

## RESEARCH ARTICLE

# Contextual effects on loudness judgments for sounds with continuous changes of intensity are reflected in nonauditory areas

Oliver Behler  | Stefan Uppenkamp

Medical Physics and Cluster of Excellence "Hearing4all", Department of Medical Physics and Acoustics, Faculty VI Medicine and Health Sciences, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

## Correspondence

Oliver Behler, Medical Physics, Department of Medical Physics and Acoustics, Faculty VI Medicine and Health Sciences, Carl von Ossietzky Universität Oldenburg, 26111 Oldenburg, Germany.  
Email: oliver.behler@uni-oldenburg.de

## Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Numbers: Cluster of Excellence EXC1077 Hearing4all, 390895286; Niedersächsisches Ministerium für Wissenschaft und Kultur, Grant/Award Number: PhD programme 'Signals; Cognition'

## Abstract

Psychoacoustic research suggests that judgments of perceived loudness change differ significantly between sounds with continuous increases and decreases of acoustic intensity, often referred to as "up-ramps" and "down-ramps." The magnitude and direction of this difference, in turn, appears to depend on focused attention and the specific task performed by the listeners. This has led to the suspicion that cognitive processes play an important role in the development of the observed context effects. The present study addressed this issue by exploring neural correlates of context-dependent loudness judgments. Normal hearing listeners continuously judged the loudness of complex-tone sequences which slowly changed in level over time while auditory fMRI was performed. Regression models that included information either about presented sound levels or about individual loudness judgments were used to predict activation throughout the brain. Our psychoacoustical data confirmed robust effects of the direction of intensity change on loudness judgments. Specifically, stimuli were judged softer when following a down-ramp, and louder in the context of an up-ramp. Levels and loudness estimates significantly predicted activation in several brain areas, including auditory cortex. However, only activation in nonauditory regions was more accurately predicted by context-dependent loudness estimates as compared with sound levels, particularly in the orbitofrontal cortex and medial temporal areas. These findings support the idea that cognitive aspects contribute to the generation of context effects with respect to continuous loudness judgments.

## KEYWORDS

auditory cortex, auditory fMRI, context effects, loudness judgments, orbitofrontal cortex

## 1 | INTRODUCTION

Most of our understanding about how perceived loudness relates to the sound pressure level and other physical characteristics (e.g., the spectral content) of acoustic events has been established through studies using stationary sounds of rather short duration in the

laboratory. In contrast, most sounds in our daily life are dynamic and may vary considerably over time, which affects their loudness.

Although elaborate models have been developed to predict the loudness of time-varying sounds (Chalupper & Fastl, 2002; Glasberg & Moore, 2002), these models are still limited in their capabilities (e.g., Oberfeld, Heeren, Rannies, & Verhey, 2012; Oberfeld &

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

Plank, 2011; Rannies, Verhey, & Fastl, 2010). This might be particularly true in regard to sounds with continuous increases or decreases of intensity, often referred to as “up-ramps” and “down-ramps.” For instance, psychoacoustic research suggests that the perceived change of loudness can differ significantly between up-ramps and down-ramps, despite identical absolute changes in level (for a review, see Olsen, 2014). The direction and magnitude of this perceptual asymmetry appears to depend on several variables, which include physical characteristics of the stimuli (e.g., dynamic range, start or end level, duration), but also the psychoacoustic assessment procedure: If listeners are asked directly to judge the loudness change following the presentation of a ramp stimulus (retrospective “global loudness change”), this change is usually judged to be greater for up-ramps relative to down-ramps (e.g., Bach, Neuhoﬀ, Perrig, & Seifritz, 2009; Neuhoﬀ, 1998; Olsen, Stevens, & Tardieu, 2010; Seifritz et al., 2002). If, on the other hand, listeners are instructed to make judgments about their current loudness perception repeatedly or continuously throughout the stimulus, the loudness change inferred from these data (i.e., the difference between judgments at the beginning and end of each ramp) tends to be greater in response to down-ramps (termed as “decrutment,” e.g., Canévet & Scharf, 1990; Canévet, Teghtsoonian, & Teghtsoonian, 2003; Olsen, Stevens, Dean, & Bailes, 2014; Schlauch, 1992; Susini, McAdams, & Smith, 2007; Teghtsoonian, Teghtsoonian, & Canévet, 2000). In an attempt to reconcile these two conﬂictive phenomena, it has been argued that judgments of global loudness change and judgments of momentary loudness may reflect different underlying mechanisms (Neuhoﬀ, 1999).

