

Title: 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^{3, 2}, Ima Trempler^{1, 2}, Karen Zentgraf⁴, Markus Raab^{5, 6}, Ricarda I. Schubotz^{1, 2}

¹Department of Psychology, University of Münster, Germany, ²Otto Creutzfeldt Center for Cognitive and Behavioural Neuroscience, University of Münster, Germany, ³Institute for Biomagnetism and Biosignalanalysis, University of Münster, Germany, ⁴Institute of Sports Sciences, Faculty of Psychology and Sports Sciences, Goethe University Frankfurt am Main, Germany, ⁵Institute of Psychology, German Sport University Cologne, Germany, ⁶School of Applied Sciences, London South Bank University, United Kingdom

Submitted to Journal:
Frontiers in Neuroscience

Specialty Section:
Auditory Cognitive Neuroscience

Article type:
Original Research Article

Manuscript ID:
527741

Received on:
17 Jan 2020

Revised on:
17 Apr 2020

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

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

Author contribution statement

NH, JP, DSK and RIS contributed conception and design of the study; NH performed the statistical analysis and wrote the first draft of the manuscript; IT and RIS wrote sections of the manuscript. KZ, MR and RIS contributed with scientific support, supervision, and coordination. All authors contributed to manuscript revision, read and approved the submitted version.

Keywords

human action sounds, auditory action effects, Action-effect association, Auditory prediction, Sensory Attenuation

Abstract

Word count: 333

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.

Contribution to the field

Dear Editor, Please find enclosed our manuscript, "Incidental or intentional? Different brain responses to one's own action sounds in hurdling vs. tap dancing", which we would like to submit for publication as a full length article. Our study provides interesting new insights on action sounds and their relevance for evaluating action performance. 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). 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. Due to the relevance of our study for the understanding of neural processes underlying action sound processing, its predictive mechanisms and relevance for action performance evaluation, we believe our findings are likely to appeal to the readership of Frontiers in Neuroscience. We look forward to your correspondence. Sincerely, Nina Heins (on behalf of all co-authors)

Ethics statements

Studies involving animal subjects

Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects

Generated 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.

Inclusion of identifiable human data

Generated Statement: No potentially identifiable human images or data is presented in this study.

In review

Data availability statement

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

In review

1 **Incidental or intentional? Different brain responses to one's own**
2 **action sounds in hurdling vs. tap dancing**

3

4 Authors: Nina Heins^{a,b}, Jennifer Pomp^{a,b}, Daniel S. Kluger^{b,c}, Ima Trempler^{a,b}, Karen Zentgraf^d,
5 Markus Raab^{e,f}, Ricarda I. Schubotz^{a,b}

6

7 ^aDepartment of Psychology, University of Muenster, Germany

8 ^bOtto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Muenster,
9 Germany

10 ^cInstitute for Biomagnetism and Biosignal Analysis, University Hospital Muenster, Germany

11 ^dDepartment of Movement Science and Training in Sports, Institute of Sport Sciences, Goethe
12 University Frankfurt, Germany

13 ^eInstitute of Psychology, German Sport University Cologne, Germany

14 ^fSchool of Applied Sciences, London South Bank University, United Kingdom

15

16 Corresponding author:

17 Nina Heins

18 Fliednerstr. 21

19 48149 Münster

20 nina.heins@uni-muenster.de

21

22

23

24

25

26

27

28

29

30

1 Abstract

2 Most human actions produce concomitant sounds. Action sounds can be either part of
3 the action goal (GAS, goal-related action sounds), as for instance in tap dancing, or a mere by-
4 product of the action (BAS, by-product action sounds), as for instance in hurdling. It is currently
5 unclear whether these two types of action sounds - incidental or intentional - differ in their
6 neural representation and whether the impact on the performance evaluation of an action
7 diverges between the two. We here examined whether during the observation of tap dancing
8 compared to hurdling, auditory information is a more important factor for positive action quality
9 ratings. Moreover, we tested whether observation of tap dancing vs. hurdling led to stronger
10 attenuation in primary auditory cortex, and a stronger mismatch signal when sounds do not
11 match our expectations.

12 We recorded individual point-light videos of newly trained participants performing tap
13 dancing and hurdling. In the subsequent functional magnetic resonance imaging (fMRI)
14 session, participants were presented with the videos that displayed their own actions, including
15 corresponding action sounds, and were asked to rate the quality of their performance. Videos
16 were either in their original form or scrambled regarding the visual modality, the auditory
17 modality, or both. As hypothesized, behavioral results showed significantly lower rating scores
18 in the GAS condition compared to the BAS condition when the auditory modality was
19 scrambled. Functional MRI contrasts between BAS and GAS actions revealed higher activation
20 of primary auditory cortex in the BAS condition, speaking in favor of stronger attenuation in
21 GAS, as well as stronger activation of posterior superior temporal gyri and the supplementary
22 motor area in GAS.

23 Results suggest that the processing of self-generated action sounds depends on whether
24 we have the intention to produce a sound with our action or not, and action sounds may be more
25 prone to be used as sensory feedback when they are part of the explicit action goal. **Our findings**
26 **contribute to a better understanding** of the function of action sounds for learning and controlling
27 sound-producing actions.

28

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

30

31

32

33

34

1 Introduction

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

18 In the present fMRI study, we addressed the potential dissociation of BAS and GAS
19 actions in the framework of predictive coding, suggesting that the brain works as a predictive
20 device and is tuned to minimize its prediction errors (Friston, 2005). According to this model,
21 action sounds are part of the predictive model that is engaged during action execution, or even
22 the observation thereof (Friston, 2012). Neural responses in primary sensory cortices are
23 attenuated for predicted self-generated sensations and the evidence is especially vast for self-
24 initiated sounds (re-aferences, Baess et al., 2011; Kennel et al., 2015; Mifsud et al., 2016;
25 Pizzera and Hohmann, 2015; Rummell et al., 2016; Timm et al., 2016), enabling the immediate
26 registration of prediction-deviant sensations (prediction errors) and effective correction of
27 sound-producing actions (Tourville et al., 2008). With regard to its neural underpinnings,
28 prediction of self-produced sounds is considered to rely on a network consisting of the primary
29 auditory cortex (Heschl's Gyrus) and the posterior superior temporal gyrus (pSTG) (Heilbron
30 and Chait, 2018; Rauschecker, 2013) and potentially also the supplementary motor area (SMA)
31 (Jo et al., 2019) and the cerebellum (Petrini et al., 2011; Waszak et al., 2012). The latter three
32 structures are suggested to deliver a predictive sound model to primary auditory cortex, causing
33 an attenuation of responses to expected sounds, and mismatch signals for unexpected sounds.

