

ORIGINAL ARTICLE

Frontoparietal, Cerebellum Network Codes for Accurate Intention Prediction in Altered Perceptual Conditions

L. Ceravolo^{1,†}, S. Schaerlaeken^{1,†}, S. Frühholz^{2,3}, D. Glowinski^{1,‡} and D. Grandjean^{1,‡}

¹Neuroscience of Emotion and Affective Dynamics Lab, Department of Psychology and Educational Sciences and Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland, ²Department of Psychology, University of Zurich, Zurich, Switzerland and ³Department of Psychology, University of Oslo, Oslo, Norway

Address correspondence to Leonardo Ceravolo, PhD, Neuroscience of Emotion and Affective Dynamics laboratory, Faculté de Psychologie et des Sciences de l'Éducation, Université de Genève, 40 bd du Pont-d'Arve, 1205 Geneva, Switzerland. Email: leonardo.ceravolo@unige.ch;

ORCID ID: <https://orcid.org/0000-0003-0638-3981>

[†]Equal contribution as first authors

[‡]Equal contribution as senior authors

Abstract

Integrating and predicting the intentions and actions of others are critical components of social interactions, but the behavioral and neural bases of such mechanisms under altered perceptual conditions are poorly understood. In the present study, we recruited expert violinists and age-matched controls with no musical training and asked them to evaluate simplified dynamic stimuli of violinists playing in a *piano* or *forte* communicative intent while undergoing functional magnetic resonance imaging. We show that expertise is needed to successfully understand and evaluate communicative intentions in spatially and temporally altered visual representations of musical performance. Frontoparietal regions—such as the dorsolateral prefrontal cortex and the inferior parietal lobule and sulcus—and various subregions of the cerebellum—such as cerebellar lobules I-IV, V, VI, VIIb, VIIIa, X—a re recruited in the process. Functional connectivity between these brain areas reveals widespread organization, particularly in the dorsolateral prefrontal cortex, inferior frontal gyrus, inferior parietal sulcus, and in the cerebellum. This network may be essential to successfully assess communicative intent in ambiguous or complex visual scenes.

Key words: communication, functional connectivity, intention, neuroimaging

The human ability to coordinate with others is a key evolutionary skill that enables us to accomplish tasks that would otherwise be impossible to manage individually. In addition to vocal and semantic communication (Kotz and Schwartze 2010), such a mechanism relies on finely tuned nonverbal expressive behaviors that must be able to communicate one's intention reliably and efficiently (Jahng et al. 2017). Intention therefore

involves the entire body as a means of communication, with a focus on upper body actions and movement dynamics (Anderesen and Cui 2009). It requires that both parties share similar representations at different levels (e.g., a common goal or intermediate steps to reach a final goal), predict common outcomes, and integrate the predicted consequences of their own actions as well as those of others (Sebanz and Knoblich 2009). Thus,

Received: 30 November 2020; **Revised:** 16 April 2021; **Accepted:** 16 April 2021

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

coordination also requires paying attention to one's own intentions and being able to predict and anticipate movement generation (Lau et al. 2004). However, the neural processes underlying such a flow of communication are not yet fully understood. Currently, the literature suggests the use of an internal forward model that optimizes one's motor control by comparing the actual and predicted sensory consequence of movements (Wolpert et al. 2003b). More broadly, such models could also be used to predict the actions of others in social interactions based on one's own action representations (Wolpert et al. 2003b). Such abilities would be supported by multiple interacting brain networks that are active in both generating actions and observing the actions of others (Rizzolatti et al. 1996). The frontoparietal network, including the inferior parietal lobe (IPL), has been associated with observation of individual movements (in both monkeys and humans), joint visual attention, and motor intention recognition (Rizzolatti and Sinigaglia 2010). Attention to one's intentions and actions also recruits the frontoparietal network, particularly the prefrontal cortex, which has strong functional coupling with the premotor cortex (Lau et al. 2004). Another brain region, the posterior parietal cortex, has also been repeatedly observed in situations involving motor intention and imagination (Jeannerod 1994), while regions in the lateral parietal cortex, such as the intraparietal sulcus (IPS) and IPL, have been more directly and precisely linked to attention and intention itself (Lau et al. 2004; Desmurget and Sirigu 2012; Eskenazi et al. 2015). While such a simulation system is described as the mirror neuron system in monkeys based on single-cell recordings, its existence in humans is still controversial (Keysers 2009; Turella et al. 2009; Campbell and Cunnington 2017). Intention recognition is particularly influenced by contextual information and prior knowledge (Brass et al. 2007). It is subject to information availability and processing (whole-body view vs. covert view), context richness (familiar vs. novel context) and expertise in a given task (Rizzolatti and Sinigaglia 2010). A few human studies have highlighted the influence of expertise on the prediction of movements in groups of experts, such as athletes (Aglioti et al. 2008), by showing their improved ability to anticipate the outcome of a given action in others. Compared to sport athletes, musicians have expertise in both specific sensorimotor skills and social signal analysis, and show structural brain changes, particularly in temporal and premotor cortex, as a strong biological basis for their expertise and training (Münste et al. 2002).

While music provides a unique solution to balance a rigorous experimental approach and ecological testing of cognition and high brain function (D'Ausilio et al. 2015), it also provides an excellent playground to explore a more refined form of coordination, joint action, and intention. Musical ensemble performance requires real-time interpersonal coordination at the level of sensorimotor, cognitive, emotional, and social processes (Keller et al. 2014). It involves nonverbal communication of information about musical structure and expressive intentions through the sounds and body movements of the performers. Musical coordination improves when co-performers share common representations of musical goals enabled by internal models in the central nervous system, as it helps them produce their parts in a way that is compatible with the other musicians and helps them anticipate the timing of each other's actions by generating online predictions during performance (Keller et al. 2014). In his 2016 review, Keller proposed that joint action and coordination in musical ensemble occur in a three-stage process: 1) integrating information related to one's own part, the parts of others, and the outcome of the joint action while maintaining a distinction between self et al., 2) representing self, others, and the joint action in predictive

internal models, and 3) recruiting the motor system to simulate self- and other-produced actions at multiple hierarchical levels. "Self" models aid in the production of one's own movement by allowing efficient action planning and execution while running slightly ahead of the movement to anticipate and correct potential errors before they occur (Wolpert and Kawato 1998; Jeannerod 2001). The "other" models simulate the observed actions of fellow players and allow a musician to predict what another will do, as well as how and when they will do it (Wolpert et al. 2003a; Keller 2008, 2012). The "other" models help to understand the intentions of the other (Wilson and Knoblich 2005; Schubotz 2007). Finally, the "joint" internal models integrate the outputs of "self" and "other" internal models and then modify their "own" inverse models to compensate for any discrepancies between these outputs (Keller et al. 2016). These internal models originate in the cerebellum, from where they communicate with other brain regions (Ito 2008). They drive simulations of goal-directed actions by recruiting brain areas normally involved in action execution and observation, but without causing overt movement and in the absence of appropriate sensory input (Pezzulo et al. 2013). Internal models are formed and adapted during repetition of a movement and enable us to move skillfully after repeated practice (Ito 2008). Therefore, "self," "other," and "joint" internal models must be trained for complex skills such as music making. In particular, the sensorimotor transformations represented in internal models must be acquired, reinforced, and refined through active experience and observational learning (Wolpert et al. 2003a; Schubotz 2007; Cross et al. 2009). While practicing instrumental technique contributes to the development of internal models that are recruited in the production of desired sounds, observing, and listening to fellow players leads to the calibration of internal models that represent other's action systems and allows an individual to learn to simulate another's playing style (Repp and Keller 2010). Musicians, through years of practice, develop a strong functional association between a musical note, its visual representation, and the movement required to produce it (Zatorre et al. 2007; D'Ausilio 2010).

