. https://doi.org/10.1007/978-3-642-04076-4_15
- Araneda, R., Renier, L., Ebner-Karestinis, D., Dricot, L., & De Volder, A. G. (2017). Hearing, feeling or seeing a beat recruits a supramodal network in the auditory dorsal stream. *European Journal of Neuroscience*, *45*(11), 1439–1450. <https://doi.org/10.1111/ejn.13349>
- Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. <https://doi.org/10.1016/j.tics.2012.05.003>
- Badian, M., Appel, E., Palm, D., Rupp, W., Sittig, W., & Taeuber, K. (1979). Standardized mental stress in healthy volunteers induced by delayed auditory feedback (DAF). *European Journal of Clinical Pharmacology*, *16*(3), 171–176. <https://doi.org/10.1007/BF00562057>
- Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology*, *48*(9), 1276–1283. <https://doi.org/10.1111/j.1469-8986.2011.01196.x>
- Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., ... Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*, *30*(3), 917–926. <https://doi.org/10.1016/j.neuroimage.2005.10.044>
- Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudnessfeedback: An automatic mechanism for stabilizing voiceamplitude. *J Acoust Soc Am*, *119*(4), 2363–2371. https://doi.org/10.1007/978-1-4939-3274-0_14
- Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for

- predictive processing in audition: A review. *International Journal of Psychophysiology*, 83(2), 120–131. <https://doi.org/10.1016/j.ijpsycho.2011.08.003>
- Bischoff, M., Zentgraf, K., Pilgramm, S., Stark, R., Krüger, B., & Munzert, J. (2014). Anticipating action effects recruits audiovisual movement representations in the ventral premotor cortex. *Brain and Cognition*, 92, 39–47. <https://doi.org/10.1016/j.bandc.2014.09.010>
- Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *NeuroReport*, 11(11), R11–R16. <https://doi.org/10.1097/00001756-200008030-00002>
- Brainard, M. S., & Doupe, A. J. (2000). Auditory feedback in learning and maintenance of vocal behaviour. *Nature Reviews. Neuroscience*, 1(1), 31–40. <https://doi.org/10.1038/35036205>
- Buffalo, E. A., Fries, P., Landman, R., Liang, H., & Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, 107(1), 361–365. <https://doi.org/10.1073/pnas.0907658106>
- Camponogara, I., Rodger, M., Craig, C., & Cesari, P. (2017). Expert players accurately detect an opponent's movement intentions through sound alone. *Journal of Experimental Psychology: Human Perception and Performance*. Camponogara, Ivan: Department of Neurological, Biomedical and Movement Sciences, University of Verona, Via Casorati, 42, Verona, Italy, 37121, ivan.camponogara@gmail.com: American Psychological Association. <https://doi.org/10.1037/xhp0000316>
- Candidi, M., Sacheli, L. M., Mega, I., & Aglioti, S. M. (2014). Somatotopic mapping of piano fingering errors in sensorimotor experts: TMS studies in pianists and visually trained musically naïves. *Cerebral Cortex*, 24(2), 435–443. <https://doi.org/10.1093/cercor/bhs325>
- Carbajal, G. V., & Malmierca, M. S. (2018). The Neuronal Basis of Predictive Coding Along the Auditory Pathway: From the Subcortical Roots to Cortical Deviance Detection. *Trends in Hearing*, 22, 1–33. <https://doi.org/10.1177/2331216518784822>
- Castiello, U., Giordano, B. L., Begliomini, C., Ansuini, C., & Grassi, M. (2010). When ears drive hands: The influence of contact sound on reaching to grasp. *PLoS ONE*, 5(8), 1–9. <https://doi.org/10.1371/journal.pone.0012240>
- Cesari, P., Camponogara, I., Papetti, S., Rocchesso, D., & Fontana, F. (2014). Might as well jump: Sound affects muscle activation in skateboarding. *PLoS ONE*, 9(3), 1–9. <https://doi.org/10.1371/journal.pone.0090156>
- Chennu, S., Noreika, V., Gueorguiev, D., Shtyrov, Y., Bekinschtein, T. A., & Henson, R. (2016). Silent expectations: Dynamic causal modeling of cortical prediction and attention to sounds that weren't. *Journal of Neuroscience*, 36(32), 8305–8316. <https://doi.org/10.1523/JNEUROSCI.1125-16.2016>
- Chesters, J., Baghai-Ravary, L., & Möttönen, R. (2015). The effects of delayed auditory and visual feedback on speech production. *The Journal of the Acoustical Society of America*, 137(2), 873–883. <https://doi.org/10.1121/1.4906266>
- Christoffels, I. K., van de Ven, V., Waldorp, L. J., Formisano, E., & Schiller, N. O. (2011). The sensory consequences of speaking: Parametric neural cancellation during speech in auditory cortex. *PLoS ONE*, 6(5). <https://doi.org/10.1371/journal.pone.0018307>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of