1 We examined potential differences between GAS and BAS as operationalized by tap
2 dancing and hurdling, respectively. In both dancing and hurdling, the lower limbs are the
3 effectors of audible sounds, and sounds produced by the lower limbs seem to show the same
4 prediction-driven sensory attenuation as the more thoroughly examined sounds produced by
5 our hands (Van Elk et al., 2014). We trained naïve participants and filmed them during motor
6 execution to create point-light displays with accompanying action sounds for both tap dancing
7 and hurdling. During a subsequent fMRI session, the same participants were presented the
8 point-light videos of their own actions and asked to rate the subjective quality of their actions
9 after each video on a 6-point Likert scale. To separately investigate the impact of visual and
10 auditory information on BAS and GAS action processing, we additionally introduced different
11 types of "scrambling" to the action videos serving as selective baseline conditions. Scrambling
12 was applied to either the visual modality, the auditory modality, or both. While leaving the
13 biological motion visually and audibly perceivable, scrambling strongly reduced the
14 information about the quality of action performance. On the behavioral level, we expected
15 overall lower rating scores for auditory scrambled videos, and due to the presumed greater
16 importance of auditory feedback in GAS actions, we hypothesized that this effect would be
17 especially pronounced in the tap-dancing condition.

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

2 Material and methods

2.1 Participants

The original sample consisted of 19 participants. One participant left the study before finishing the nine-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) 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, Ossmy, & Mukamel, 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.

2.2 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 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.

- Insert Figure 1 here -

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 Fig. 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.

1 Sound data were processed using Reaper v5.28 (Cockos Inc., New York, United States).
2 In a first step, stimulus intensities of hurdling and tap-dancing recordings were normalized
3 separately. In order to equalize the spectral distributions of both types of recordings, the
4 frequency profiles of hurdling and tap-dancing sounds were then captured using the Reaper
5 plugin Ozone 5 (iZotope Inc, Cambridge, United States). Finally, the difference curve (hurdling
6 – tap-dancing) was used by the plugin’s match function to adjust the tap-dancing spectrum to
7 the hurdling reference (see Supplementary Material, Fig. 1; examples of the sounds are given
8 in the videos of the Supplementary Material). Point-light displays and sound were
9 synchronized, and the subsequent videos were cut using Adobe Premiere Pro CC (Adobe
10 Systems Software, Dublin, Ireland). The final videos had a size of 640x400 pixels, 25 frames
11 per second, and an audio rate of 44 100 Hz. A visual fade-in and fade-out of 1 s (25 frames)
12 were added with Adobe Premiere. Video length ranged from 3 to 6 s, with an average length of
13 5 s.

14
15 - Insert Figure 2 here -
16

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

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

26 27 2.3 Procedure

28 2.3.1 Training and filming sessions

29 Participants engaged in a nine-week training period during which they were trained in
30 hurdling and tap dancing by professional instructors (Fig. 3). The training in both hurdling and
31 tap dancing was conducted two times a week, with each training session having a length of 90
32 minutes, so that participants trained both action types for three hours a week. Before this
33 training, none of the participants ever practiced hurdling or tap dancing. During the nine-weeks
34 training period, participants had to take part in four filming sessions, taking place at different

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

11

12

- Insert Figure 3 here -

13

14 2.3.2 Behavioral test and retest sessions

15 Behavioral test-retest sessions were conducted to find the videos with the highest
16 reliability of participants' rating. Both sessions were conducted in a computer lab in the
17 Department of Psychology at the University of Muenster. Participants were seated in front of a
18 computer and instructed to rate the quality of their actions on a scale from 1 (*“not well at all”*)
19 to 6 (*“very well”*) based on their subjective impression. The instructions were kept intentionally
20 liberal as to not influence participants to favor specific aspects of the action for their evaluation.
21 The experiment consisted of two blocks with self-paced responses, both lasting between 20 to
22 30 minutes. The same videos were presented in a different order in the second block of the
23 experiment. Videos were pseudorandomized so that not more than three videos in a row showed
24 the same action type (hurdling vs. tap dancing). Overall durations of the test session ranged
25 from 40 to 60 minutes, depending on the participants' response speed. Two weeks after the test
26 session, participants were presented the same videos once more (in pseudo-randomized order).
27 Twenty-seven videos for both hurdling and tap dancing were chosen per participant and were
28 used in the subsequent fMRI session. The videos with the highest reliability in rating were
29 chosen. Every video was rated a total of four times (two times in the test and two times in the
30 retest sessions). Of all chosen videos (702 videos in total, 54 per participant), 23.79 % received
31 the same rating on all four repetitions, in 69.8 % ratings varied by a score of either +1 or -1 in
32 one or two of the repetitions, and in 6.41 % ratings varied by a score of ± 1 .

33

1 2.3.3 fMRI session

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

9 Transition probabilities ensured that every condition was preceded by every condition
10 (including the same condition) in the same number of trials over the whole experiment. The
11 first trial of a block was a repetition of the last trial of the preceding block, to avoid losing a
12 transition. The remaining 27 trials consisted of 3 trials for each of the nine conditions. With the
13 first trial after each pause discarded, 243 trials remained, 216 video trials and 27 null events,
14 where a fixation cross was presented (27 trials for each of the nine conditions). The duration of
15 the null events was fixed at 5 s. Before every trial, a fixation cross was presented as an
16 interstimulus interval, varying between 3 to 4 s in length. After every video trial, the six-point
17 rating scale, including the rating question, was presented. The experiment continued upon the
18 participants’ button press.

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

22 2.4 fMRI recordings and preprocessing

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

30 Gradient-echo echoplanar imaging (EPI) was used to measure blood-oxygen-level-
31 dependent (BOLD) contrast for functional imaging data of the whole brain. There were 10 EPI
32 sequences in total. One sequence for the volume adjustment and one sequence for each of the
33 nine experimental blocks. Scanning parameters were set to a TE of 30 msec, a TR of 2000 msec,
34 a flip angle of 90°, 33 slices with a slice thickness of 3 mm, and a FoV of 192x192 mm².

1 Imaging data were processed using SPM12 (Wellcome Trust, London, England).
2 Preprocessing consisted of slice time correction to the middle slice, realignment to the mean
3 image, co-registration of the functional data to the individual's structural scan, normalization
4 into the standard MNI space (Montreal Neurological Institute, Montreal, QC, Canada) based on
5 segmentation parameters, and spatial smoothing with a Gaussian kernel of full-width at half
6 maximum (FWHM) of 8 mm. A high-pass temporal filter equivalent to 128 s was applied to
7 the data.