Music has been used extensively in this regard to study brain processes involved in action representation, particularly focusing on visuomotor (Stewart et al. 2003; Buccino et al. 2004) and audiomotor (Bangert et al. 2006; Baumann et al. 2007; Lahav et al. 2007) processes. Similarly, dance has also been used extensively to study movement perception and production (for a review, see Sevdalis and Keller 2011). In both cases, point-light biological motion (Johansson 1973) has already been used to focus on brain processes driven by visual motion cues of actions (Saygin et al. 2004; Saygin 2007) and audiovisual integration (Brooks et al. 2007; Klin et al. 2009). However, previous studies (Aglioti et al. 2008; Abreu et al. 2012) have not specifically tested the utility of expertise on the recognition of expressive intentions under suboptimal conditions—when information is missing or altered (Rahnev and Denison 2018)—although this should be thoroughly investigated to better define and expand our knowledge on optimal vs. suboptimal perceptual decision making. Our focus on degraded sensory information helps to isolate the potential of kinematic information as a key source for decoding others' intentions in a very reduced time window, a situation typical of the ones tackled by musician experts during their performance. Indeed, such a line of research would contribute critically to understanding the behavioral and neural mechanisms underlying better coordination among multiple individuals. It would also provide a better understanding of the brain mechanisms that are necessary and sufficient to enable a correct assessment of intended communication. Under suboptimal conditions, the cerebellum

could play an important role in accurately predicting intention. This assumption is based on several functions of the cerebellum such as action comprehension, planning, timing, and, most importantly, because of its finely tuned connections with the basal ganglia and cerebral cortex (Caligiore et al. 2017). Indeed, the cerebellum is part of the networks for action observation (Sokolov et al. 2010) and voluntary movement (Hülsmann et al. 2003), and it is also an important hub for timing and rhythm processing (Molinari et al. 2007; Chen et al. 2008). Cerebellar activity is also enhanced during sensorimotor coordination in violinists (Krause et al. 2010) and is recruited extensively by abstract social cognition—such as fine-grained communicative intentions in short segments of more vs. less emphatic pieces—with involved subregions overlapping with sensorimotor cerebellar territories (Van Overwalle et al. 2014; Sokolov 2018). While the critical role of the cerebellum in optimal perceptual conditions of action and intention decoding is emphasized in the literature above, the likelihood of its involvement in suboptimal conditions also appears to be very high.

To shed light on the mechanisms of communicative intention, we recruited violinists and matched control participants who rated the visual dynamics of short pieces of violin solos presented with the violinist as a point-light display (PLD) after motion-capture recordings on an independent group of expert violinists. Communicative intent was materialized by categorizing piano vs. forte intentional gestures in these short pieces of music while continuous brain scans were performed using functional magnetic resonance imaging (fMRI). These short PLD videos were manipulated to include both original (unmodified, but with visual information only) and modified PLD segments: These modified segments included 1) a condition with spatially randomized initial positions of the dots (namely, the “spatial shuffling” condition) and 2) a condition in which the pieces were cropped so that only the first moments of the segments were preserved (namely, the “temporal cropping” condition). Spatial shuffling allowed us to remove information related to the musician’s body while preserving the kinetic energy associated with the movements. Kinetic energy associated with speed is an important component of expressive gestures and music perception (Eitan and Granot 2006; Schaerlaeken et al. 2017). Musicians in a musical ensemble may not have the attentional skills to focus on every detail of each other’s performance. One strategy might be to focus their attention on the movement of the others, as it has been shown that the brain regions responsible for movement decoding (MT/V5) can be modulated by attention (Treue and Maunsell 1996, 1999). The temporal excerpt mimics a different real-world situation in which the musician may not have access to the full unfolding of the musical gestures and therefore strategizes to respond to only a portion of them (usually the preparatory gesture). Based on these two situations, we then compared the performance of violinists and controls in terms of accurate recognition of expressive play for both normal and altered versions of the segments during fMRI. In other words, our experimental paradigm and population were chosen to test the general hypothesis that experts—that is, violinists—would perform better compared to control participants when asked to identify the communicative intent of both the original and altered musical pieces. We, therefore, hypothesized that there would be a clear and consistent performance advantage of expert violinists over control participants in judging the communicative intention of piano vs. forte for both original and modified (temporally cropped and spatially shuffled) pieces of music. With respect to neuroimaging data, we hypothesized increased activity in the frontoparietal network and cerebellum as a function

of expertise that might be modulated by interindividual differences. We expected increased activations in these brain regions, particularly for violinists when evaluating altered musical pieces (temporally cropped and spatially shuffled) compared to control participants. Finally, we predicted stronger coupling between the prefrontal and premotor cortex and between the cerebellum and the frontoparietal network for violinists compared to control participants, both when evaluating the communicative intent (piano vs. forte) of original and altered (temporally cropped and spatially shuffled) musical pieces.

Material and Methods

Participants

Thirty-seven right-handed participants took part in this study (25 women, 19 experts, 17 violinists, M age = 27 ± 7 years). Two participants repeatedly fell asleep during data collection and were therefore excluded from the final sample (N = 35). Participants were comprised of a group of expert violinists who had received at least 8 years of training at a high musical institution (Geneva School of Music) and a control group of non-violinists who had no musical training. While the number of males and females differed significantly between the two groups, ($\chi^2(1, N = 35) = 4.83, P = 0.027$), with more female than male violinists, age did not differ between the two groups ($t(30.26, N = 35) = -0.38, P = 0.7$). However, the variance induced by gender did not seem to affect our models (Supplementary Tables 1 and 2). All participants were naïve to the experimental procedure and material and had no known history of psychiatric or neurological disorders. Finally, all participants reported normal hearing and normal or corrected-to-normal vision (contact lenses or MRI-compatible plastic glasses). This study was conducted in accordance with the protocol, the current version of the Declaration of Helsinki, ICH-GCP, or ISO EN 14155 (as applicable), and all legal and regulatory requirements of the State Ethics Committee of the University Hospital of Geneva.

Experimental Stimuli: Communicative Intent Task

The complete set of stimuli consisted of 64 videos, repeated four times (256 stimuli in total, 128 *piano* and 128 *forte*) and split into four runs of equal duration (~5 min). Across runs, the stimuli were organized in eight conditions as follows: 1) spatially unmodified, temporally unmodified, and piano (N = 32); 2) spatially unmodified, temporally unmodified, and forte (N = 32); 3) spatially unmodified, temporally cropped, piano (N = 32); 4) spatially unmodified, temporally cropped, forte (N = 32); 5) spatially shuffled, temporally unmodified, piano (N = 32); 6) spatially shuffled, temporally unmodified, and forte (N = 32); 7) spatially shuffled, temporally cropped, and piano (N = 32); and 8) spatially shuffled, temporally cropped, and forte (N = 32). Temporally cropped conditions had a duration of 1.2 s, while temporally unmodified conditions had a duration of 2.7 s. Inter-trial interval had a mean duration of 1.5 s with a range of [1 s; 2 s]. Stimulus order within a run was pseudo-randomized so that the same condition would not occur three times in a row.

These stimuli were produced according to the following procedure. First, using a motion capture system (Qualisys, time sampling), we filmed the first violinist of a professional string quartet during 16 rehearsals of the same piece of music, *Death and the Maiden* by Schubert, which was chosen because it offers a wide variety of writing and expressive styles. For half of the rehearsals, the first violinist played alone; for the other half, he played with

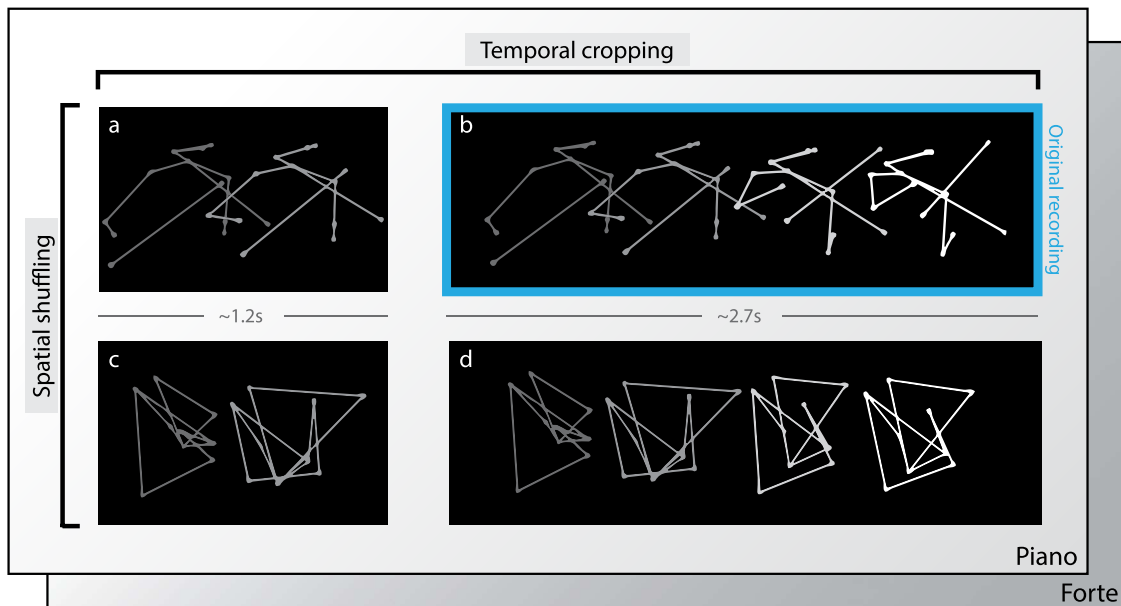


Figure 1. Overview of the stimuli displayed to the participant. The sequence of expressive motion is indicated by a dot-light display based on the collected motion capture of a first violinist of a string quartet. The two temporal segments used in the experiment refer, respectively, to the preparation (temporally cropped condition, **a**) and the preparation plus the entry part (temporally unmodified condition, **b**). For each of these temporal conditions, spatial shuffling was also applied as a separate condition (**c** and **d**, respectively). After viewing the selected sequence, participants were asked to indicate the perceived communicative intent of the music (forte or piano) by pressing a key.