Irrespective of this distinction, it has been deemed unlikely that the perceptual overestimation of up- or down-ramps can be fully explained by early sensory mechanisms such as temporal masking (Olsen & Stevens, 2012) or simple adaptation (Teghtsoonian et al., 2000), which typically evolves more gradually over longer time scales. Moreover, there is evidence that perceptual outcomes may be inﬂuenced by the order in which diﬀerent ramps are presented (Olsen et al., 2010, 2014). Importantly, the focus of attention toward the acoustic stimulus seems to play a critical role in the decrutmment caused by down-ramps (Schlauch, 1992). Hence, it rather appears that some combination of sensory and cognitive (e.g., memory) mechanisms are at play (e.g., Olsen, 2014; Olsen et al., 2010; Schlauch, 1992), whose relative contributions are not yet completely understood.

Noninvasive neuroimaging techniques such as functional magnetic resonance imaging (fMRI) have provided a detailed picture about how perceived loudness in response to stationary sounds is represented in the human auditory pathway, including the cortex. Speciﬁcally, several auditory fMRI studies suggest that individual loudness perception for these sounds is most closely (and linearly) related to activation, as inferred from blood oxygenation level dependent (BOLD) signals, in auditory cortex (AC, Behler & Uppenkamp, 2016; Hall et al., 2001; Langers et al., 2007; Röhl & Uppenkamp, 2012), more precisely in the posteromedial part of Heschl's gyrus (HG, Behler & Uppenkamp, 2016). Much less is known about the neural representation of loudness for time varying sounds.

Two fMRI studies that have compared sounds (pulsed tones) with intensity up-ramps versus down-ramps over 2 s found increased activation for the up-ramps in temporal and parietal areas thought to be involved in the processing of auditory motion (Bach et al., 2008; Seifritz et al., 2002), and in the amygdala (Bach et al., 2008). These results were considered as evidence of an evolved perceptual bias to looming auditory motion, which provides a warning cue for an approaching (and potentially dangerous) sound source (Neuhoﬀ, 1998). Yet, even though the up-ramp stimuli used in both studies produced greater estimates of loudness change relative to the corresponding down-ramps, which agrees with the theory of looming motion, it is not clear to what degree the fMRI activation patterns reﬂected diﬀerences in perceived loudness or other reactions and associations elicited by the stimuli. To our knowledge, only one auditory fMRI study (Lehne, Rohrmeier, & Koelsch, 2013) and two studies that combined electro- and magneto-encephalography (EEG and MEG, Thwaites et al., 2015; Thwaites, Glasberg, Nimmo-Smith, Marslen-Wilson, & Moore, 2016) have explicitly investigated brain activation in relation to ﬂuctuations of loudness for time varying sounds (musical pieces in Lehne et al., 2013; speech in Thwaites et al., 2015, 2016). All three studies found correlates of loudness mainly in primary AC and adjacent auditory association areas. However, it should be noted that in the analysis conducted by Lehne et al. (2013), the variable of main interest was musical tension as judged by the participants, whereas loudness estimates were only included as a confounding variable. Since both variables were highly intercorrelated, their results are somewhat diﬃcult to interpret. More importantly, in all three studies, estimates of perceived loudness were not derived from individual judgments, but instead calculated via loudness models, which neither capture diﬀerences between listeners in terms of perception, nor possible contextual effects as those discussed above.

In the present study, we sought to speciﬁcally address these issues. For this purpose, normal hearing listeners continuously judged their perceived loudness of acoustic stimuli with dynamic intensity while auditory fMRI was performed. The inﬂuence of acoustic scanner noise was reduced by means of an active noise cancellation headphone system (Chambers, Bullock, Kahana, Kots, & Palmer, 2007). Acoustic stimuli were square waves with sequences of slow (10-s) intensity up-ramps and down-ramps, stringed together by shorter (3-s) stationary parts. Based on the psychoacoustic literature delineated above, we expected this composition to elicit strong contextual effects, and thereby signiﬁcant divergence between presented levels and corresponding loudness judgments. Although it has been demonstrated that these effects can also be elicited by more natural sounds such as vowels, instrumental notes and melodies (Olsen et al., 2010, 2014), we opted for rather simple synthetic stimuli to avoid possible confounds due to changes of spectral content, lexical semantics or other associations. Using a cross-validation approach, we compared regression models which included presented levels or individual loudness estimates as explanatory variables in terms of their accuracy and robustness in predicting fMRI activation throughout the brain.