9 2.5 Statistical data analysis

10 2.5.1 Behavioral data analysis

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

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

22 2.5.2 fMRI design specification

23 The design was implemented in SPM12, following a general linear model approach
24 (GLM, Friston et al., 1994; Worsley and Friston, 1995). The modelled activation was time-
25 locked to the onsets of the videos or null events. Epochs contained the full presentation period
26 ranging from 3 to 6 s for the videos, and 5 s for the null events. Since tap dancing and hurdling
27 differ with regard to their auditory event density, i.e. the number of distinguishable auditory
28 sounds occurring per second, we controlled for this source of variance by introducing regressors
29 of nuisance. To this end, we used the MIR Toolbox (Lartillot, Toivainen, & Eerola, 2008) to
30 calculate the action sounds' event densities. The GLM consisted of 23 regressors in total: eight
31 regressors for the experimental conditions, eight parametric regressors modelling the event
32 densities for each of the eight experimental conditions, one regressor for the null events, and
33 six regressors for the motion parameters (three translations and three rotations). Activation for
34 27 trials was considered for the modeling of each of the experimental regressors as well as for

1 the null event regressor. All regressors were convolved with the hemodynamic response
2 function.

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

8 First, t-contrasts for the unscrambled conditions were calculated (BAS_normal >
9 GAS_normal) to assess basic differences between tap dancing and hurdling precluding potential
10 effects of visual scrambling on the use of auditory information. Resulting t-maps were corrected
11 using false discovery rate (FDR) correction and a threshold of $p < .05$. Additionally, we defined
12 F-contrasts for the main effect of ACTION, thus including both the normal and the scrambled
13 condition (BAS > GAS), and for the ACTION x SOUND interaction effect. T-tests were calculated
14 to examine the direction of effects. Region of interest (ROI) analyses were performed to test
15 our anatomically specified hypotheses, using FDR-correction with a threshold of $p < .001$.
16 Structural ROIs were defined using the automated anatomical labelling (AAL) atlas and created
17 using the WFU PickAtlas toolbox (Maldjian et al., 2003) in SPM12. Firstly, we performed a
18 ROI analysis for the primary auditory cortex (Heschl's gyrus) for the BAS>GAS contrast, to
19 test for the hypothesized stronger sensory attenuation for tap dancing than for hurdling.
20 Secondly, ROI analyses for the secondary auditory cortex (pSTG), the SMA, and cerebellum
21 for the GAS>BAS contrast were performed, to investigate a stronger activation for tap dancing
22 due to explicit sound predictions over and above visual predictions. Thirdly, we performed a
23 ROI analysis for the ACTION x SOUND interaction effect, using structural ROIs for the pSTG, to
24 examine the differential involvement of secondary auditory cortices in the sound-scrambled
25 versions of hurdling and tap dancing. We expected more activation in the tap-dancing condition
26 due to a more pronounced mismatch whenever the sound did not fit the perceived action. We
27 additionally extracted beta values from pSTG and the primary auditory cortex, to examine more
28 subtle differences between the conditions.

29 3 Results

30 3.1 Behavioral Results

31 The rating scores for all conditions were normally distributed, as checked by a
32 Kolmogorov-Smirnov test. For the analysis of the rating scores, a 2 x 2 x 2 within-subject
33 ANOVA with the factors ACTION (BAS, GAS), PICTURE (*picture normal*, *picture scrambled*)

1 and SOUND (*sound normal*, *sound scrambled*) was calculated. Our final sample size was $n = 13$.
2 We calculated a post-hoc power analysis using GPower (Erdfelder et al., 1996), determining a
3 critical F-value (defining the boundary for the rejection of the null hypothesis) of 4.75. All
4 observed effects were well above this critical value, indicating that our findings were reliable
5 even in this relatively small sample size.

6 First of all, we did not find a main effect for the factor ACTION ($F(1,12) = 0.009$, $p =$
7 $.928$), indicating that participants were not biased to rate either their hurdling or tap-dancing
8 performance as superior. Importantly, this balanced rating provides a solid basis to interpret
9 differences between BAS and GAS without a confounding bias by preference.

10 As hypothesized, there was a significant main effect for the factor SOUND ($F(1,12) =$
11 22.01 , $p < .001$, $\eta^2 = 0.647$), driven by lower rating for *sound scrambled* ($M = 3.13$, $SD = 0.6$,
12 Fig. 4B) than *sound normal* ($M = 3.67$, $SD = 0.53$). Likewise, a significant main effect for the
13 factor PICTURE ($F(1,12) = 11.86$, $p = .005$, $\eta^2 = 0.497$) was explained by higher rating scores
14 in the *picture normal* ($M = 3.71$, $SD = 0.53$) vs. *picture scrambled* condition ($M = 3.09$, $SD =$
15 0.7 , Fig. 4B).

16 There was a significant ACTION x SOUND interaction ($F(1,12) = 11.67$, $p = .005$, $\eta^2 =$
17 0.493). Paired *t*-tests revealed lower ratings for *sound scrambled* vs. *normal* for hurdling ($M =$
18 3.23 , $SD = 0.61$ vs. $M = 3.55$, $SD = 0.49$; $t(12) = 2.91$, $p = .013$) as well as for tap dancing (M
19 $= 3.03$, $SD = 0.82$ vs. $M = 3.79$, $SD = 0.81$; $t(12) = 5.08$, $p < .001$). The three-way interaction
20 between ACTION, PICTURE and SOUND reached significance ($F(1,12) = 6.66$, $p = .024$, $\eta^2 = 0.357$;
21 for post-hoc *t*-tests, see Table 1 of the Supplementary Material), corroborating that the impact
22 of auditory scrambling was stronger on tap dancing than on hurdling.

23 We performed a post-experimental survey where we asked participants to rate on a 6-
24 point Likert scale how difficult it was for them to evaluate hurdling and tap-dancing videos.
25 The rating difficulty did not differ between hurdling ($M = 3.77$, $SD = 1.17$) and tap-dancing
26 videos ($M = 3.54$, $SD = 1.13$, $t(12) = 0.507$, $p = .621$).

27 Finally, as for the fMRI data, we aimed to control for potential confounds of the
28 behavioral rating by stimulus event density. To this end, we tested whether the rating scores
29 correlated with the event density of the corresponding condition (all $p > .05$), implementing the
30 same $2 \times 2 \times 2$ within-subject ANOVA for the event density values. Here, we did not find a
31 significant ACTION x SOUND interaction (ACTION x SOUND, $F(1,12) = 0.34$, $p = .573$).

32
33
34 - Insert Figure 4 here -
35

1 3.2 fMRI results

2 The whole brain contrast BAS_normal>GAS_normal yielded higher activity in right
3 primary auditory cortex (hypothesis 1) as well as in the occipital pole. Corroborating hypothesis
4 2, the reverse contrast GAS_normal>BAS_normal showed significant effects in SMA and right
5 pSTG (Fig. 5 and Tab. 1). Note that the same significant effects were observed when contrasting
6 BAS and GAS aggregated for all their sub-conditions, i.e., effects were independent of auditory
7 or visual scrambling.