the other members of the string quartet. During the recording sessions, the violinists were instructed to play as expressively as if they were giving a concert performance. The recording sessions took place over two days in a concert hall, as it offers naturalistic conditions perfectly suited to the musicians' needs and expectations (e.g., the quality of the acoustics). After filming, all clips were processed, and motion capture data were preprocessed to eliminate interfering data through a standard filtering process (despiking and smoothing using MATLAB (The Mathworks, Inc., Natick, MA, USA)) to produce cleaned PLD of the performances. The next step was to select specific moments in each performance where the first violinist indicated his intention to the fellow players with two communicative intentions: piano and forte. To do this, we worked with the violinists to define these key moments and we obtained the corresponding 16 short sequences (8 piano, 8 forte). Two experimental visual manipulations were then applied to each edited point-light sequence. The first manipulation consisted of segmenting the sequence into two parts: the first part referred to the preparation of the entry, just before the generation of the piano or forte gesture (duration = ~ 1.2 s), and the second part referred to the entire sequence (i.e., movement preparation plus entry (duration = ~ 2.7 s)). We refer to these sequences as "temporally cropped" vs. "temporally unmodified." For the temporally cropped sequences, we focused on the first moments of the sequence because we believe it is critical to understand the specific ability of the violinists to coordinate with each other. The second manipulation consisted of destroying the anthropomorphic shape of the stimuli through a single spatial scrambling process that preserved the dynamics of the individual points but shifted their relative relationships to other points. The final shape, which retained the same kinetic energy as the original, had no relation to an anthropomorphic shape. We refer to these sequences as "spatially shuffled" vs. "spatially unmodified." They were designed to highlight the violinists' higher processing of dynamic visual information in the absence of visual corporeal anthropomorphic references.

The final excerpts presented to participants were a combination of all three conditions in a pseudorandom order: communicative intent, temporal cropping (Fig. 1a and b), and spatial shuffling (Fig. 1c and d). For example, one excerpt might present a piano gesture, and be temporally cropped but spatially unmodified, while another might present a forte gesture, with temporal cropping and spatial shuffling. In all sequences, only visual information was used and the accompanying audio tracks were not presented (see an example for each condition in Fig. 1).

Experimental Procedure: Communicative Intent Task

Participants, divided into two groups (violinists vs. control participants), were subjected to a 2 (communicative intentions: piano vs. forte) \times 2 (temporal cropping: temporally cropped vs. temporally unmodified) \times 2 (spatial shuffling: spatially shuffled vs. spatially unmodified) within-subject factorial design, resulting in eight experimental conditions (Fig. 1). Prior to scanning, participants were introduced to and familiarized with the experimental task and confirmed they understood the study and what it required of them. This included performing two trials from each condition (these stimuli were excluded from the fMRI task), while the experimenter monitored performance. During the MRI session, participants watched point-light display movies of the violinist's movements. After each trial, participants had to report whether the performance related to the communicative intent forte or piano. After the fMRI session, all participants reported that they had no conceptual or technical problem performing the experimental task. The experimental procedure was based on a two-alternative forced-choice task. It aimed to reveal brain mechanisms underlying how groups of violinists (task experts) may differ in enacting fast and accurate decisions in perceptually altered conditions with respect to control participants (non-expert) (Gold and Shadlen 2002; Ratcliff and McKoon 2008; Deneve 2012).

Behavioral Data Analysis

The statistical software R was used to analyze all behavioral data. We computed a generalized linear mixed model (GLMM) to estimate the variance explained by the piano/forte and violinists/controls fixed factors on the percentage of correct responses. GLMM makes use of random effects modeling to improve the accuracy of the model and to allow the computation of models with non-normal distribution, here a binomial distribution. We tested our predictions for the effect of several fixed effect factors, including participant expertise, communicative intentions, spatial shuffling, and temporal cropping. Random intercept effects encapsulated variability with respect to each participant. We used a step-up strategy in building the model to test the different combinations of fixed effects. Based on the marginality principle, we present the highest order interaction effects (Nelder 1977), namely the interaction between expertise and the other experimental conditions mentioned above. We used chi-square difference tests to examine the contribution of each of our variables and their interactions. We report effect sizes in accordance with the approach of Nakagawa and Schielzeth, implemented in the R package “MuMIn” (Nakagawa and Schielzeth 2013). They created an approach based on two indicators, a marginal and a conditional R² (R²_m and R²_c, respectively). R²_m is the variance explained by the fixed factors, while R²_c is the variance explained by the entire model (both fixed and random effects). These two indicators allow comparability with standard methods, taking into account the variance explained by the random effects. We calculated and reported them for each statistical model.

Neuroimaging Image Acquisition

Neuroimaging data were acquired using a Siemens Trio 3.0 Tesla MRI scanner at the Brain and Behavioral Laboratory (BBL), University Medical Center, University of Geneva (Geneva, Switzerland). For each participant and for each run of the experimental task, 290 functional T2* -weighted echo planar image volumes (EPIs; slice thickness = 3 mm, gap = 1 mm, 36 slices, TR = 650 ms, TE = 30 ms, flip angle = 90°, matrix = 6464, FOV = 200 mm) were acquired. In addition, a T1-weighted, magnetization-prepared, rapid acquisition gradient echo scan (slice thickness = 1 mm, 176 slices, TR = 2530 ms, TE = 3.31 ms, flip angle = 7°, matrix = 256 256, FOV = 256 mm) was acquired. Thus, 290 volumes with 36 slices were acquired for each participant, resulting in a total of 10 440 slices. The total for all participants was 10,150 volumes and 365,400 slices.

Neuroimaging Data Analysis

Functional data were analyzed using Statistical Parametric Mapping version 12 (<https://www.fil.ion.ucl.ac.uk/spm>) in MATLAB (The Mathworks, Inc., Natick, MA, USA). Preprocessing steps included realignment to the first volume of the time series to correct for head motion, slice timing, normalization to the Montreal Neurological Institute (MNI) template (resampled at 3 × 3 × 3 mm), and finally spatial smoothing with an isotropic Gaussian kernel of 8-mm full width at half-maximum. A high-pass filter of 128 s was used to remove low-frequency components. A general linear model (first-level analysis) was then defined for each participant separately (within-subject statistics). For the experimental task, correctly scored trials were modeled by specific boxcar functions defined by the duration of the video stimuli from stimulus onset to offset and

convolved with the canonical hemodynamic response function. Group-level statistics were then performed using a flexible factorial design to account for variance across conditions and participants. Two different group-level models were calculated for the present data: Model 1 included eight conditions (1) spatially unmodified, temporally unmodified, and piano; 2) spatially unmodified, temporally unmodified, and forte; 3) spatially unmodified, temporally cropped, piano; 4) spatially unmodified, temporally cropped, forte; 5) spatially shuffled, temporally unmodified, piano; 6) spatially shuffled, temporally unmodified, and forte; 7) spatially shuffled, temporally cropped, and piano; 8) spatially shuffled, temporally cropped, and forte) and two groups (violinist; control) without covariates [group × conditions], whereas model 2 included task performance as a group-level covariate of interest that interacted with the factor conditions [performance × conditions]. Both models also included a mandatory “participant factor” that allowed for the calculation of between-subject variability. For both group-level models, the specification of independence was set to true for the “participant” and “group” factors, whereas it was set to false for the other factor conditions. Regarding variance estimation, it was set to unequal for all factors including group, because homoscedasticity criteria cannot usually be met for fMRI data (default setting in SPM12). For both models, group-level results in SPM12 were estimated voxel-wise using corrected statistics with $P < 0.05$ false discovery rate (FDR) and an arbitrary cluster threshold of $k > 10$ voxels. For all analyses, regions of significant activation increase were labeled based on probabilistic cytoarchitectonic atlases (Automated Anatomical Labelling Atlas (Tzourio-Mazoyer et al. 2002), Cerebellum Atlas (Diedrichsen et al. 2009; Diedrichsen et al. 2011)), and rendered on semi-inflated brains from the CONN toolbox (<http://www.nitrc.org/projects/conn>), see Fig. 2.

Model 1: Conditions Training Group-Level Statistics

For this first model, regressors were created for each experimental condition and for each participant ($N = 35$), resulting in a first-level planning matrix that contained a total of 16 regressors, including 8 regressors of interest (conditions, see above) and 8 regressors of no interest (incorrect trials, the six motion parameters, and the constant term). Each regressor of interest was used to compute main effect contrast vectors, which were then carried forward into a second-level, group analysis using the flexible factorial design specification that we will describe in detail here. The group-level analysis included the following factors: conditions (see above) and group. The condition factor was used to compare violinists with control participants on their ability to judge the communicative intent of the stimuli, regardless of whether they were spatially shuffled and/or temporally cropped. The following contrasts were therefore computed using the factorial architecture of the data mentioned above: [piano > forte * violinists > controls], [temporally cropped > temporally unmodified * violinists > control], and [spatially shuffled > spatially unmodified * violinists > control] (see Fig. 2).