- cognitive science. *The Behavioral and Brain Sciences*, 36(3), 181–204.
<https://doi.org/10.1017/S0140525X12000477>
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3), 201–215.
<https://doi.org/10.1038/nrn755>
- Denham, S. L., & Winkler, I. (2020). Predictive coding in auditory perception: challenges and unresolved questions. *European Journal of Neuroscience*, 51(5), 1151–1160.
<https://doi.org/10.1111/ejn.13802>
- Dercksen, T. T., Widmann, A., Schröger, E., & Wetzels, N. (2020). Omission related brain responses reflect specific and unspecific action-effect couplings. *NeuroImage*, 215(March). <https://doi.org/10.1016/j.neuroimage.2020.116840>
- Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, 9(6), 719–721. <https://doi.org/10.1068/p090719>
- Dogge, M., Custers, R., & Aarts, H. (2019). Moving Forward: On the Limits of Motor-Based Forward Models. *Trends in Cognitive Sciences*, 23(9), 743–753.
<https://doi.org/10.1016/j.tics.2019.06.008>
- Drost, U. C., Rieger, M., Brass, M., Gunter, T. C., & Prinz, W. (2005a). Action-effect coupling in pianists. *Psychological Research*, 69(4), 233–241.
<https://doi.org/10.1007/s00426-004-0175-8>
- Drost, U. C., Rieger, M., Brass, M., Gunter, T. C., & Prinz, W. (2005b). When hearing turns into playing: Movement induction by auditory stimuli in pianists. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 58(8), 1376–1389. <https://doi.org/10.1080/02724980443000610>
- Drost, U. C., Rieger, M., & Prinz, W. (2007). Instrument specificity in experienced musicians. *Quarterly Journal of Experimental Psychology*, 60(4), 527–533.
<https://doi.org/10.1080/17470210601154388>
- Dürschmid, S., Edwards, E., Reichert, C., Dewar, C., Hinrichs, H., Heinze, H. J., ... Knight, R. T. (2016). Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 113(24), 6755–6760. <https://doi.org/10.1073/pnas.1525030113>
- Effenberg, A. O., Fehse, U., Schmitz, G., Krueger, B., & Mechling, H. (2016). Movement sonification: Effects on motor learning beyond rhythmic adjustments. *Frontiers in Neuroscience*, 10(MAY), 1–18. <https://doi.org/10.3389/fnins.2016.00219>
- Eg, R., & Behne, D. M. (2015). Perceived synchrony for realistic and dynamic audiovisual events. *Frontiers in Psychology*, 6(JUN). <https://doi.org/10.3389/fpsyg.2015.00736>
- Fairbanks, G. (1954). Systematic Research In Experimental Phonetics: A Theory Of The Speech Mechanism As A Servosystem. *Journal of Speech and Hearing Disorders*, 19(2), 133–139. <https://doi.org/10.1044/jshd.1902.133>
- Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4(December), 1–23. <https://doi.org/10.3389/fnhum.2010.00215>
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, 1(1), 1–47.
<https://doi.org/10.1093/cercor/1.1.1>
- Fischer, C., Morlet, D., & Giard, M. H. (2000). Mismatch negativity and N100 in comatose

- patients. *Audiology and Neuro-Otology*, 5(3–4), 192–197.
<https://doi.org/10.1159/000013880>
- Foldal, M. D., Blenkmann, A. O., Llorens, A., Knight, R. T., Solbakk, A. K., & Endestad, T. (2020). The brain tracks auditory rhythm predictability independent of selective attention. *Scientific Reports*, 10(1), 1–13. <https://doi.org/10.1038/s41598-020-64758-y>
- Fonken, Y. M., Mukerji, A., Jimenez, R., Lin, J., Brunner, P., Schalk, G., & Knight, R. T. (2019). Unexpected sound omissions are signaled in human posterior superior temporal gyrus: an intracranial study. *BioRxiv*, 62(1), 55–55.
https://doi.org/10.11453/orltokeo.62.1_55
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836.
<https://doi.org/10.1098/rstb.2005.1622>
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>
- Fujisaki, W., & Nishida, S. (2005). Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals. *Experimental Brain Research*, 166(3–4), 455–464. <https://doi.org/10.1007/s00221-005-2385-8>
- Garrido, M. I., Friston, K. J., Kiebel, S. J., & Stephan, K. E. (2009). Europe PMC Funders Group The functional anatomy of the MMN : A DCM study of the roving paradigm. *NeuroImage*, 42(2), 936–944. <https://doi.org/10.1016/j.neuroimage.2008.05.018>.The
- Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, 120(3), 453–463. <https://doi.org/10.1016/j.clinph.2008.11.029>
- Gates, A., Bradshaw, J. L., & Nettleton, N. C. (1974). Effect of different delayed auditory feedback intervals on a music performance task. *Perception & Psychophysics*, 15(1), 21–25. <https://doi.org/10.3758/BF03205822>
- Ghio, M., Scharmach, K., & Bellebaum, C. (2018). ERP correlates of processing the auditory consequences of own versus observed actions. *Psychophysiology*, 55(6).
<https://doi.org/10.1111/psyp.13048>
- Hasegawa, T., Matsuki, K. I., Ueno, T., Maeda, Y., Matsue, Y., Konishi, Y., & Sadato, N. (2004). Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Cognitive Brain Research*, 20(3), 510–518. <https://doi.org/10.1016/j.cogbrainres.2004.04.005>
- Hashimoto, Y., & Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. *Human Brain Mapping*, 20(1), 22–28. <https://doi.org/10.1002/hbm.10119>
- 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(2), 282–293.
<https://doi.org/10.1162/0898929053124893>
- Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, 389, 54–73.
<https://doi.org/10.1016/j.neuroscience.2017.07.061>
- Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain.