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

13 The ROI analysis for the interaction effect between ACTION and SOUND, testing for a
14 more pronounced prediction error in the auditorily scrambled GAS condition, did not show
15 significant results in pSTG. Beta value extraction was performed to follow up on this, yielding
16 a non-significant trend ($p = .061$) for the interaction effect (Fig. 2 of Supplementary Material).
17 Additionally, we also extracted beta values from primary auditory cortex, to determine whether
18 auditory scrambling of GAS would offset attenuation by increasing the prediction error. Indeed,
19 we found a non-significant trend for the GAS_sound_scrambled>GAS_sound_normal
20 comparison ($p = .071$).

21
22 - Insert Figure 5 here -
23

24 **Table 1 – Whole-brain activation of the main effects of ACTION**

	x	y	z	t-value	voxels
BAS_normal>GAS_normal					
Occipital pole	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

25 Regions of activation, MNI (Montreal Neurological Institute) coordinates (x, y, z), t -values for the local maxima
26 (FDR-corrected, $p < .05$), activation extent in voxels (clusters larger than $k=20$).

27

1

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

2

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

6

7

8

9 4 Discussion

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

1 To determine the impact of visual and auditory information on the evaluation of one's
2 own performance, we manipulated the visual and the auditory information by scrambling. As
3 expected, both visual scrambling and auditory scrambling led to a significant reduction of the
4 rating scores for hurdling and tap-dancing performance, with scores being the lowest when both
5 modalities were scrambled at the same time. In line with our predictions, auditory scrambling
6 had a stronger impact on the rating of tap-dancing than hurdling performance, **with rating scores**
7 **decreasing in both actions, but to a stronger degree in tap dancing.** This finding supports the
8 particular relevance of action sounds as an error-monitoring tool in GAS actions (Murgia et al.,
9 2017). While auditory scrambling had an effect on the rating of the hurdling trials as well, the
10 effect was more pronounced in tap dancing, showing that, with the auditory output being an
11 explicit action goal, perception of action quality was especially reduced by incoherent auditory
12 feedback. It is however important to note that action sounds were also important for the rating
13 of hurdling performance, although the created sounds were not explicitly intended. Together,
14 behavioral findings suggest that the brain generates predictions of how our actions should sound
15 like, both in case of tap-dancing (GAS actions) as well as hurdling (BAS actions). Subtle
16 differences in the level of interference, however, point to a more prominent role of auditory
17 expectations in the former.

18 In agreement with our hypotheses, we found stronger activation in Heschl's gyri for the
19 hurdling (BAS) compared to tap dancing (GAS) trials, reflecting a more pronounced sensory
20 attenuation when auditory action consequences are predominantly used in the predictive model
21 (GAS). Sensory attenuation to self-initiated sounds is based on an existing association between
22 the initiated movement and the resulting sound (Keyesers and Gazzola, 2014; Ticini et al., 2012).
23 Self-produced sounds elicit a smaller amplitude in early EEG or MEG components, presumably
24 due to the feeling of self-agency (Aliu et al., 2009; Baess et al., 2011; Timm et al., 2014). The
25 precise origin of sensory attenuation of self-produced sounds has not yet been completely
26 unraveled (Horváth, 2015; Hughes et al., 2013). A more recent study shows that sensory
27 attenuation is not solely due to the self-generation of action sounds, but relies on the
28 predictability of sensory input (Kaiser and Schütz-Bosbach, 2018). Self-generated sounds have
29 high predictability, as the sensory effects are part of our motor plan when initiating and
30 performing a movement (Shin et al., 2010). Accordingly, we suppose that tap-dancing sounds,
31 being an intentional part of our motor plan, are more efficiently attenuated by internal predictive
32 models, as reflected by stronger pSTG and SMA activity. In contrast, hurdling sounds may be
33 less relevant in the predictive model, and are therefore not attenuated to the same extend in
34 primary auditory cortices early cortical responses to the same extend as the tap-dancing sounds.

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

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

23 While BOLD contrasts did not confirm the differential effect of auditory scrambling on
24 tap-dancing and hurdling, beta estimates extracted from the pSTG did indicate a descriptive
25 interaction effect for the imaging data. Thus, the beta weights for hurdling did not differ
26 between the *sound normal* and *sound scrambled* condition, whereas there was a small difference
27 between these two conditions for tap dancing. Both these findings and the behavioral results
28 speak in favor of our interaction hypothesis: While auditory scrambling has an effect on both
29 hurdling and tap dancing, the effect on tap dancing is larger, indicating a greater relevance of
30 the auditory domain to positively evaluate action quality and a stronger predictive mismatch in
31 GAS actions.

32 The overall stronger activation in occipital visual areas for hurdling compared to tap-
33 dancing was not hypothesized. Obviously, evaluating one's hurdling performance yielded a
34 more extensive visual processing of the observed action, as indicated by increased BOLD

1 activity in occipito-temporal cortices (Jastorff et al., 2010). This also matches the behavioral
2 finding of a stronger impact of visual scrambling on the rating of hurdling quality.

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

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

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

32 To further clarify the role of action sounds for monitoring action performance, future
33 studies may examine the effects of deprivation and interference on the entire spectrum of sound-
34 producing actions, as has been done for language production, musical performance, and some

1 other actions (Howell, 2004; Kennel et al., 2015; Keough and Jones, 2009; Pfordresher and
2 Beasley, 2014). Our study is a first step into a more systematic approach to understanding action
3 sounds, while establishing ideas for additional research to deepen the comprehension of this
4 relevant topic. Both musicians and athletes might benefit from a better understanding of the
5 role of action sounds for optimizing action performance, giving them the opportunity to
6 adequately train their skills. Research in schizophrenia might also gain from a better insight
7 into the connections between action sounds and motor control, especially regarding the sense
8 of agency, and the failure to attribute self-produced sensations to oneself in people with
9 schizophrenia.

11 5 Conclusion

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

26 Acknowledgements

27 We would like to thank Monika Mertens for her help during data collection. Furthermore, we
28 thank Theresa Eckes, Alina Eisele, Marie Kleinbielen, Katharina Thiel and Stefan Vinbrück for
29 their assistance during filming and with creating the stimulus material. Finally, we would like
30 to thank Nadiya El-Sourani, Amelie Huebner, Klara Hagelweide, Laura Quante, Marlen Roehe
31 and Lena Schliephake for rewarding discussions, and our three reviewers for insightful input
32 on this manuscript.