Lastly, in addition to judgments of context-speciﬁc loudness at every moment, we also included individual estimates of “context-

unspecific loudness” in our analyses that were calculated by averaging across data in different contexts throughout the experiment. This step was done to provide results that are more directly comparable to previous auditory fMRI studies investigating individual loudness perception for stationary sounds, which have typically used estimates based on fitted data across multiple stimulus presentations outside the scanner. Based on the existing literature, we hypothesized that activation in AC is more accurately predicted by both types of loudness estimates relative to the physical levels of the acoustic stimuli. If, however, cognitive contributions play an important role in the development of contextual effects on loudness judgments, we should expect this to be reflected by activation in frontal and other non-auditory areas of the brain, which are thought to be involved in higher cognitive processing.

## 2 | METHODS

### 2.1 | Participants

Twenty-five normal hearing volunteers were recruited at the University of Oldenburg and gave written informed consent to participate in this study. One male and one female participant were excluded from the analysis due to excessive head movement during the fMRI experiment and/or ambiguous loudness judgments (more than 10% of data rejected based on the criteria described below). The remaining sample comprised 14 female and 9 male participants, ranging from 18 to 33 years of age (mean: 24 years). All participants had hearing thresholds of 20 dB HL or better in the frequency range from 125 Hz to 8 kHz, as tested by means of standard pure tone audiometry with a clinical audiometer and Sennheiser HDA 200 Headphones (Sennheiser electronic GmbH & Co. KG, Wedemark, Germany). A questionnaire was used to ensure that subjects had no conditions contraindicative for MRI. The study was approved by the ethics committee of the University of Oldenburg.

### 2.2 | Acoustic setup and stimuli

Stimuli were delivered binaurally via an MRI compatible, opto-acoustic headphone system capable of providing a wide frequency response (OptoActive™, Optoacoustics Ltd, Or Yehuda, ISR). During the functional MRI sequences, this headphone system was also used for active noise cancelation (Chambers et al., 2007) to further reduce the scanner gradient noise beyond passive attenuation provided by the headphones and additional foam pillows used for head fixation. Specifically, the active noise cancelation system achieves a frequency-dependent attenuation of the scanning noise, with about 30 dB at its spectral peak at 1 kHz and a broadband attenuation of around 10 dB (from 60 Hz to 12 kHz), which results in a much flatter noise spectrum (these numbers represent averages obtained from acoustic measurements during a separate study in our lab with the respective functional MRI sequence).

All stimuli were square waves composed of odd harmonics 1–11, with a fundamental frequency of 440 Hz, created at a sampling rate of 44.1 kHz and 24 bit depth. Stimulus levels either increased or decreased linearly over time on a dB scale, with stationary or silent parts in between intensity ramps. More details are given below.

All experiments were programmed and presented using MATLAB 2014 (Mathworks, Natick, MA) and the Cogent 2000 toolbox (v125, <http://www.vislab.ucl.ac.uk/cogent.php>, London, UK).

### 2.3 | fMRI setup and data acquisition

The fMRI measurements were done using a 3-Tesla scanner (Magnetom Prisma 3 T, Siemens AG, Erlangen, Germany), equipped with a 20-channel head coil. The response scale and additional instructions (see below) were projected onto a screen in the scanner bore. Participants saw the screen via a mirror construction mounted onto the head coil. Behavioral responses were collected by means of an fMRI-compatible response pad (LXPAD-2x5-10 M, NAtA technologies, Coquitlam, Canada).

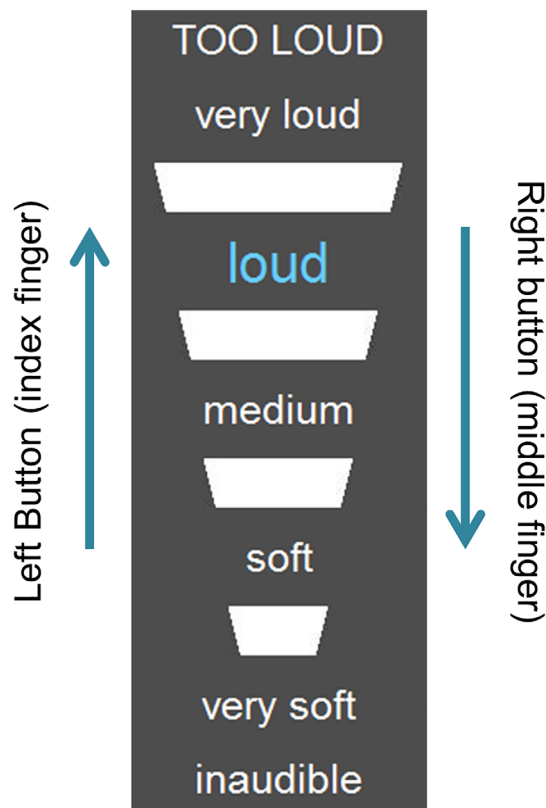
Functional images were obtained using a T2\*-weighted gradient echo planar imaging (EPI) sequence (continuous imaging at TR 1.5 s, echo time 30 ms, flip angle 90°). Every image comprised 24 slices (in-plane field of view 204 × 204 mm<sup>2</sup>, 94 × 94 voxels, voxel size 2.17 × 2.17 × 5 mm<sup>3</sup>, distance factor 10%) in axial orientation, acquired in ascending interleaved order.