- Frontiers in Psychology*, 3(APR), 1–14. <https://doi.org/10.3389/fpsyg.2012.00096>
- 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(5), 849–878. <https://doi.org/10.1017/S0140525X01000103>
- Horváth, J. (2015). Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain Research*. <https://doi.org/10.1016/j.brainres.2015.03.038>
- Houde, J. F., & Jordan, M. I. (2002). Sensorimotor Adaptation of Speech I. *Journal of Speech, Language, and Hearing Research*, 45(2), 295–310. [https://doi.org/10.1044/1092-4388\(2002/023\)](https://doi.org/10.1044/1092-4388(2002/023))
- Howell, P. (2004). Effects of delayed auditory feedback and frequency-shifted feedback on speech control and some potentials for future development of prosthetic aids for stammering. *Stammering Research : An on-Line Journal Published by the British Stammering Association*, 1(1), 31–46. <https://doi.org/10.1126/science.1182395.Evolution>
- 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(1), 133–151. <https://doi.org/10.1037/a0028566>
- Jo, H. G., Habel, U., & Schmidt, S. (2019). Role of the supplementary motor area in auditory sensory attenuation. *Brain Structure and Function*, (0123456789). <https://doi.org/10.1007/s00429-019-01920-x>
- Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research*, 190(3), 279–287. <https://doi.org/10.1007/s00221-008-1473-y>
- Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: the role of auditory feedback. *The Journal of the Acoustical Society of America*, 113(1), 532–543. <https://doi.org/10.1121/1.1529670>
- Kaiser, J., & Schütz-Bosbach, S. (2018). Sensory attenuation of self-produced signals does not rely on self-specific motor predictions. *European Journal of Neuroscience*, 47(11), 1303–1310. <https://doi.org/10.1111/ejn.13931>
- Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive Processing: A Canonical Cortical Computation. *Neuron*, 100(2), 424–435. <https://doi.org/10.1016/j.neuron.2018.10.003>
- 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(1), 102–111. <https://doi.org/10.1016/j.concog.2005.12.004>
- 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(2), 157–165. <https://doi.org/10.1080/20445911.2013.869226>
- Kennel, C., Streese, L., Pizzera, A., Justen, C., Hohmann, T., & Raab, M. (2015). Auditory reafferences: The influence of real-time feedback on movement control. *Frontiers in Psychology*, 6(JAN), 1–6. <https://doi.org/10.3389/fpsyg.2015.00069>
- Keough, D., & Jones, J. A. (2009). The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing. *The Journal of the Acoustical Society of America*, 126(2), 837–846. <https://doi.org/10.1121/1.3158600>
- Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & De Lange, F. P. (2012). Attention reverses

- the effect of prediction in silencing sensory signals. *Cerebral Cortex*, 22(9), 2197–2206. <https://doi.org/10.1093/cercor/bhr310>
- Kompus, K., Volehaugen, V., Todd, J., & Westerhausen, R. (2019). Hierarchical modulation of auditory prediction error signaling is independent of attention. *Cognitive Neuroscience*, 1–11. <https://doi.org/10.1080/17588928.2019.1648404>
- Krala, M., van Kemenade, B., Straube, B., Kircher, T., & Bremmer, F. (2019). Predictive coding in a multisensory path integration task: An fMRI study. *Journal of Vision*, 19(11), 13. <https://doi.org/10.1167/19.11.13>
- Kulpa, J. D., & Pfordresher, P. Q. (2013). Effects of delayed auditory and visual feedback on sequence production. *Experimental Brain Research*, 224(1), 69–77. <https://doi.org/10.1007/s00221-012-3289-z>
- Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action Representation of Sound: Audiomotor Recognition Network While Listening to Newly Acquired Actions. *Journal of Neuroscience*, 27(2), 308–314. <https://doi.org/10.1523/jneurosci.4822-06.2007>
- Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, 32(27), 9351–9358. <https://doi.org/10.1523/JNEUROSCI.0404-12.2012>
- Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *The Journal of the Acoustical Society of America*, 89(2), 859–866. <https://doi.org/10.1121/1.1894647>
- Lane, Harlan, & Tranel, B. (1971). The Lombard Sign and the Role of Hearing in Speech. *Journal of Speech and Hearing Research*, 14(4), 677–709. <https://doi.org/10.1044/jshr.1404.677>
- Lee, H. L., & Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proceedings of the National Academy of Sciences of the United States of America*, 108(51). <https://doi.org/10.1073/pnas.1115267108>
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–75. <https://doi.org/10.1017/S0140525X99001776>
- Liberman, M. C., & Guinan, J. J. (1998). Feedback control of the auditory periphery: Anti-masking effects of middle ear muscles vs. olivocochlear efferents. *Journal of Communication Disorders*, 31(6), 471–483. [https://doi.org/10.1016/S0021-9924\(98\)00019-7](https://doi.org/10.1016/S0021-9924(98)00019-7)
- Liebenthal, E., Ellingson, M. L., Spanaki, M. V., Prieto, T. E., Ropella, K. M., & Binder, J. R. (2003). Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *NeuroImage*, 19(4), 1395–1404. [https://doi.org/10.1016/S1053-8119\(03\)00228-3](https://doi.org/10.1016/S1053-8119(03)00228-3)
- Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of Supplementary Motor Areas in Auditory Processing and Auditory Imagery. *Trends in Neurosciences*, 39(8), 527–542. <https://doi.org/10.1016/j.tins.2016.06.003>
- Lincoln, M., Packman, A., & Onslow, M. (2006). Altered auditory feedback and the treatment of stuttering: A review. *Journal of Fluency Disorders*. <https://doi.org/10.1016/j.jfludis.2006.04.001>
- Loehr, J. D., Large, E. W., & Palmer, C. (2011). Temporal coordination and adaptation to rate