Model 2: Conditions Training Group-Level Statistics with Task Performance as Second-Level Covariate

The second model used exactly the same settings and factorial structure as Model 1, in addition to a group-level covariate that accounted for task performance for each participant (percentage of hits during the experimental task). This covariate was set to interact with the conditions factor. Therefore, Model 2 contained the following factorial structure: condition * task performance.

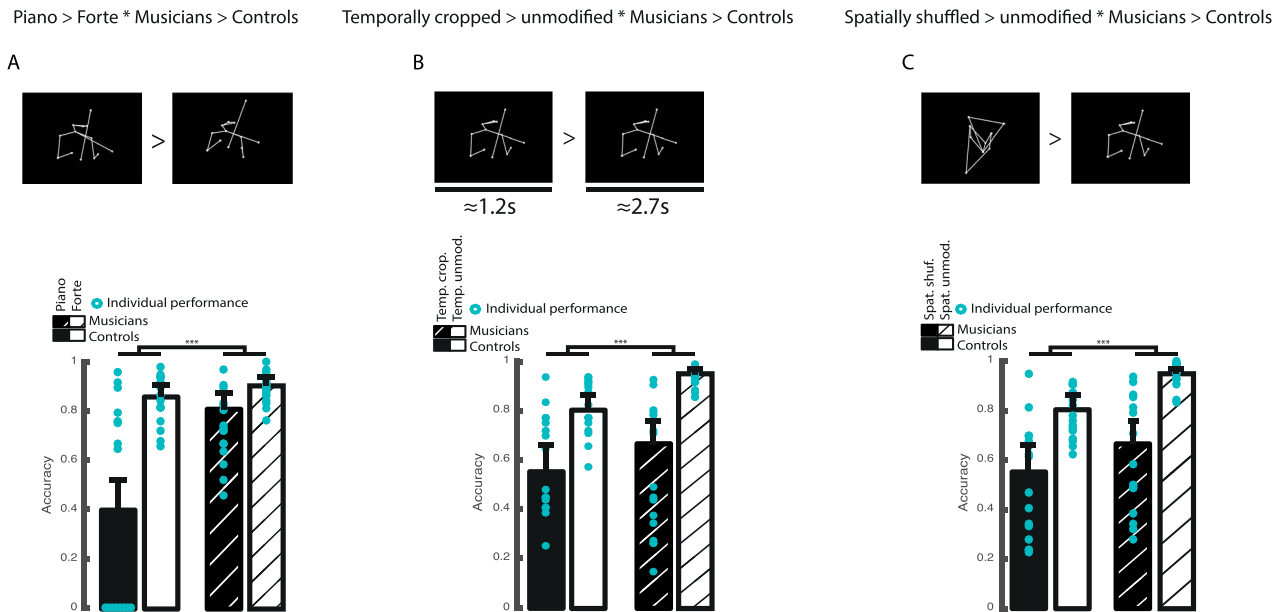


Figure 2. Experimental stimuli and behavioral results for the impact of expertise on intention evaluation. (a) Example of piano vs. forte PLD and (d) averaged performance and individual points per group for piano vs. forte piece dynamics. (b) Example of temporally cropped vs. temporally unmodified PLD and (e) averaged performance and individual points per group for piano vs. forte piece dynamics in temporally cropped vs. temporally unmodified sequences. (c) Example of spatially shuffled vs. spatially unmodified PLD and (f) averaged performance and individual points per group for piano vs. forte piece dynamics in spatially shuffled vs. spatially unmodified sequences. Error bars represent 95% confidence interval. [***: $P < 0.001$, **: $P < 0.01$, *: $P < 0.05$;.: $P < 0.1$; Spat.: spatial, Temp.: temporal, shuf.: shuffling, crop.: cropping, unmodif.: unmodified].

This model was used to constrain our statistical results and to observe how some brain regions are sensitive to task performance in violinists as opposed to control participants for the following contrasts: [piano > forte * performance], [temporally cropped > temporal unmodified * performance], and [spatially shuffled > spatially unmodified * performance]. The results of this second model have been overlaid in green-to-blue in Figure 2.

Functional Connectivity Analyses

Functional connectivity analyses were performed using CONN Toolbox v18.a (Whitfield-Gabrieli and Nieto-Castanon 2012). Interfering noise sources were estimated and removed using the automated toolbox preprocessing algorithm, and the remaining BOLD time series was bandpass filtered using a low frequency window ($0.008 < f < 0.09$ Hz). Correlation maps were then created for each condition of interest by taking the remaining BOLD time course for each condition from the atlas ROIs and calculating bivariate Pearson's correlation coefficients between the time courses of each voxel of each region of the atlas. These correlations were then converted to normally distributed values using the Fisher transform. Finally, group-level analyses were performed using these Fisher-transformed correlation maps to test for main effects within groups and significant connectivity differences between groups for the contrasts of interest. Type I errors were controlled for by using the seed-level FDR correction with $P < 0.05$ to correct for multiple comparisons.

Results

Behavior

Behavioral results showed a generalized and reliable advantage of violinists over control participants in correctly discriminating

between the communicative intent piano or forte (Fig. 2). Interaction effects of our factors (group [violinists>controls] and conditions [forte>piano, temporally cropped>temporally unmodified, spatially shuffled>spatially unmodified]) explained a greater proportion of the variance for each statistical model, compared to models with only the main effects (all $P < 0.001$, full statistics in Supplementary Table 2). More specifically, for each computed model, we observed that the performance of all our participants decreased significantly when the information was altered (temporally cropped, spatially shuffled; Fig. 2b and c) or more subtle (piano, Fig. 2a) (all $P < 0.001$, full statistics in Supplementary Tables 2 and 3). At the group level, violinists outperformed control participants in estimating communicative intentions regardless of the condition presented (Fig. 2a-f, all $P < 0.001$). Finally, we describe a significant interaction effect between group and conditions for each model (all $P < 0.001$). No differences were observed between genders (Supplementary Fig. 1, Supplementary Table 1).

Neuroimaging

Whole Brain Data

Neuroimaging results focused on regions showing enhanced activations for violinists compared to control participants, when also comparing communicative intention, temporal cropping, and spatial shuffling to unmodified excerpts. Analyses focused exclusively on trials in which participants correctly identified the intention presented in the PLD. When focusing on the excerpts that expressed a piano nuance (piano > forte * violinists > control), we observed enhanced activations in the pre-supplementary area (preSMA) (MNI coordinates in Supplementary Tables 4 and 5) and left dorsolateral prefrontal cortex (DLPFC) (Fig. 3a and j, Supplementary Tables 4 and 5, Supplementary Figures 2 and 3). When we focused on the