1 References

- 2 Aliu, S.O., Houde, J.F., Nagarajan, S.S., 2009. Motor-induced suppression of the auditory
3 cortex. *J. Cogn. Neurosci.* 21, 791–802. <https://doi.org/10.1162/jocn.2009.21055>
- 4 Baess, P., Horváth, J., Jacobsen, T., Schröger, E., 2011. Selective suppression of self-initiated
5 sounds in an auditory stream: An ERP study. *Psychophysiology* 48, 1276–1283.
6 <https://doi.org/10.1111/j.1469-8986.2011.01196.x>
- 7 Bischoff, M., Zentgraf, K., Pilgramm, S., Stark, R., Krüger, B., Munzert, J., 2014.
8 Anticipating action effects recruits audiovisual movement representations in the ventral
9 premotor cortex. *Brain Cogn.* 92, 39–47. <https://doi.org/10.1016/j.bandc.2014.09.010>
- 10 Erdfelder, E., Faul, F., Buchner, A., 1996. GPOWER: A general power analysis program.
11 *Behav. Res. Methods, Instruments Comput.* <https://doi.org/10.3758/BF03203630>
- 12 Finney, S.A., 1997. Auditory Feedback and Musical Keyboard Performance. *Music Percept.*
13 *An Interdiscip. J.* 15, 153–174.
- 14 Friston, K., 2012. Prediction, perception and agency. *Int. J. Psychophysiol.*
15 <https://doi.org/10.1016/j.ijpsycho.2011.11.014>
- 16 Friston, K., 2005. A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 815–
17 836. <https://doi.org/10.1098/rstb.2005.1622>
- 18 Friston, K.J., Daunizeau, J., Kilner, J., Kiebel, S.J., 2010. Action and behavior: a free-energy
19 formulation. *Biol. Cybern.* 102, 227–260. <https://doi.org/10.1007/s00422-010-0364-z>
- 20 Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J. -P, Frith, C.D., Frackowiak, R.S.J.,
21 1994. Statistical parametric maps in functional imaging: A general linear approach.
22 *Hum. Brain Mapp.* 2, 189–210. <https://doi.org/10.1002/hbm.460020402>
- 23 Fu, C.H.Y., Vythelingum, G.N., Brammer, M.J., Williams, S.C.R., Amaro, E., Andrew, C.M.,
24 Yágüez, L., Van Haren, N.E.M., Matsumoto, K., McGuire, P.K., 2006. An fMRI study
25 of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cereb. Cortex*
26 16, 969–977. <https://doi.org/10.1093/cercor/bhj039>
- 27 Gates, A., Bradshaw, J.L., Nettleton, N.C., 1974. Effect of different delayed auditory
28 feedback intervals on a music performance task. *Percept. Psychophys.* 15, 21–25.
29 <https://doi.org/10.3758/BF03205822>
- 30 Heilbron, M., Chait, M., 2018. Great Expectations: Is there Evidence for Predictive Coding in
31 Auditory Cortex? *Neuroscience* 389, 54–73.
32 <https://doi.org/10.1016/j.neuroscience.2017.07.061>
- 33 Herrington, J.D., Nymberg, C., Schultz, R.T., 2011. Biological motion task performance
34 predicts superior temporal sulcus activity. *Brain Cogn.* 77, 372–381.

- 1 <https://doi.org/10.1016/j.bandc.2011.09.001>
- 2 Horváth, J., 2015. Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain*
- 3 *Res.* <https://doi.org/10.1016/j.brainres.2015.03.038>
- 4 Howell, P., 2004. Effects of delayed auditory feedback and frequency-shifted feedback on
- 5 speech control and some potentials for future development of prosthetic aids for
- 6 stammering. *Stammering Res.* 1, 31–46.
- 7 <https://doi.org/10.1126/science.1182395.Evolution>
- 8 Hughes, G., Desantis, A., Waszak, F., 2013. Mechanisms of intentional binding and sensory
- 9 attenuation: The role of temporal prediction, temporal control, identity prediction, and
- 10 motor prediction. *Psychol. Bull.* 139, 133–151. <https://doi.org/10.1037/a0028566>
- 11 Jastorff, J., Begliomini, C., Rizzolatti, G., Orban, G.A., Fabbri-Destro, M., Jastorff, J., 2010.
- 12 Coding Observed Motor Acts: Different Organizational Principles in the Parietal and
- 13 Premotor Cortex of Humans. *J. Neurophysiol.* 104, 128–140.
- 14 <https://doi.org/10.1152/jn.00254.2010>
- 15 Jo, H.G., Habel, U., Schmidt, S., 2019. Role of the supplementary motor area in auditory
- 16 sensory attenuation. *Brain Struct. Funct.* <https://doi.org/10.1007/s00429-019-01920-x>
- 17 Jones, J.A., Keough, D., 2008. Auditory-motor mapping for pitch control in singers and
- 18 nonsingers. *Exp. Brain Res.* 190, 279–287. <https://doi.org/10.1007/s00221-008-1473-y>
- 19 Kaiser, J., Schütz-Bosbach, S., 2018. Sensory attenuation of self-produced signals does not
- 20 rely on self-specific motor predictions. *Eur. J. Neurosci.* 47, 1303–1310.
- 21 <https://doi.org/10.1111/ejn.13931>
- 22 Kennel, C., Streese, L., Pizzera, A., Justen, C., Hohmann, T., Raab, M., 2015. Auditory
- 23 reafferences: The influence of real-time feedback on movement control. *Front. Psychol.*
- 24 6, 1–6. <https://doi.org/10.3389/fpsyg.2015.00069>
- 25 Keough, D., Jones, J.A., 2009. The sensitivity of auditory-motor representations to subtle
- 26 changes in auditory feedback while singing. *J. Acoust. Soc. Am.* 126, 837–846.
- 27 <https://doi.org/10.1121/1.3158600>
- 28 Keyzers, C., Gazzola, V., 2014. Hebbian learning and predictive mirror neurons for actions,
- 29 sensations and emotions. *Philos. Trans. R. Soc. B Biol. Sci.* 369.
- 30 <https://doi.org/10.1098/rstb.2013.0175>
- 31 Lartillot, O., Toivainen, P., Eerola, T., 2008. A Matlab Toolbox for Music Information
- 32 Retrieval, in: Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R. (Eds.), *Data*
- 33 *Analysis, Machine Learning and Applications*. Springer Berlin Heidelberg, Berlin,
- 34 Heidelberg, pp. 261–268.