- change in music performance. *Journal of Experimental Psychology: Human Perception and Performance*. Palmer, Caroline: Department of Psychology, McGill University, 1205 Dr Penfield Ave, Montreal, PQ, Canada, H3A 1B1, caroline.palmer@mcgill.ca: American Psychological Association. <https://doi.org/10.1037/a0023102>
- Martikainen, M. H., Kaneko, K. I., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex*, *15*(3), 299–302. <https://doi.org/10.1093/cercor/bhh131>
- Mathiak, K., Rapp, A., Kircher, T. T. J., Grodd, W., Hertrich, I., Weiskopf, N., ... Ackermann, H. (2002). Mismatch responses to randomized gradient switching noise as reflected by fMRI and whole-head magnetoencephalography. *Human Brain Mapping*, *16*(3), 190–195. <https://doi.org/10.1002/hbm.10041>
- 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*(3), 184–192. <https://doi.org/10.1080/17588921003743581>
- Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*. [https://doi.org/10.1016/S0893-6080\(96\)00035-4](https://doi.org/10.1016/S0893-6080(96)00035-4)
- Mifsud, N. G., & Whitford, T. J. (2017). Sensory attenuation of self-initiated sounds maps onto habitual associations between motor action and sound. *Neuropsychologia*, *103*(March), 38–43. <https://doi.org/10.1016/j.neuropsychologia.2017.07.019>
- Mukerji, S., Windsor, A. M., & Lee, D. J. (2010). Auditory Brainstem Circuits That Mediate the Middle Ear Muscle Reflex. *Trends in Amplification*, *14*(3), 170–191. <https://doi.org/10.1177/1084713810381771>
- Murgia, M., Hohmann, T., Galmonte, A., Raab, M., & Agostini, T. (2012). Recognising one's own motor actions through sound: The role of temporal factors. *Perception*, *41*(8), 976–987. <https://doi.org/10.1068/p7227>
- Murgia, M., Prpic, V., Jenny, O., McCullagh, P., Santoro, I., Galmonte, A., & Agostini, T. (2017). Modality and perceptual-motor experience influence the detection of temporal deviations in tap dance sequences. *Frontiers in Psychology*, *8*(AUG), 1–8. <https://doi.org/10.3389/fpsyg.2017.01340>
- Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. [https://doi.org/10.1016/0001-6918\(78\)90006-9](https://doi.org/10.1016/0001-6918(78)90006-9)
- Näätänen, Risto, & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*, *24*(4).
- Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience*, *9*(11), 856–869. <https://doi.org/10.1038/nrn2478>
- Nazimek, J. M., Hunter, M. D., Hoskin, R., Wilkinson, I., & Woodruff, P. W. (2013). Neural basis of auditory expectation within temporal cortex. *Neuropsychologia*, *51*(11), 2245–2250. <https://doi.org/10.1016/j.neuropsychologia.2013.07.019>
- Neszmélyi, B., & Horváth, J. (2017). Consequences matter: Self-induced tones are used as feedback to optimize tone-eliciting actions. *Psychophysiology*, *54*(6), 904–915. <https://doi.org/10.1111/psyp.12845>

- Neszmélyi, B., & Horváth, J. (2019). The role of auditory context in action-effect-related motor adaptation. *Human Movement Science*, 67, 102503. <https://doi.org/https://doi.org/10.1016/j.humov.2019.102503>
- Novembre, G., & Keller, P. E. (2014). A conceptual review on action-perception coupling in the musicians' brain: What is it good for? *Frontiers in Human Neuroscience*, 8(AUG), 1–11. <https://doi.org/10.3389/fnhum.2014.00603>
- Parizet, E., Guyader, E., & Nosulenko, V. (2008). Analysis of car door closing sound quality. *Applied Acoustics*, 69(1), 12–22. <https://doi.org/10.1016/j.apacoust.2006.09.004>
- Parrell, B., & Houde, J. (2019). Modeling the role of sensory feedback in speech motor control and learning. *Journal of Speech, Language, and Hearing Research*, 62(8S), 2963–2985. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0127
- Perkell, J. S. (2013). Mechanisms in Speech Production. *J. Neurolinguis.*, 25(5), 382–407. <https://doi.org/10.1016/j.jneuroling.2010.02.011.Movement>
- Persichetti, A. S., & Dilks, D. D. (2016). Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 77, 155–163. <https://doi.org/10.1016/j.cortex.2016.02.006>
- Pfordresher, P. Q. (2005). Auditory Feedback in Music Performance: The Role of Melodic Structure and Musical Skill. *Journal of Experimental Psychology: Human Perception and Performance*. Pfordresher, Peter Q.: Department of Psychology and Institute for Music Research, University of Texas at San Antonio, 6900 North Loop 1604 West, San Antonio, TX, US, 78249-0652, peter.pfordresher@utsa.edu: American Psychological Association. <https://doi.org/10.1037/0096-1523.31.6.1331>
- Pfordresher, P. Q. (2006). Coordination of perception and action in music performance. *Advances in Cognitive Psychology*, 2(2), 183–198. <https://doi.org/10.2478/v10053-008-0054-8>
- Pfordresher, P. Q. (2012). Musical training and the role of auditory feedback during performance. *Annals of the New York Academy of Sciences*, 1252, 171–178. <https://doi.org/10.1111/j.1749-6632.2011.06408.x>
- Pfordresher, P. Q., & Beasley, R. T. E. (2014). Making and monitoring errors based on altered auditory feedback. *Frontiers in Psychology*, 5(AUG), 1–13. <https://doi.org/10.3389/fpsyg.2014.00914>
- Pfordresher, P. Q., Mantell, J. T., Brown, S., Zivadinov, R., & Cox, J. L. (2014). Brain responses to altered auditory feedback during musical keyboard production: an fMRI study. *Brain Research*, 1556, 28–37. <https://doi.org/10.1016/j.brainres.2014.02.004>
- Pfordresher, P. Q., & Palmer, C. (2006). Effects of hearing the past, present and future during music performance. *Perception & Psychophysics*, 68(3), 362–376. Retrieved from <https://link.springer.com/content/pdf/10.3758/BF03193683.pdf>
- Pizzera, A., & Hohmann, T. (2015). Acoustic Information During Motor Control and Action Perception: A Review. *The Open Psychology Journal*, 8(1), 183–191. <https://doi.org/10.2174/1874350101508010183>
- Pizzera, A., Hohmann, T., Streese, L., Habbig, A., & Raab, M. (2017). Long-term effects of acoustic refference training (ART). *European Journal of Sport Science*, 17(10), 1279–1288. <https://doi.org/10.1080/17461391.2017.1381767>
- Poulet, J. F. A., & Hedwig, B. (2002). A corollary discharge maintains auditory sensitivity