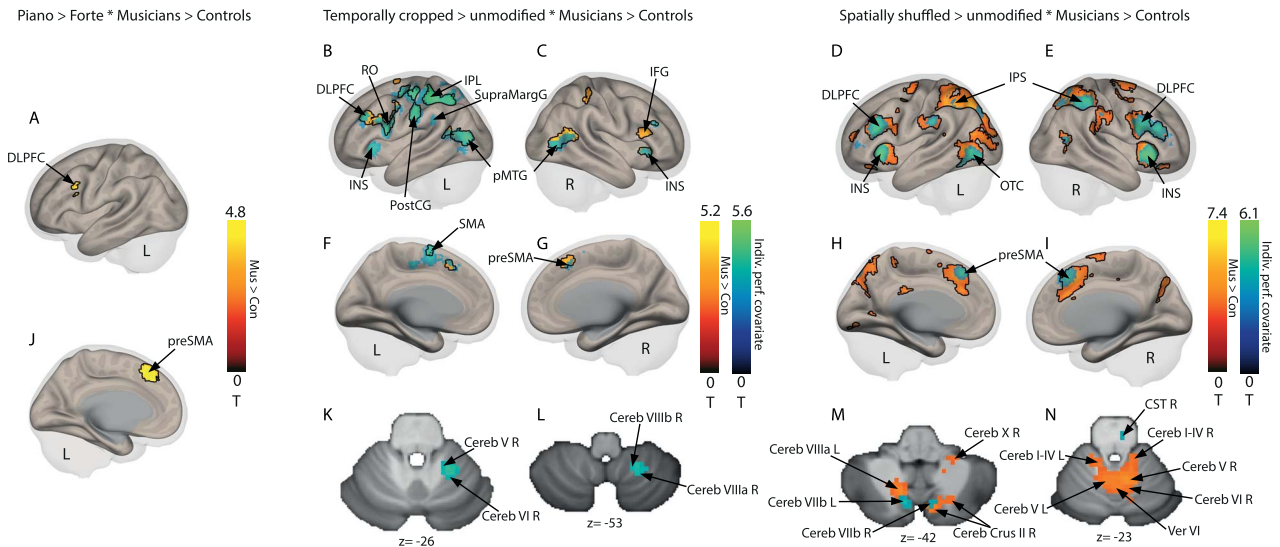


Figure 3. Neural evidence for decoding communicative intentions in violinists and control participants. (a and j) Increased activity in preSMA and DLPFC for piano vs. forte sequences in violinists vs. control participants. (b,c,f,g) Increased activity for temporally cropped vs. temporally unmodified sequences in violinists vs. control participants in red-to-yellow in IPL, and with overall task performance as group covariates in blue-to-green in IPL, DLPFC, pMTG, preSMA, and cerebellar subregions (k and l). (d,e,h,i) Increased activity for spatially shuffled vs. spatially unmodified sequences for violinists vs. control participants in red-to-yellow in IPS, preSMA, DLPFC, and INS, and with overall task performance as group-level covariates in blue-to-green in IPS, preSMA, DLPFC, INS, and OTC and in cerebellum (m and n). Color bars represent statistical T values of contrast. Black outlines delineate the regions of the group analysis of model 1—not the performance based analysis of model 2. [Cereb: cerebellum lobule; Cereb Crus: cerebellum crus of ansiform lobule; CST: corticospinal tract of the brainstem; DLPFC: dorsolateral prefrontal cortex; IFG: inferior frontal gyrus; INS: insula; IPL: inferior parietal lobule; IPS: inferior parietal sulcus; lingual gyrus; OTC: occipito-temporal cortex; pMTG: medial temporal gyrus; posterior part; preSMA: pre supplementary motor area; PostCG: postcentral gyrus; RO: Rolandic operculum; SupraMargG: supra marginal gyrus; Ver: vermis; Violon.: violinists; Con.: control participants; L: left; R:right]. Voxel-wise $P < 0.05$ FDR corrected.

temporally cropped sections (temporally cropped > temporally unmodified * violinists > control), we observed increased activations in the left IPL, cerebellar lobes V, VI, VIIb, and VIII, bilateral posterior middle temporal gyrus (MTG), and inferior frontal gyrus (IFG) (Fig. 3b,c,f,g,k,l; Supplementary Tables 4 and 5, Supplementary Figures 2 and 3). Finally, when we focused on the spatially shuffled excerpts (spatially shuffled > spatially unmodified * violinists > control), we observed enhanced activations in the bilateral intraparietal sulcus (IPS), the right preSMA, and the bilateral DLPFC, which extended to the IFG pars opercularis and triangularis, the cerebellum (vermis areas IV,V, crus II, lobules I-IV, V, VI, VIIb, VIIla, VIIlb, X), and the bilateral insula (Fig. 3d,e,h,i,m,n; Supplementary Tables 4 and 5, Supplementary Figures 2 and 3). To further highlight these increased activations in the context of each participant's global performance, we also calculated another second-level analysis with the participants' individual performance as group-level covariate. Consequently, we were able to account for brain regions with enhanced activations in relation to overall performance (group-level analysis) for our contrasts of interest. These results include interindividual variability with respect to task performance across groups (continuous, one average value per participant). Displayed in green-to-blue activations (Fig. 3b-e,f-i,k-n; Supplementary Tables 4 and 5, Supplementary Figure 2), the global individual performance analysis showed a strong overlap with the above brain regions for the temporally cropped condition, particularly in the IPL, SMA, and cerebellum (temporally cropped > temporally unmodified * violinists > control with individual performance, Fig. 3b,c,f,g,k,l; Supplementary Tables 4 and 5, Supplementary Figure 2). This result suggests that this network plays an important role in the accurate assessment of communicative intention of temporally cropped PLD as a function of performance and expertise,

and is also sensitive to individual performance differences. This overlap between analyses was smaller for the spatially shuffled excerpts (spatially shuffled > spatially unmodified * violinists > control with individual performance), especially in the cerebellum. Indeed, the results show that parts of the bilateral insula, bilateral preSMA, bilateral DLPFC, and right IPS overlapped, highlighting the important role of these regions in both performance and expertise. Interestingly, the left IPS showed much less overlap between analyses than the right IPS, suggesting interhemispheric dissociation (Fig. 3d,e,h,i,m,n), with left IPS activity massively enhanced in correctly assessed piece dynamics, whereas the right IPS was modulated by interindividual differences in performance. Such an overlap between analyses was also observed bilaterally for cerebellar area VIIb (Fig. 3m). Additional brainstem activity was observed in the right corticospinal tract (Fig. 3n).

Functional Connectivity Data

Atlas-based analyses of seed-to-seed functional connectivity (FC) were performed to highlight the existence of widespread coupled brain activity targeting frontoparietal and cerebellar regions related to both communicative intention processing and expertise. These analyses revealed the involvement of numerous regions observed in our whole-brain contrasts of interest, in addition to subcortical and cerebellar connectivity (Fig. 4). A general effect of expertise across conditions (violinists > control, main effect of all conditions) showed functional coupling between the left IPL and the left postcentral gyrus (Supplementary Fig. 4). As for our contrasts of interest, communicative intention and expertise (piano > forte * violinist > control) interacted and resulted in both coupled and negatively coupled functional networks (Fig. 4a; Supplementary Table 6).

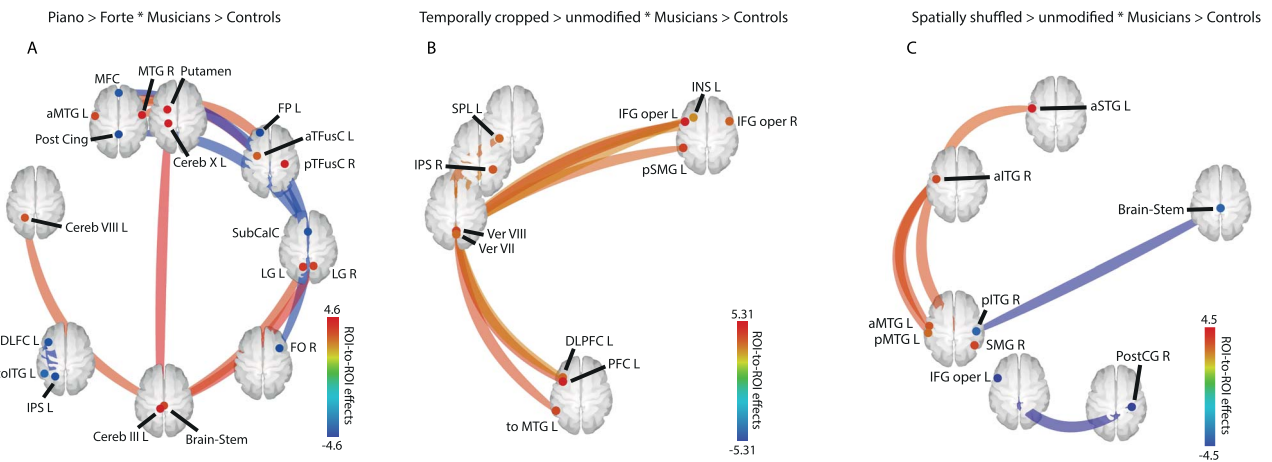


Figure 4. Functional connectivity of decoding communicative intentions in violinists as compared to control participants. (a) Increased connectivity for piano vs. forte excerpts in violinists vs. control participants. (b) Increased connectivity for temporally cropped vs. unmodified sequences in violinists vs. control participants. (c) Increased connectivity for spatially shuffled vs. unmodified sequences for violinists vs. control participants. [aITG: inferior temporal gyrus, anterior part; aSTG: superior temporal gyrus, anterior part; aTFus: temporal fusiform, anterior part; Cereb: cerebellum lobule; DLPFC: dorsolateral prefrontal cortex; FO: frontal operculum; FP: frontal pole; IFG oper: inferior frontal gyrus operculum; INS: insula; IPL: inferior parietal lobule; IPS: inferior parietal sulcus; LG: lingual gyrus; MFC: medial frontal cortex; MTG: medial temporal gyrus; PFC: prefrontal cortex; pITG: inferior temporal gyrus, posterior part; pMTG: medial temporal gyrus, posterior part; Post Cing: posterior cingulate; PostCG: posterior central gyrus; pSMG: superior medial gyrus, posterior part; pTFusC: temporal fusiform cortex, posterior part; R: right; SMG: superior medial gyrus; SPL: superior parietal lobule; SubCalC: subcallosal cortex; toITG: inferior temporal gyrus temporo-occipital part; toMTG: medial temporal gyrus, temporo-occipital part; Ver: vermis]. Seed-level $P < 0.05$ FDR corrected.