- 1 Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H., 2003. An automated method for
2 neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets 19,
3 1233–1239. [https://doi.org/10.1016/S1053-8119\(03\)00169-1](https://doi.org/10.1016/S1053-8119(03)00169-1)
- 4 Menzer, F., Brooks, A., Halje, P., Faller, C., Vetterli, M., Blanke, O., 2010. Feeling in control
5 of your footsteps: Conscious gait monitoring and the auditory consequences of footsteps.
6 *Cogn. Neurosci.* 1, 184–192. <https://doi.org/10.1080/17588921003743581>
- 7 Miall, R.C., Wolpert, D.M., 1996. Forward models for physiological motor control. *Neural*
8 *Networks*. [https://doi.org/10.1016/S0893-6080\(96\)00035-4](https://doi.org/10.1016/S0893-6080(96)00035-4)
- 9 Mifsud, N.G., Oestreich, L.K.L., Jack, B.N., Ford, J.M., Roach, B.J., Mathalon, D.H.,
10 Whitford, T.J., 2016. Self-initiated actions result in suppressed auditory but amplified
11 visual evoked components in healthy participants. *Psychophysiology* 53.
12 <https://doi.org/10.1111/psyp.12605>
- 13 Murgia, M., Prpic, V., Jenny, O., McCullagh, P., Santoro, I., Galmonte, A., Agostini, T.,
14 2017. Modality and perceptual-motor experience influence the detection of temporal
15 deviations in tap dance sequences. *Front. Psychol.* 8, 1–8.
16 <https://doi.org/10.3389/fpsyg.2017.01340>
- 17 Oldfield, R.C., 1971. The assessment and analysis of handedness. *Neuropsychologia* 9, 97–
18 113.
- 19 Petrini, K., Pollick, F.E., Dahl, S., McAleer, P., McKay, L., Rocchesso, D., Waadeland, C.H.,
20 Love, S., Avanzini, F., Puce, A., 2011. Action expertise reduces brain activity for
21 audiovisual matching actions: An fMRI study with expert drummers. *Neuroimage* 56,
22 1480–1492. <https://doi.org/10.1016/j.neuroimage.2011.03.009>
- 23 Pfordresher, P.Q., Beasley, R.T.E., 2014. Making and monitoring errors based on altered
24 auditory feedback. *Front. Psychol.* 5, 1–13. <https://doi.org/10.3389/fpsyg.2014.00914>
- 25 Phillips, H.N., Blenkman, A., Hughes, L.E., Bekinschtein, T.A., Rowe, J.B., 2015.
26 Hierarchical organization of frontotemporal networks for the prediction of stimuli across
27 multiple dimensions. *J. Neurosci.* 35, 9255–9264.
28 <https://doi.org/10.1523/JNEUROSCI.5095-14.2015>
- 29 Pizzera, A., Hohmann, T., 2015. Acoustic Information During Motor Control and Action
30 Perception: A Review. *Open Psychol. J.* 8, 183–191.
31 <https://doi.org/10.2174/1874350101508010183>
- 32 Rauschecker, J.P., 2013. Neural Correlates of Auditory Cognition.
33 <https://doi.org/10.1007/978-1-4614-2350-8>
- 34 Reznik, Daniel, Henkin, Y., Levy, O., Mukamel, R., 2015. Perceived loudness of self-

- 1 generated sounds is differentially modified by expected sound intensity. *PLoS One* 10.
2 <https://doi.org/10.1371/journal.pone.0127651>
- 3 Reznik, D., Ossmy, O., Mukamel, R., 2015. Enhanced Auditory Evoked Activity to Self-
4 Generated Sounds Is Mediated by Primary and Supplementary Motor Cortices. *J.*
5 *Neurosci.* 35, 2173–2180. <https://doi.org/10.1523/JNEUROSCI.3723-14.2015>
- 6 Rummell, B.P., Klee, J.L., Sigurdsson, T., 2016. Attenuation of responses to self-generated
7 sounds in auditory cortical neurons. *J. Neurosci.* 36.
8 <https://doi.org/10.1523/JNEUROSCI.1564-16.2016>
- 9 Schaffert, N., Janzen, T.B., Mattes, K., Thaut, M.H., 2019. A Review on the Relationship
10 Between Sound and Movement in Sports and Rehabilitation. *Front. Psychol.* 10, 244.
11 <https://doi.org/10.3389/fpsyg.2019.00244>
- 12 Schröger, E., Kotz, S.A., SanMiguel, I., 2015a. Bridging prediction and attention in current
13 research on perception and action. *Brain Res.*
14 <https://doi.org/10.1016/j.brainres.2015.08.037>
- 15 Schröger, E., Marzecová, A., Sanmiguel, I., 2015b. Attention and prediction in human
16 audition: A lesson from cognitive psychophysiology. *Eur. J. Neurosci.* 41, 641–664.
17 <https://doi.org/10.1111/ejn.12816>
- 18 Shin, Y.K., Proctor, R.W., Capaldi, E.J., 2010. A Review of Contemporary Ideomotor
19 Theory. *Psychol. Bull.* 136, 943–974. <https://doi.org/10.1037/a0020541>
- 20 Ticini, L.F., Schütz-Bosbach, S., Weiss, C., Casile, A., Waszak, F., 2012. When sounds
21 become actions: Higher-order representation of newly learned action sounds in the
22 human motor system. *J. Cogn. Neurosci.* 24, 464–474.
23 https://doi.org/10.1162/jocn_a_00134
- 24 Timm, J., Schönwiesner, M., SanMiguel, I., Schröger, E., 2014. Sensation of agency and
25 perception of temporal order. *Conscious. Cogn.* 23, 42–52.
26 <https://doi.org/10.1016/j.concog.2013.11.002>
- 27 Timm, J., Schönwiesner, M., Schröger, E., SanMiguel, I., 2016. Sensory suppression of brain
28 responses to self-generated sounds is observed with and without the perception of
29 agency. *Cortex* 80, 5–20. <https://doi.org/10.1016/j.cortex.2016.03.018>
- 30 Tourville, J.A., Reilly, K.J., Guenther, F.H., 2008. Neural mechanisms underlying auditory
31 feedback control of speech. *Neuroimage* 39, 1429–1443.
32 <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- 33 Van Elk, M., Salomon, R., Kannape, O., Blanke, O., 2014. Suppression of the N1 auditory
34 evoked potential for sounds generated by the upper and lower limbs. *Biol. Psychol.* 102,

- 1 108–117. <https://doi.org/10.1016/j.biopsycho.2014.06.007>
- 2 Waszak, F., Cardoso-Leite, P., Hughes, G., 2012. Action effect anticipation:
3 Neurophysiological basis and functional consequences. *Neurosci. Biobehav. Rev.* 36,
4 943–959. <https://doi.org/10.1016/j.neubiorev.2011.11.004>
- 5 Wollman, I., Morillon, B., 2018. Organizational principles of multidimensional predictions in
6 human auditory attention. *Sci. Rep.* 8, 1–11. <https://doi.org/10.1038/s41598-018-31878-5>
- 7 Wolpert, D.M., Ghahramani, Z., Jordan, M.I., 1995. An internal model for sensorimotor
8 integration. *Science* (80-.). 269, 1880–1882. <https://doi.org/10.1126/science.7569931>
- 9 Woods, E.A., Hernandez, A.E., Wagner, V.E., Beilock, S.L., 2014. Expert athletes activate
10 somatosensory and motor planning regions of the brain when passively listening to
11 familiar sports sounds. *Brain Cogn.* 87, 122–133.
12 <https://doi.org/10.1016/j.bandc.2014.03.007>
- 13 Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited — Again.
14 *Neuroimage.* <https://doi.org/10.1006/nimg.1995.1023>
- 15 Wutte, M.G., Glasauer, S., Jahn, K., Flanagan, V.L., 2012. Moving and being moved:
16 Differences in cerebral activation during recollection of whole-body motion. *Behav.*
17 *Brain Res.* 227, 21–29. <https://doi.org/10.1016/j.bbr.2011.09.042>
- 18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