- during sound production. *Nature*, 418(6900), 872–876.
<https://doi.org/10.1038/nature00919>
- Pruitt, T. A., & Pfordresher, P. Q. (2015). The role of auditory feedback in speech production. *Journal of Experimental Psychology: Human Performance and Perception*, 41(1), 152–166.
- Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, 120(2), 966–977. <https://doi.org/10.1121/1.2217714>
- Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, 119(4), 2288–2297. <https://doi.org/10.1121/1.2173514>
- Raij, T., McEvoy, L., Mäkelä, J. P., & Hari, R. (1997). Human auditory cortex is activated by omission of auditory stimuli. *Brain Research*, 745(1–2), 134–143.
[https://doi.org/10.1016/S0006-8993\(96\)01140-7](https://doi.org/10.1016/S0006-8993(96)01140-7)
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87. <https://doi.org/10.1038/4580>
- Recasens, M., Grimm, S., Capilla, A., Nowak, R., & Escera, C. (2014). Two sequential processes of change detection in hierarchically ordered areas of the human auditory cortex. *Cerebral Cortex*, 24(1), 143–153. <https://doi.org/10.1093/cercor/bhs295>
- Renier, L. A., Anurova, I., De Volder, A. G., Carlson, S., VanMeter, J., & Rauschecker, J. P. (2010). Preserved functional specialization for spatial processing in the middle occipital gyrus of the early blind. *Neuron*, 68(1), 138–148.
<https://doi.org/10.1016/j.neuron.2010.09.021>
- Repp, B.H. (1999). Effects of Auditory Feedback Deprivation on Expressive Piano Performance, 16(4), 409–438.
- Repp, Bruno H. (1987). The sound of two hands clapping: An exploratory study. *J Acoust Soc Am*, 81(4), 1100–1109.
- Reznik, D., Ossmy, O., & Mukamel, R. (2015). Enhanced Auditory Evoked Activity to Self-Generated Sounds Is Mediated by Primary and Supplementary Motor Cortices. *Journal of Neuroscience*, 35(5), 2173–2180. <https://doi.org/10.1523/JNEUROSCI.3723-14.2015>
- Reznik, Daniel, Henkin, Y., Levy, O., & Mukamel, R. (2015). Perceived loudness of self-generated sounds is differentially modified by expected sound intensity. *PLoS ONE*, 10(5). <https://doi.org/10.1371/journal.pone.0127651>
- Rizzolatti, G., & Craighero, L. (2004). the Mirror-Neuron System. *Annual Review of Neuroscience*, 27(1), 169–192. <https://doi.org/10.1146/annurev.neuro.27.070203.144230>
- Roberts, T. F., Hisey, E., Tanaka, M., Kearney, M., Chatree, G., Yang, C. F., ... Mooney, R. (2017). Identification of motor to auditory pathway important for vocal learning. *Nat Neurosci*, 20(7), 978–986. <https://doi.org/10.1038/nn.4563>. Identification
- Rummell, B. P., Klee, J. L., & Sigurdsson, T. (2016). Attenuation of responses to self-generated sounds in auditory cortical neurons. *Journal of Neuroscience*, 36(47). <https://doi.org/10.1523/JNEUROSCI.1564-16.2016>
- SanMiguel, I., Saupe, K., & Schröger, E. (2013). I know what is missing here: Electrophysiological prediction error signals elicited by omissions of predicted “what”