Specifically, we observed coupled FC between the bilateral MTG, left putamen, bilateral fusiform cortex, brainstem, and several subregions within the cerebellum, such as cerebellar lobules III, VIII, and X of the left hemisphere. Negatively coupled FC was observed between the medial frontal cortex, posterior cingulate gyrus, frontal pole, left DLPFC, and left IPS (see details in Fig. 4a; Supplementary Table 6). For temporally cropped excerpts (temporally cropped > temporally unmodified * violinists > control), only coupled FC was observed. More specifically, the analyses revealed widespread fronto-parieto-cerebellar FC in the bilateral inferior frontal gyrus pars opercularis (IFGop), left DLPFC, left superior parietal lobule (SPL), right IPS, and vermis areas VII and VIII (see details in Fig. 4b; Supplementary Table 6). The final contrast of interest with visually shuffled PLD (spatially shuffled > spatially unmodified * violinists > control) highlighted coupled FC between the anterior part of the left inferior, middle, and superior temporal gyri (aITG, aMTG, and aSTG), the left posterior MTG, and the right supramarginal gyrus, whereas negatively coupled FC characterized connectivity between the left IFGop, right posterior ITG, posterior cingulate gyrus, and brainstem (see details in Fig. 4c; Supplementary Table 6).

Discussion

The present study aimed to gain a clearer understanding of the interaction between expertise and assessment of communicative intention as a potential proxy for social interactions and coordination. Using point-light representations of violinists as stimuli, we asked experts and non-experts, namely violinists and control participants with no musical training, to rate the expressive intention of the performances. Expressive intent could materialize as piano or forte and could be visually modified, temporally modified, both, or neither. Our results show that violinists consistently performed better than control participants, whether for unmodified or modified stimuli. Premotor and lateral parietal areas together with the dorsolateral prefrontal cortex, as well as numerous cerebellar regions, appear to be the critical players in

the violinists' advantage in successfully assessing communicative intent.

Behavioral results confirmed the role of expertise in perceiving the intentions of others. Violinists were more consistent and accurate in perceiving expressive gestures, indicating a close link between perception and action skills (Wöllner and Cañal-Bruland 2010; Küssner et al. 2014). Importantly, these results held true even when information was absent or altered, highlighting the advantages of action expertise in understanding, integrating, and predicting actions based on the short dynamics of segments and even when anthropomorphic information was altered. Even in the absence of visual contact, musicians have been shown to be capable of interpersonal coordination because they can rely on auditory imagery to promote the operation of internal models that simulate their own and others' actions during ensemble performance (Keller and Appel 2010). This underscores the importance of combining perceptual information about one's own and others' goal-directed movements for action processing and prediction (Spunt and Lieberman 2012).

Our neuroimaging results revealed a widely ramified network of brain regions depending on expertise in decoding or inferring communicative intent, especially under highly altered perceptual conditions. Violinists' advantage in understanding and estimating communicative intent recruited the DLPFC, which is associated with action observation (Rizzolatti and Sinigaglia 2010) and influenced by training and expertise (Moore et al. 2006), and the preSMA, which is mainly involved in internally and externally selected actions (Mueller et al. 2007). Therefore, preSMA and DLPFC appear to be sufficient for accurately internalizing actions and extracting communicative intentions in experienced participants, respectively. These regions would also explain violinists' ability to integrate the temporal structure of rhythm via working memory (Chen et al. 2008). Functional connectivity analyses also revealed that expert participants relied on both positively and negatively coupled networks to successfully infer communicative intention. Positively coupled networks include areas involved in intention probability (putamen; Zapparoli et al.

2018) as well as motion prediction and motor imagination (cerebellum; Sokolov et al. 2017) and time encoding (brainstem and cerebellum; Rao et al. 2001; Molinari et al. 2007). More specifically, a large region of the cerebellum, including lobule VIII (and VIIIa), has been shown to covary with instrumental expertise, particularly for temporal complexity (Chen et al. 2008). This finding raises the question of whether general or instrument-specific skills of violinists directly influence cerebellar activity; this important distinction should be investigated in future studies. In a lesion study, lobule III—as well as lobules I, IV, and V—was also involved in action observation (Sokolov et al. 2010), whereas lobule IX and, to a lesser extent, lobule X were repeatedly associated with verbal working memory (Van Overwalle et al. 2014). Thus, as a functionally connected cerebellar network and through connections with the brainstem and basal ganglia, these lobules confer strong weight on the cerebellum as a crucial player in action observation, processing, and prediction. On the other hand, negatively coupled activity recruited brain regions known for mental states attributed to moving shapes and for memory for intentions (medial prefrontal cortex or frontal pole; Blakemore and Decety 2001), for action observation (ITG and DLPFC; Blakemore and Decety 2001; Rizzolatti and Sinigaglia 2010), and for attention to and understanding of intentions (IPS; Blakemore and Decety 2001).

The behavioral advantage in the evaluation of communicative intention under altered perceptual conditions by experienced participants relied essentially on several of the brain areas mentioned above with some additions. In temporally cropped sequences, expertise additionally recruited positively connected regions involved in interoception and motor intention perception (insula; Craig and Craig 2009), intentional action production (IFGop; Zapparoli et al. 2018), intention comprehension (IPS and IPL; Blakemore and Decety 2001), and temporal processing related to actions (vermis, especially areas VIII and VIIIa; Rao et al. 2001). Communicative intention comprehension as a function of expertise in spatially shuffled PLD recruited very similar brain areas at the whole-brain level compared with temporal cropping, but with larger clusters. This result could be explained by an advantage in assessing complex visual inputs in violinists compared to non-violinists for assessing instrumental performance (Griffiths and Reay 2018). Specifically, in the spatially altered condition, violinists made greater use of bilateral regions such as IPL and IPS. While IPL helps establish a stable body-centered reference system for movement planning that uses visual and kinesthetic information, IPS is responsible for mental rotation, especially in creative processes such as music composition (Wöllner 2017). Both regions helped violinists make sense of the altered visual information, reconstruct a stable representation of the expressive gestures and associated intention. In addition, general task performance as a group-level covariate constrained our whole-brain results and showed a difference between left and right IPS, with the latter showing a larger cluster of increased activity as a function of task performance in experts vs. controls. This suggests a specific role for the right IPS region in the context of interindividual differences in performance, and indeed this region has been reported to contribute to interpersonal synchronization in the context of actions (Bhat et al. 2017). In addition, the inferior parietal cortex plays a role in discriminating between self-generated actions and actions generated by others. The right inferior parietal cortex is activated when participants mentally simulate actions from the perspective of another person but not from their own perspective (Ruby and Decety 2001). Because this distinction between self and other is a key process for

understanding intentions in musical ensemble (Keller et al. 2016), we propose that more accomplished musicians use their ability to distinguish between self and other to better predict intentions. It is also worth mentioning the role of the superior temporal cortex, particularly the posterior STS, which receives important information from the dorsal visual pathway but also from the ventral visual pathway and therefore has a dual role in both perception for identification and action perception (Blakemore and Decety 2001). It was therefore surprising that our data showed no activity in the STS for biological motion perception—comparing piano vs. forte (Blakemore and Decety 2001; Beauchamp 2015), especially since the contrast included correctly assessed PLD. Indeed, posterior STS activity was predicted by performance in a biological movement task (Herrington et al. 2011), and the same region was also shown to communicate with the left cerebellum (Sokolov et al. 2012), several subregions of which showed increased activity in our data, particularly for spatially shuffled PLD that were correctly assessed by violinists. However, STS activity was increased in altered perceptual conditions, and it was interesting to observe greater activity of the right posterior STS for temporally cropped—and not spatially shuffled—PLD in violinists compared with control participants. This result may help to specify the active window of the posterior STS' function in biological motion perception, namely that this region is activated early in experts and leads to accurate evaluation even when stimulation is present for only a short period of time. However, such assumptions should be tested in more detail using imaging techniques with higher temporal resolution. In addition, functional connectivity analyses revealed coupled temporal cortices (anterior STG, MTG, ITG; posterior MTG, ITG) and negatively coupled connectivity in IFGop, brainstem, and posterior cingulate cortex. The coupled networks showed processing of biological motion independent of motor information (superior and middle temporal cortex; Rizzolatti and Sinigaglia 2010) and attribution of intentions to spatially displaced stimuli (posterior STG; Lee et al. 2012), regions interestingly known for their feed-forward connections to the IPS and IPL (Rizzolatti and Sinigaglia 2010). Negatively coupled functional connectivity in turn emphasized intentional action production or unusual action intention processing (IFGop; Zapparoli et al. 2018) and movement initiation and control (brainstem and cerebellum; Rao et al. 2001; Nandi et al. 2002).