1 **Fig. 1. Camera positions and set-up during the point light recordings.** (A) Camera positions
2 during hurdling from a top view perspective. Green dots represent the cameras, red lines the
3 hurdles, and the yellow arrow the hurdling track. (B) Camera positions during tap dancing from
4 a top view perspective. Green dots represent the cameras, the yellow square the area in which
5 the tap dancer performed the sequence (C) Set-up during the recording of hurdling. Three hurdle
6 transitions had to be performed during the recording. The two last hurdles are visible in the
7 figure above. The yellow arrow indicates the hurdling track.

8

9 **Fig. 2. Position of the point-light markers.** (A) Twelve point-light markers were used and
10 positioned at the shoulders, elbows, wrists, hips, knees, ankles, and the tips of the toes.
11 Exemplary videos can be found in the Supplementary Material (Video 1 – Video 8). (B) Three
12 snapshots of the hurdling action over the course of one video. (C) Three snapshots of the tap-
13 dancing action over the course of one video.

14

15 **Fig. 3. Procedure of the study.** Participants were filmed on several occasions during their
16 nine-week training in hurdling and tap dancing. Two behavioral sessions were conducted
17 before the MRI session.

18

19 **Fig. 4. Trial composition and action quality rating scores.** (A) A trial consisted of a video (3
20 – 6 s in length), followed by the video rating question (“How well did you perform the action?”
21 in German), and a fixation cross (3.5 – 4.5 s in length). The total duration of one trial was
22 approximately 10 s. (B) Mean rating scores for the evaluation of the quality of the action
23 performance presented in the observed videos, obtained during the MRI sessions. Rating scores
24 could range from 1 to 6 (1 representing a low, 6 a high rating of quality). Error bars show
25 standard errors. BAS conditions are represented in yellow, GAS conditions in blue. Vertical
26 stripes represent the scores for *picture-scrambled* conditions, whereas horizontal stripes
27 represent *sound-scrambled* conditions. Columns with both vertical and horizontal stripes
28 represent the conditions with both picture and sound scrambled.

29

30 **Fig. 5. Whole-brain activation of the main effects of action.** (A) FDR-corrected t-maps ($p <$
31 $.05$) for the $BAS_normal > GAS_normal$ contrast. (B) FDR-corrected t-maps ($p <$
32 $.05$) for the $GAS_normal > BAS_normal$ contrast.