- but not “when.” *Frontiers in Human Neuroscience*, 7(JUL), 1–10.
<https://doi.org/10.3389/fnhum.2013.00407>
- Sasisekaran, J. (2012). Effects of delayed auditory feedback on speech kinematics in fluent speakers. *Perceptual and Motor Skills*, 115(3), 845–864.
<https://doi.org/10.2466/15.22.PMS.115.6.845-864>
- Sato, A. (2008). Action observation modulates auditory perception of the consequence of others’ actions. *Consciousness and Cognition*, 17(4), 1219–1227.
<https://doi.org/10.1016/j.concog.2008.01.003>
- Schaffert, N., Janzen, T. B., Mattes, K., & Thaut, M. H. (2019). A Review on the Relationship Between Sound and Movement in Sports and Rehabilitation. *Frontiers in Psychology*, 10(February), 244. <https://doi.org/10.3389/fpsyg.2019.00244>
- Schaffert, N., Oldag, B., & Cesari, P. (2020). Sound matters: The impact of auditory deprivation on movement precision in rowing. *European Journal of Sport Science*, 0(0), 1–8. <https://doi.org/10.1080/17461391.2019.1710265>
- Schneider, D. M., & Mooney, R. (2018). How Movement Modulates Hearing. *Annual Review of Neuroscience*, 41(1), 553–572. <https://doi.org/10.1146/annurev-neuro-072116-031215>
- Schneider, D. M., Sundararajan, J., & Mooney, R. (2018). A cortical filter that learns to suppress the acoustic consequences of movement. *Nature*, 561(7723), 391–395.
<https://doi.org/10.1038/s41586-018-0520-5>
- Schönwiesner, M., Novitski, N., Pakarinen, S., Carlson, S., Tervaniemi, M., & Näätänen, R. (2007). Heschl’s gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. *Journal of Neurophysiology*, 97(3), 2075–2082. <https://doi.org/10.1152/jn.01083.2006>
- Schröger, E., Kotz, S. A., & SanMiguel, I. (2015). Bridging prediction and attention in current research on perception and action. *Brain Research*.
<https://doi.org/10.1016/j.brainres.2015.08.037>
- Schröger, E., Marzecová, A., & Sanmiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *European Journal of Neuroscience*, 41(5), 641–664. <https://doi.org/10.1111/ejn.12816>
- Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in Cognitive Sciences*, 11(5), 211–218.
<https://doi.org/10.1016/j.tics.2007.02.006>
- Schubotz, R. I., & Von Cramon, D. Y. (2003). Functional-anatomical concepts of human premotor cortex: Evidence from fMRI and PET studies. In *NeuroImage*.
<https://doi.org/10.1016/j.neuroimage.2003.09.014>
- Sculthorpe, L. D., Ouellet, D. R., & Campbell, K. B. (2009). MMN elicitation during natural sleep to violations of an auditory pattern. *Brain Research*, 1290, 52–62.
<https://doi.org/10.1016/j.brainres.2009.06.013>
- Smout, C. A., Tang, M. F., Garrido, M. I., & Mattingley, J. B. (2019). Attention promotes the neural encoding of prediction errors. *PLoS Biology*, 17(2), 1–22.
<https://doi.org/10.1371/journal.pbio.2006812>
- Sors, F., Lath, F., Bader, A., Santoro, I., Galmonte, A., Agostini, T., & Murgia, M. (2018). Predicting the length of volleyball serves: The role of early auditory and visual information. *PLoS ONE*, 13(12). <https://doi.org/10.1371/journal.pone.0208174>

- Sors, F., Prpic, V., Santoro, I., Galmonte, A., Agostini, T., & Murgia, M. (2018). Loudness, but not shot power, influences simple reaction times to soccer penalty sounds. *Psihologija*, *51*(1), 127–141. <https://doi.org/10.2298/PSI170926016S>
- Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, *43*(6), 482–489. <https://doi.org/10.1037/h0055479>
- Straube, B., Van Kemenade, B. M., Arikan, B. E., Fiehler, K., Leube, D. T., Harris, L. R., & Kircher, T. (2017). Predicting the multisensory consequences of one's own action: Bold suppression in auditory and visual cortices. *PLoS ONE*, *12*(1), 1–25. <https://doi.org/10.1371/journal.pone.0169131>
- Stuart, A., Kalinowski, J., Rastatter, M. P., & Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *The Journal of the Acoustical Society of America*, *111*(5 Pt 1), 2237–2241. <https://doi.org/10.1121/1.1466868>
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. <https://doi.org/10.1016/j.tics.2009.06.003>
- Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *J Acoust Soc Am*, *September*(84(3)), 917–928. <https://doi.org/10.1038/jid.2014.371>
- Takeuchi, T. (1993). Auditory information in playing tennis. *Perceptual and Motor Skills*, *76*(3 Pt 2), 1323–1328. <https://doi.org/10.2466/pms.1993.76.3c.1323>
- Timm, J., Schönwiesner, M., SanMiguel, I., & Schröger, E. (2014). Sensation of agency and perception of temporal order. *Consciousness and Cognition*, *23*(1), 42–52. <https://doi.org/10.1016/j.concog.2013.11.002>
- Timm, J., Schönwiesner, M., Schröger, E., & SanMiguel, I. (2016). Sensory suppression of brain responses to self-generated sounds is observed with and without the perception of agency. *Cortex*, *80*, 5–20. <https://doi.org/10.1016/j.cortex.2016.03.018>
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443. <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- van der Steen, M. C., Molendijk, E. B. D., Altenmüller, E., & Furuya, S. (2014). Expert pianists do not listen: The expertise-dependent influence of temporal perturbation on the production of sequential movements. *Neuroscience*, *269*, 290–298. <https://doi.org/10.1016/j.neuroscience.2014.03.058>
- Van Elk, M., Salomon, R., Kannape, O., & Blanke, O. (2014). Suppression of the N1 auditory evoked potential for sounds generated by the upper and lower limbs. *Biological Psychology*, *102*(1), 108–117. <https://doi.org/10.1016/j.biopsycho.2014.06.007>
- van Kemenade, B. M., Arikan, B. E., Kircher, T., & Straube, B. (2016). Predicting the sensory consequences of one's own action: First evidence for multisensory facilitation. *Attention, Perception, and Psychophysics*, *78*(8), 2515–2526. <https://doi.org/10.3758/s13414-016-1189-1>
- Van Vugt, F. T., & Tillmann, B. (2015). Auditory feedback in error-based learning of motor regularity. *Brain Research*. <https://doi.org/10.1016/j.brainres.2015.02.026>
- Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neuroscience Letters*, *393*(1), 40–44.