While our data shed new light on intention decoding and highlighted additional behavioral contexts that favor experts over non-experts, several limitations should be considered. First, the sample size could have been larger, although it was difficult to recruit several additional violinists who would meet our inclusion criteria. Second, previous studies have highlighted structural brain differences between violinists and non-violinists (Gaser and Schlaug 2003), and therefore, we could have further explored our results by, for example, capturing diffusion tensor imaging to better characterize anatomical variations among our participants. Third, our stimuli contained only point-light representations of violinists, which limits our conclusions regarding other types of expertise (e.g., professional athletes or dancers). However, we chose to study communicative intention in music for its ecological validity and rigorous experimental approach (D'Ausilio et al. 2015). Fourth, although both spatial and temporal changes reflect potential real-world situations, other types of stimulus modification, such as modifying rhythmicity or adding sublevels of visual shuffling, could have been used to further specify the influence of expertise on more subtle perceptual changes and their effect on decoding communicative intent. Fifth, while we used a group-level performance covariate to

characterize interindividual differences in ratings of communicative intent between groups, the task-related fMRI data were modeled using only trials with correct intention evaluation. The reason for this decision was that we were primarily interested in correct ratings of communicative intention by experts compared with non-expert participants—but this decision has the major drawback of setting aside variance explained by incorrect ratings, which hinders our interpretation of fMRI results compared with behavioral results. The use of task-specific computational modeling of communicative intent using model-based fMRI analysis could circumvent such problems by the inclusion of every trial and should be a method of choice in the future (Lebreton et al. 2019). Finally, the use of functional—rather than effective—connectivity can be criticized, as Pearson's correlations between regions do not allow for a test of direct, causal relationship(s) between the regions of interest compared to partial correlations or multivariate regressions (Reid et al. 2019).

Considering our behavioral and neuroimaging data, as well as study limitations, our results suggest a strong role of expertise in understanding and predicting actions and communicative intentions. This claim is especially true under altered perceptual conditions, namely, visually or temporally altered stimuli, and this advantage of violinists over non-violinists would rely on regions of the frontoparietal network, in addition to various areas of the cerebellum, basal ganglia, and brainstem. Such neural systems could also play a role in numerous other conditions in everyday social interactions, as humans are experts at predicting others as they interact with them on a daily basis.

Supplementary Material

Supplementary material can be found at *Cerebral Cortex Communications* online.

Data availability statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on request.

Code availability statement

The codes used to analyze the data of the current study are available from the corresponding author on request.

Funding

A project grant from the Swiss National Science Foundation (51NF40-104897—DG). The EU ICT SIEMPRE project gratefully acknowledges financial support from the Future and Emerging Technologies (FET) program within the European Commission's Seventh Framework Programme for Research (under FET-Open grant number 250026-2).

Notes

We thank Bruno Bonet of the Brain and Behavior Laboratory of the Medical Center of the University Hospital of Geneva for his assistance in data collection. Finally, we would like to thank the violinists of the Quartetto di Cremona for recording the stimuli and the contribution of violinist experts Chiara Noera and Florence Malgoire for their valuable suggestions. *Conflict of Interest*: None declared.

References

- Abreu AM, Macaluso E, Azevedo R, Cesari P, Urgesi C, Aglioti SM. 2012. Action anticipation beyond the action observation network: a functional magnetic resonance imaging study in expert basketball players. *Eur J Neurosci*. 35(10):1646–1654.
- Aglioti SM, Cesari P, Romani M, Urgesi C. 2008. Action anticipation and motor resonance in elite basketball players. *Nat Neurosci*. 11(9):1109–1116.
- Andersen RA, Cui H. 2009. Intention, action planning, and decision making in parietal-frontal circuits. *Neuron*. 63(5):568–583.
- Bangert M, Peschel T, Schlaug G, Rotte M, Drescher D, Hinrichs H, Heinze H-J, Altenmüller E. 2006. Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *NeuroImage*. 30(3):917–926.
- Baumann S, Koeneke S, Schmidt CF, Meyer M, Lutz K, Jancke L. 2007. A network for audio-motor coordination in skilled pianists and non-musicians. *Brain Res*. 1161:65–78.
- Beauchamp MS. 2015. The social mysteries of the superior temporal sulcus. *Trends Cogn Sci*. 19(9):489–490.
- Bhat AN, Hoffman MD, Trost SL, Culotta ML, Eilbott J, Tsuzuki D, Pelphey KA. 2017. Cortical activation during action observation, action execution, and interpersonal synchrony in adults: a functional near-infrared spectroscopy (fNIRS) study. *Front Hum Neurosci*. 11:431.
- Blakemore S-J, Decety J. 2001. From the perception of action to the understanding of intention. *Nat Rev Neurosci*. 2(8):561–567.
- Brass M, Schmitt RM, Spengler S, Gergely G. 2007. Investigating action understanding: inferential processes versus action simulation. *Curr Biol*. 17(24):2117–2121.
- Brooks A, Van Der Zwan R, Billard A, Petreska B, Clarke S, Blanke O. 2007. Auditory motion affects visual biological motion processing. *Neuropsychologia*. 45(3):523–530.
- Buccino G, Vogt S, Ritzl A, Fink GR, Zilles K, Freund H-J, Rizzolatti G. 2004. Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. *Neuron*. 42(2):323–334.
- Caligiore D, Pezzulo G, Baldassarre G, Bostan AC, Strick PL, Doya K, Helmich RC, Dirx M, Houk J, Jörntell H. 2017. Consensus paper: towards a systems-level view of cerebellar function: the interplay between cerebellum, basal ganglia, and cortex. *Cerebellum*. 16(1):203–229.
- Campbell ME, Cunnington R. 2017. More than an imitation game: Top-down modulation of the human mirror system. *Neurosci Biobehav Rev*. 75:195–202.
- Chen JL, Penhune VB, Zatorre RJ. 2008. Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J Cogn Neurosci*. 20(2):226–239.
- Craig AD, Craig A. 2009. How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci*. 10(1):59–70.
- Cross ES, Hamilton AFC, Kraemer DJ, Kelley WM, Grafton ST. 2009. Dissociable substrates for body motion and physical experience in the human action observation network. *Eur J Neurosci*. 30(7):1383–1392.
- D'Ausilio A. 2010. *Reviews: The Oxford handbook of auditory science: The ear, the auditory brain, hearing*. London, England: SAGE Publications Sage UK.
- D'Ausilio A, Novembre G, Fadiga L, Keller PE. 2015. What can music tell us about social interaction? *Trends Cogn Sci*. 19(3):111–114.
- Deneve S. 2012. Making decisions with unknown sensory reliability. *Front Neurosci*. 6:75.
- Desmurget M, Sirigu A. 2012. Conscious motor intention emerges in the inferior parietal lobule. *Curr Opin Neurobiol*. 22(6):1004–1011.