The fMRI experiment comprised three functional runs. In the first run, the number of scans varied across participants depending on their behavioral responses (see below). In the second and third run, 541 images were collected per run. Due to the time required by the active noise cancelation system for “learning” the EPI noise and adjusting its countermeasures, the first 16 images of every run were removed from the data set before preprocessing and analyses.

After completion of the functional MRI experiment, high-resolution structural images were acquired using a T1-weighted magnetization-prepared rapid acquisition gradient echo sequence (voxel size 0.7 × 0.7 × 0.9 mm<sup>3</sup>, distance factor 50%, TR = 2 s, TE = 2.41 ms, FA = 9°, FoV = 230 × 194 × 187 mm<sup>3</sup>).

### 2.4 | Continuous categorical loudness scaling

The response scale used in the present study followed the ISO 16832:2006 standard, as originally proposed by Brand and Hohmann (2002). It comprised 11 response alternatives, including seven named loudness categories—“inaudible,” “very soft,” “soft,” “medium,” “loud,” “very loud,” and “too loud”—and four unnamed intermediate categories. The response scale was displayed in white on a dark gray background. Whenever a stimulus was playing, one response alternative was highlighted by means of a different color (light blue) and slightly increased size (Figure 1). Participants were asked to continuously adjust the highlighted response alternative via button presses, so that it always reflected their current perception of the stimulus. They were instructed to press one button with their right index



**FIGURE 1** Categorical loudness scale with 11 response alternatives. The trapezoids represent four unnamed intermediate categories. The category “loud” is currently highlighted. Participants continuously adjusted the highlighted category via button presses, as indicated, so that it always reflected their current perception of the stimulus. The responses were transformed into “categorical units” ranging from 0 (“inaudible”) to 50 (“too loud”) and collected at a sampling rate of 10 Hz

finger and another with their right middle finger. Pressing the first one once moved the highlighted category up the scale by one category, pressing the other moved it down the scale. The category did not change by more than one position if either button was held depressed.

The time course of these continuous loudness judgments was collected with a sampling rate of 10 Hz. To quantify individual loudness perception, the 11 response alternatives were transformed into their corresponding numerical values ranging from 0 to 50 categorical units (cu) in steps of 5 cu.

All participants completed a short training on a personal computer outside of the MR scanner room to get familiar with the continuous loudness scaling procedure and the stimuli played during the fMRI experiment. The training included shortened versions of the detection threshold and UDL estimation as well as the main experiment (see below) and took around 4 min.

## 2.5 | Assessment of detection thresholds and UDL

In the first functional run, participants were presented with stimuli that either increased or decreased continuously in level at a rate of

1 dB per second, with a starting level of 70 dB SPL. When the stimulus was judged as “too loud” or “inaudible,” the current level was considered as detection threshold level (DTL) or uncomfortable loudness level (UCL) and the stimulus immediately stopped playing. The stimulus also stopped playing if it reached the maximum level of 105 or 0 dB SPL, which however did not occur in this group of listeners. After a silent interval of 5 s, the next stimulus started playing. The direction of intensity changes alternated with every presentation until three up- and three down-ramps were played. The medians of the corresponding DTLs and UCLs were used as individual parameter estimates for the main experiment.