33

34

Figure 1.TIFF

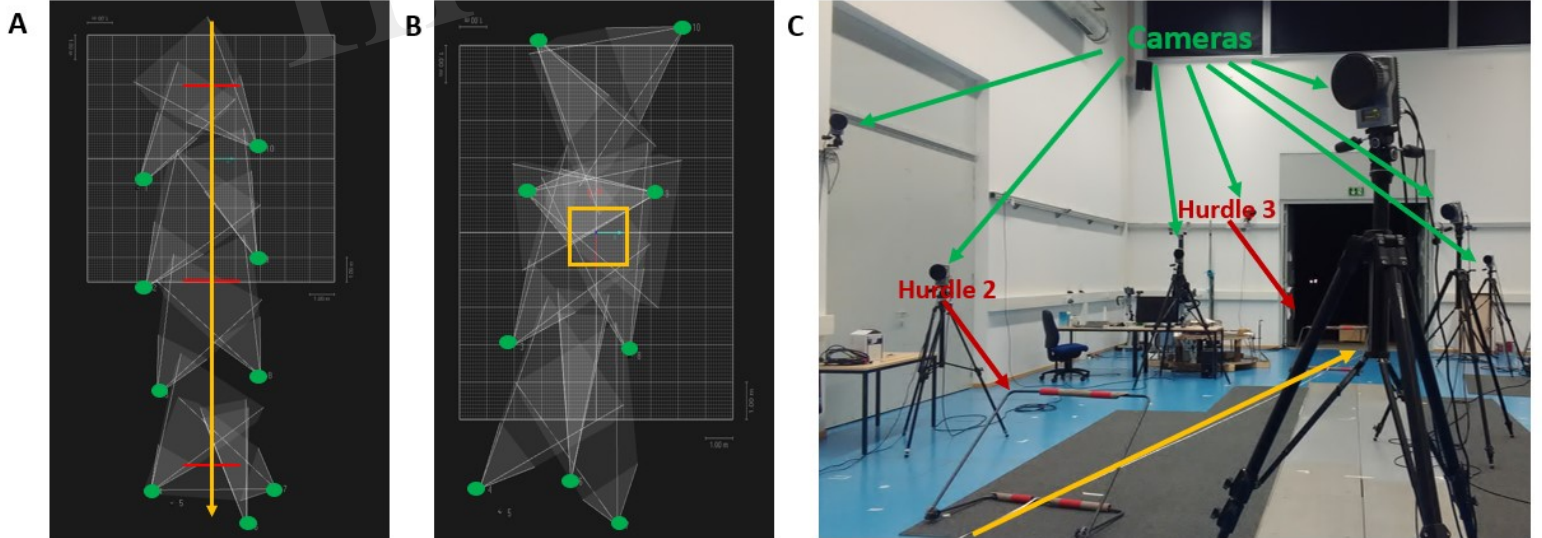


Figure 2.TIFF

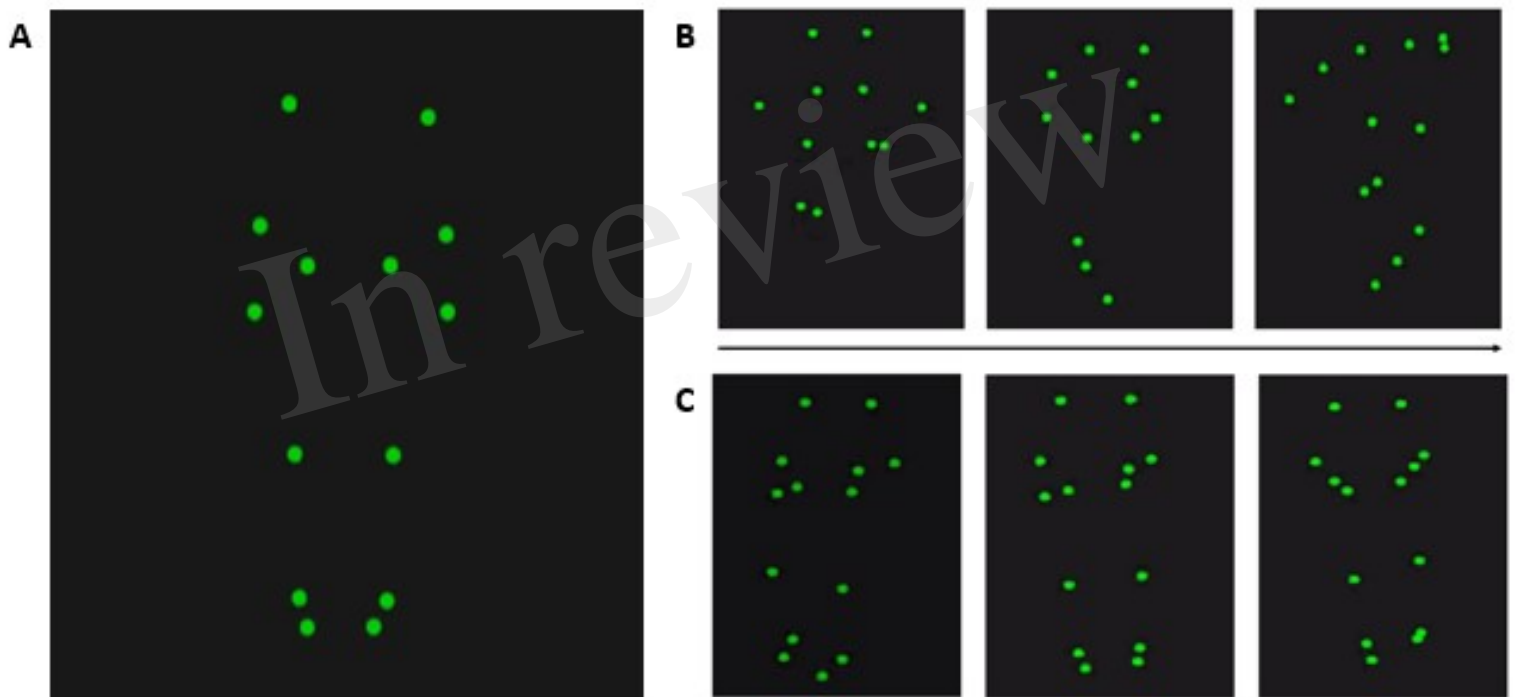


Figure 3.TIFF

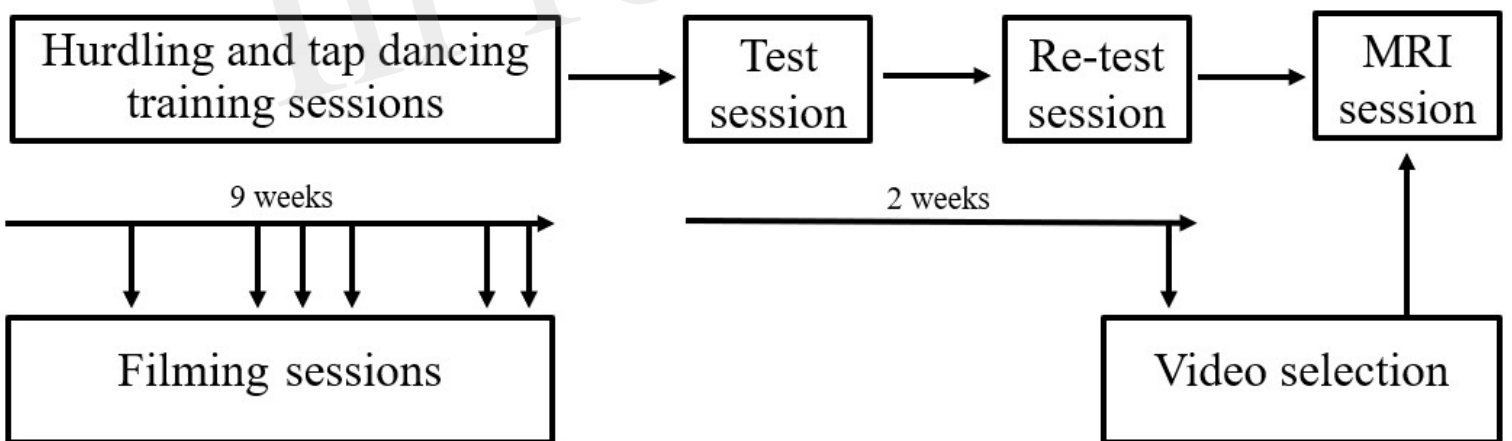
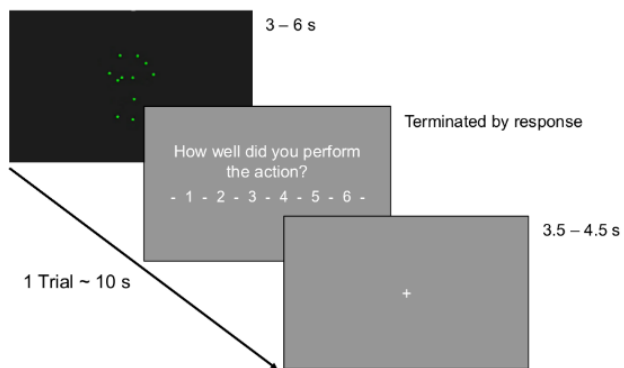


Figure 4.TIF

In review

A Trial composition



B Mean rating scores per condition

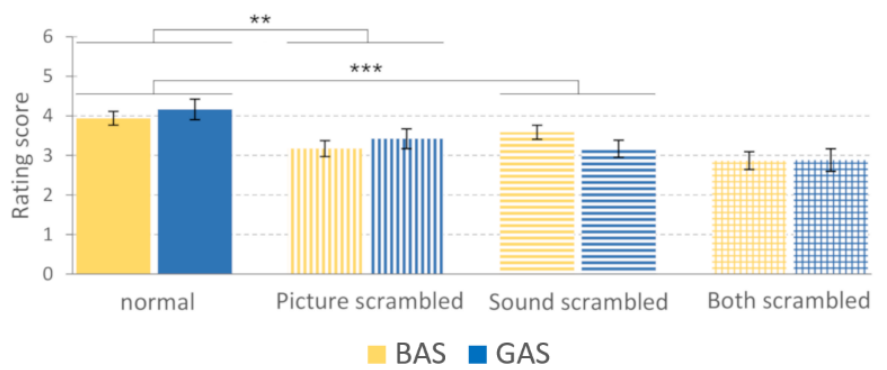


Figure 5.TIFF