- Diedrichsen J, Balsters JH, Flavell J, Cussans E, Ramnani N. 2009. A probabilistic MR atlas of the human cerebellum. *NeuroImage*. 46(1):39–46.
- Diedrichsen J, Maderwald S, Küper M, Thürling M, Rabe K, Gizewski E, Ladd ME, Timmann D. 2011. Imaging the deep cerebellar nuclei: a probabilistic atlas and normalization procedure. *NeuroImage*. 54(3):1786–1794.
- Eitan Z, Granot RY. 2006. How music moves:: Musical parameters and listeners images of motion. *Music Percept*. 23(3):221–248.
- Eskenazi T, Rueschemeyer S-A, de Lange FP, Knoblich G, Sebanz N. 2015. Neural correlates of observing joint actions with shared intentions. *Cortex*. 70:90–100.
- Gaser C, Schlaug G. 2003. Brain structures differ between musicians and non-musicians. *J Neurosci*. 23(27):9240–9245.
- Gold JI, Shadlen MN. 2002. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*. 36(2):299–308.
- Griffiths N, Reay J. 2018. The Relative Importance of Aural and Visual Information in the Evaluation by Musicians and Non-Musicians of Classical Music Performance. *Music Percept*. 35(3):364–375.
- Herrington JD, Nymberg C, Schultz RT. 2011. Biological motion task performance predicts superior temporal sulcus activity. *Brain and Cognition*. 77(3):372–381.
- Hülsmann E, Erb M, Grodd W. 2003. From will to action: sequential cerebellar contributions to voluntary movement. *NeuroImage*. 20(3):1485–1492.
- Ito M. 2008. Control of mental activities by internal models in the cerebellum. *Nat Rev Neurosci*. 9(4):304–313.
- Jahng J, Kralik JD, Hwang D-U, Jeong J. 2017. Neural dynamics of two players when using nonverbal cues to gauge intentions to cooperate during the Prisoner's Dilemma Game. *NeuroImage*. 157:263–274.
- Jeannerod M. 1994. The representing brain: Neural correlates of motor intention and imagery. *Behav Brain Sci*. 17(2):187–202.
- Jeannerod M. 2001. Neural simulation of action: a unifying mechanism for motor cognition. *NeuroImage*. 14(1):S103–S109.
- Johansson G. 1973. Visual perception of biological motion and a model for its analysis. *Percept Psychophys*. 14(2):201–211.
- Keller PE. 2008. Joint action in music performance. In: Morganti F, Carassa A, Riva G, editors. *Emerging communication: Studies on new technologies and practices in communication*. Vol. 10. Enacting intersubjectivity: A cognitive and social perspective on the study of interactions. Amsterdam, Netherlands: IOS Press. pp. 205–221.
- Keller PE. 2012. Mental imagery in music performance: underlying mechanisms and potential benefits. *Ann N.Y. Acad. Sci.* 1252(1):206–213.
- Keller PE, Appel M. 2010. Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Percept*. 28(1):27–46.
- Keller PE, Novembre G, Hove MJ. 2014. Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philos Trans R Soc Lond B Biol Sci*. 369(1658):20130394.
- Keller PE, Novembre G, Loehr J. 2016. *14 musical ensemble performance: representing self, other and joint action outcomes*. Vol. 280. Shared representations: Sensorimotor foundations of social life. New York, NY: Cambridge University Press.
- Keysers C. 2009. Mirror neurons. *Current Biology*. 19(21):R971–R973.
- Klin A, Lin DJ, Gorrindo P, Ramsay G, Jones W. 2009. Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*. 459(7244):257–261.
- Kotz SA, Schwartze M. 2010. Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn Sci*. 14(9):392–399.
- Krause V, Schnitzler A, Pollok B. 2010. Functional network interactions during sensorimotor synchronization in musicians and non-musicians. *NeuroImage*. 52(1):245–251.
- Küssner MB, Tidhar D, Prior HM, Leech-Wilkinson D. 2014. Musicians are more consistent: Gestural cross-modal mappings of pitch, loudness and tempo in real-time. *Front Psychol*. 5:789.
- Lahav A, Saltzman E, Schlaug G. 2007. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J Neurosci*. 27(2):308–314.
- Lau HC, Rogers RD, Haggard P, Passingham RE. 2004. Attention to intention. *Science*. 303(5661):1208–1210.
- Lebreton M, Bavard S, Daunizeau J, Palminteri S. 2019. Assessing inter-individual differences with task-related functional neuroimaging. *Nat Hum Behav*. 3(9):897–905.
- Lee SM, Gao T, McCarthy G. 2012. Attributing intentions to random motion engages the posterior superior temporal sulcus. *Soc Cogn Affect Neurosci*. 9(1):81–87.
- Molinari M, Leggio MG, Thaut MH. 2007. The cerebellum and neural networks for rhythmic sensorimotor synchronization in the human brain. *Cerebellum*. 6(1):18–23.
- Moore CD, Cohen MX, Ranganath C. 2006. Neural mechanisms of expert skills in visual working memory. *J Neurosci*. 26(43):11187–11196.
- Mueller VA, Brass M, Waszak F, Prinz W. 2007. The role of the preSMA and the rostral cingulate zone in internally selected actions. *NeuroImage*. 37(4):1354–1361.
- Münste TF, Altenmüller E, Jäncke L. 2002. The musician's brain as a model of neuroplasticity. *Nat Rev Neurosci*. 3(6):473.
- Nakagawa S, Schielzeth H. 2013. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods Ecol Evol*. 4(2):133–142.
- Nandi D, Aziz TZ, Liu X, Stein JF. 2002. Brainstem motor loops in the control of movement. *Mov Disord*. 17(S3):S22–S27.
- Nelder J. 1977. A reformulation of linear models. *J R Stat Soc Ser A Stat Soc*. 140(1):48–63.
- Pezzulo G, Candidi M, Dindo H, Barca L. 2013. Action simulation in the human brain: twelve questions. *New Ideas Psychol*. 31(3):270–290.
- Rahnev D, Denison RN. 2018. Suboptimality in perceptual decision making. *Behav Brain Sci*. 41:e223.
- Rao SM, Mayer AR, Harrington DL. 2001. The evolution of brain activation during temporal processing. *Nat Neurosci*. 4(3):317.
- Ratcliff R, McKoon G. 2008. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput*. 20(4):873–922.
- Reid AT, Headley DB, Mill RD, Sanchez-Romero R, Uddin LQ, Marinazzo D, Lurie DJ, Valdés-Sosa PA, Hanson SJ, Biswal BB. 2019. Advancing functional connectivity research from association to causation. *Nat Neurosci*. 22(11):1751–1760.
- Repp BH, Keller PE. 2010. Self versus other in piano performance: detectability of timing perturbations depends on personal playing style. *Exp Brain Res*. 202(1):101–110.
- Rizzolatti G, Fadiga L, Gallese V, Fogassi L. 1996. Premotor cortex and the recognition of motor actions. *Cogn Brain Res*. 3(2):131–141.
- Rizzolatti G, Sinigaglia C. 2010. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat Rev Neurosci*. 11(4):264–274.
- Ruby P, Decety J. 2001. Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat Neurosci*. 4(5):546–550.

- Saygin AP. 2007. Superior temporal and premotor brain areas necessary for biological motion perception. *Brain*. **130**(9): 2452–2461.
- Saygin AP, Wilson SM, Hagler DJ, Bates E, Sereno MI. 2004. Point-light biological motion perception activates human premotor cortex. *J Neurosci*. **24**(27):6181–6188.
- Schaerlaeken S, Grandjean D, Glowinski D. 2017. Playing for a Virtual Audience: The Impact of a Social Factor on Gestures, Sounds and Expressive Intents. *Appl Sci*. **7**(12): 1321.
- Schubotz RI. 2007. Prediction of external events with our motor system: towards a new framework. *Trends Cogn Sci*. **11**(5):211–218.
- Sebanz N, Knoblich G. 2009. Prediction in joint action: What, when, and where. *Top Cogn Sci*. **1**(2):353–367.
- Sevdalis V, Keller PE. 2011. Captured by motion: Dance, action understanding, and social cognition. *Brain Cogn*. **77**(2): 231–236.
- Sokolov AA. 2018. The cerebellum in social cognition. *Front Cell Neurosci*. **12**:145.
- Sokolov AA, Erb M, Gharabaghi A, Grodd W, Tatagiba MS, Pavlova MA. 2012. Biological motion processing: the left cerebellum communicates with the right superior temporal sulcus. *NeuroImage*. **59**(3):2824–2830.
- Sokolov AA, Gharabaghi A, Tatagiba MS, Pavlova M. 2010. Cerebellar engagement in an action observation network. *Cereb Cortex*. **20**(2):486–491.
- Sokolov AA, Miall RC, Ivry RB. 2017. The cerebellum: adaptive prediction for movement and cognition. *Trends Cogn Sci*. **21**(5):313–332.
- Spunt RP, Lieberman MD. 2012. Dissociating modality-specific and supramodal neural systems for action understanding. *J Neurosci*. **32**(10):3575–3583.
- Stewart L, Henson R, Kampe K, Walsh V, Turner R, Frith U. 2003. Brain changes after learning to read and play music. *NeuroImage*. **20**(1):71–83.
- Treue S, Maunsell JH. 1996. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*. **382**(6591):539–541.
- Treue S, Maunsell JH. 1999. Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J Neurosci*. **19**(17):7591–7602.
- Turella L, Pierno AC, Tubaldi F, Castiello U. 2009. Mirror neurons in humans: Consisting or confounding evidence? *Brain Lang*. **108**(1):10–21.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*. **15**(1):273–289.
- Van Overwalle F, Baetens K, Mariën P, Vandekerckhove M. 2014. Social cognition and the cerebellum: a meta-analysis of over 350 fMRI studies. *NeuroImage*. **86**:554–572.
- Whitfield-Gabrieli S, Nieto-Castanon A. 2012. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect*. **2**(3):125–141.
- Wilson M, Knoblich G. 2005. The case for motor involvement in perceiving conspecifics. *Psychol Bull*. **131**(3):460.
- Wöllner C. 2017. *Body, sound and space in music and beyond: Multimodal explorations*. New York, NY: Routledge (Taylor & Francis Group).
- Wöllner C, Cañal-Bruland R. 2010. Keeping an eye on the violinist: motor experts show superior timing consistency in a visual perception task. *Psychol Res*. **74**(6):579–585.
- Wolpert DM, Doya K, Kawato M. 2003a. A unifying computational framework for motor control and social interaction. *Philos Trans R Soc Lond B Biol Sci*. **358**(1431):593–602.
- Wolpert DM, Doya K, Kawato M. 2003b. A unifying computational framework for motor control and social interaction. *Philos Trans R Soc Lond B Biol Sci*. **358**(1431):593–602.
- Wolpert DM, Kawato M. 1998. Multiple paired forward and inverse models for motor control. *Neural Networks*. **11**(7–8):1317–1329.
- Zapparoli L, Seghezzi S, Scifo P, Zerbi A, Banfi G, Tettamanti M, Paulesu E. 2018. Dissecting the neurofunctional bases of intentional action. *PNAS U S A*. **115**(28):7440–7445.
- Zatorre RJ, Chen JL, Penhune VB. 2007. When the brain plays music: auditory–motor interactions in music perception and production. *Nat Rev Neurosci*. **8**(7):547–558.