- <https://doi.org/10.1016/j.neulet.2005.09.032>
- Ventura, M. I., Nagarajan, S. S., & Houde, J. F. (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience*, *10*, 1–11. <https://doi.org/10.1186/1471-2202-10-58>
- von Helmholtz, H. (1867). *Handbuch der physiologischen Optik*. Stuttgart.
- von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip. *Naturwissenschaften*, *37*(20), 464–476. <https://doi.org/10.1007/bf00622503>
- Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, *108*(51), 20754–20759. <https://doi.org/10.1073/pnas.1117807108>
- Wacongne, Catherine, Changeux, J. P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience*, *32*(11), 3665–3678. <https://doi.org/10.1523/JNEUROSCI.5003-11.2012>
- Waldstein, R. S. (1990). Effects of postlingual deafness on speech production: Implications for the role of auditory feedback. *Journal of the Acoustical Society of America*. US: Acoustical Society of American. <https://doi.org/10.1121/1.400107>
- Walsh, K. S., McGovern, D. P., Clark, A., & O’Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. <https://doi.org/10.1111/nyas.14321>
- Wan, Y., Prinnet, J. C., & Sarter, N. (2017). Visual and auditory feedback to improve touchscreen usability in turbulence. *Proceedings of the Human Factors and Ergonomics Society*, *2017-October*, 89–93. <https://doi.org/10.1177/1541931213601488>
- Ward, N. S. (2006). The neural substrates of motor recovery after focal damage to the central nervous system. *Archives of Physical Medicine and Rehabilitation*, *87*(12 Suppl 2), S30–5. <https://doi.org/10.1016/j.apmr.2006.08.334>
- Weiss, C., Herwig, A., & Schütz-Bosbach, S. (2011). The self in action effects: Selective attenuation of self-generated sounds. *Cognition*. <https://doi.org/10.1016/j.cognition.2011.06.011>
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882. <https://doi.org/10.1126/science.7569931>
- Woods, E. A., Hernandez, A. E., Wagner, V. E., & Beilock, S. L. (2014). Expert athletes activate somatosensory and motor planning regions of the brain when passively listening to familiar sports sounds. *Brain and Cognition*, *87*(1), 122–133. <https://doi.org/10.1016/j.bandc.2014.03.007>
- Zampini, M., Guest, S., & Spence, C. (2003). The role of auditory cues in modulating the perception of electric toothbrushes. *Journal of Dental Research*, *82*(11), 929–932. <https://doi.org/10.1177/154405910308201116>
- Zampini, Massimiliano, & Spence, C. (2004). The role of auditory cues in modulating the perceived crispness and staleness of potato chips. *Journal of Sensory Studies*, *19*(5), 347–363. <https://doi.org/10.1111/j.1745-459x.2004.080403.x>
- Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one’s own voice during speech

production. *Journal of Cognitive Neuroscience*, 22(8), 1770–1781.
<https://doi.org/10.1162/jocn.2009.21324>

Abbreviations

A1	primary auditory cortex
B	by-product
BOLD	blood-oxygen-level-dependent
DAF	delayed auditory feedback
DCM	dynamic causal modelling
EEG	electroencephalography
fMRI	functional magnetic resonance imaging
G	goal-related
MEG	magnetoencephalography
MMN	mismatch negativity
MRI	magnetic resonance imaging
pSTG	posterior superior temporal gyrus
SMA	supplementary motor area
STG	superior temporal gyrus
STS	superior temporal sulcus

List of Figures

Figure 1. Predictive hierarchy of action sound processing.....	15
Figure 2. Procedure of the three studies.....	25

Curriculum Vitae

Personal Details

Name	Nina Heins
Date of birth	22/06/1992
Place of birth	Düsseldorf

Education

PhD Psychology Westfälische Wilhelms-Universität, Münster	September 2016 - Present
M.Sc. Psychology and Cognitive Neuroscience Ruhr-Universität Bochum Grade: 1.2	October 2014 - July 2016
B.Sc. Psychology Heinrich-Heine-Universität, Düsseldorf Grade: 1.3	October 2011 - July 2014
Abitur Nelly-Sachs-Gymnasium, Neuss Grade: 1.4	August 2002 – July 2011

Research Experience

Internship at the department of Biopsychology - Head: Prof. Dr. Dr. h.c. Onur Güntürkün Ruhr-Universität Bochum, Germany <i>Supervisor: Dr. Erhan Genç</i>	February – April 2015
Internship at the Research Centre for Cognitive Neuroscience - Head: Prof. Dr. Karen Waldie University of Auckland, New Zealand <i>Supervisor: Haeme Park, MSc</i>	February – March 2014

**Internship at the Institute of Neuroscience
and Medicine (INM-1)**

August – September 2013

- Head: Prof. Dr. Katrin Amunts

Research Centre Jülich, Germany

Supervisor: Prof. Dr. Silke Lux

**Student assistant at the Department of
Cognitive Psychology, Institute of
Experimental Psychology**

July 2012 – June 2015

- Head: Prof. Dr. Martin Heil

Heinrich-Heine-Universität Düsseldorf,
Germany

Supervisor: Dr. Nora Schaal

Teaching Experience

**Seminar „Neurophysiological methods of the
cognitive neurosciences: fMRI”**

Winter term 2019/2020

Westfälische-Wilhelms-Universität Münster

**Seminar „Neurophysiological methods of the
cognitive neurosciences: fMRI”**

Winter term 2018/2019

Westfälische-Wilhelms-Universität Münster

**Tutor for the undergrad course „Evolution &
Emotion“, Biopsychology**

Summer term 2015

Ruhr-Universität Bochum

Declarations

Declarations by the candidate documenting **open-science activities** and **on the consideration of ethical aspects** as part of the doctoral process and disclosure of **personal contribution** for manuscripts completed by two or more authors (cumulative dissertations)

Doctoral candidate: Nina Heins

Title of dissertation:

1. Documentation of open-science activities

Manuscript 1

	Yes	No	If yes, please specify the source
Pre-registration		X	
Publication of data		X	
Publication of analysis scripts		X	
Publication of materials		X	
Open access publication		X	

Manuscript 2

	Yes	No	If yes, please specify the source
Pre-registration		X	
Publication of data		X	
Publication of analysis scripts		X	
Publication of materials		X	
Open access publication		X	

Manuscript 3

	Yes	No	If yes, please specify the source
Pre-registration		X	
Publication of data		X	
Publication of analysis scripts		X	
Publication of materials		X	
Open access publication		X	

2. Declaration on the consideration of ethical aspects

Study number	Source (manuscript / chapter of dissertation): e.g. study 1 in paper 2, study 1 described in chapter 4	Was the study reviewed by an ethics commission?	
		yes	No
1	Manuscript 1	X	
2	Manuscript 2	X	
3	Manuscript 3	X	

3. Declaration of one's personal contribution to the submitted academic manuscripts by two or more authors**Manuscript 1**

Title	Incidental or Intentional? Different Brain Responses to One's Own Action Sounds in Hurdling vs. Tap Dancing		
Author(s)	Nina Heins, Jennifer Pomp, Daniel S. Kluger, Ima Trempler, Karen Zentgraf, Markus Raab, Ricarda I. Schubotz.		
Publication status:	not yet submitted	<input type="checkbox"/>	(please mark with X)
	submitted	<input type="checkbox"/>	
	in review	<input type="checkbox"/>	
	in revision	<input type="checkbox"/>	
	accepted	<input type="checkbox"/>	
	published	<input checked="" type="checkbox"/>	
Journal	Frontiers in Neuroscience		
Year of publication	2020		
Description of your own contribution in the case of joint authorship:			
<ul style="list-style-type: none"> - partly responsible for the study's conception and design - collecting, processing, analysing and interpreting data - mainly responsible for drafting and revising the manuscript - marked as corresponding author 			

Manuscript 2

Title	Where is my sound? Omission of one's own intentional and incidental action sounds and its effect on BOLD fMRI		
Author(s)	Nina Heins, Jennifer Pomp, Karen Zentgraf, Markus Raab, Ricarda I. Schubotz		
Publication status:	not yet submitted	<input type="checkbox"/>	(please mark with X)
	submitted	<input checked="" type="checkbox"/>	
	in review	<input type="checkbox"/>	
	in revision	<input type="checkbox"/>	
	accepted	<input type="checkbox"/>	
	published	<input type="checkbox"/>	
Journal	Frontiers in Neuroscience		
Year of publication	2020		
Description of your own contribution in the case of joint authorship:			
<ul style="list-style-type: none"> - partly responsible for the study's conception and design - mainly responsible for data collection - processing, analysing and interpreting data - mainly responsible for drafting and revising the manuscript - marked as corresponding author 			

Manuscript 3

Title	Too late! Influence of temporal delay on the neural processing of one's own incidental and intentional action sounds.		
Author(s)	Nina Heins, Ima Trempler, Karen Zentgraf, Markus Raab, Ricarda I. Schubotz		
Publication status:	not yet submitted	<input type="checkbox"/>	(please mark with X)
	submitted	<input checked="" type="checkbox"/>	
	in review	<input type="checkbox"/>	
	in revision	<input type="checkbox"/>	
	accepted	<input type="checkbox"/>	
	published	<input type="checkbox"/>	
Journal			
Year of publication			
Description of your own contribution in the case of joint authorship:			
<ul style="list-style-type: none"> - partly responsible for the study's conception and design - mainly responsible for data collection - processing, analysing and interpreting data - mainly responsible for drafting and revising the manuscript - marked as corresponding author 			

Place, Date

Signature of the doctoral candidate