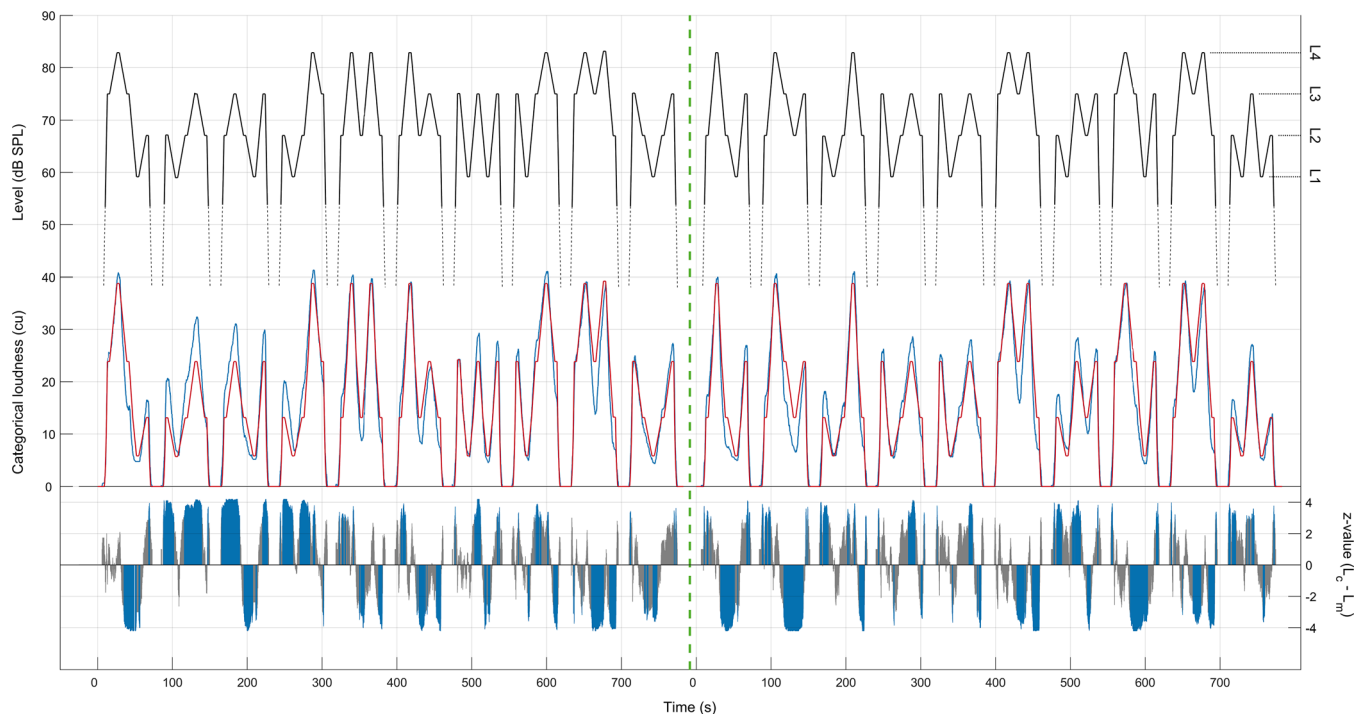
## 2.6 | Main experiment

In the second and third functional run, participants were presented with continuous stimuli composed of stationary parts with 3 s duration, interconnected by up- or down-ramps with 10 s duration that increased or decreased linearly in dB over time. Every sixth stationary part was silent (0 dB SPL), which effectively created stimulus “blocks” with a duration of 75 s, separated by silent intervals. Over the course of both functional runs, 20 such blocks were presented (10 per run). At every nonsilent stationary part, one of four different levels (L1, L2, L3, L4) was presented. These levels were adjusted to the individual dynamic range of the listener. Specifically, for each participant, four levels were chosen that were equally spaced in level between their estimated DTL plus 6 dB and their estimated UCL minus 6 dB. The order of levels at the stationary parts was the same for all participants and was created pseudo-randomized with the following constraints: (a) The first and last level of each block had to be either L2 or L3; (b) L1 must not be followed by L4, or vice versa; (c) Each of the remaining 10 possible up- and down-ramps (L1 → L2, L2 → L1, L1 → L3, etc.) is presented 10 times per run (see Figure 2, panel a). The first two constraints were included to ensure that changes of perceived loudness were slow enough to allow the participants to faithfully track them. This was deemed necessary since every change from one loudness category to the next higher or lower one required an individual button press. Transitions from silent parts and L1 to the top level L4 were avoided to prevent uneasiness and head movement induced by very sharp rises of loudness. Starting or ending stimulus blocks with the softest level, on the other hand, was avoided because pilot experiments showed that it introduced high uncertainty when judging the presence or absence of stimuli even after L1 or “silence” had been reached.

## 2.7 | Preprocessing of level and loudness data

The presented sound levels and loudness judgments obtained in the main experiment were preprocessed to produce the following three explanatory variables that were entered into the statistical analyses (see Figure 3):

1. Sensation level ( $L_5$ ): The median DTL from the data in the first run were subtracted from the presented sound pressure levels. All



**FIGURE 2** Time course of level and loudness estimates in the main experiment. The upper trace shows the group averaged sound pressure levels (in dB, black lines) as a function of time (in seconds). Levels below the group averaged detection threshold are represented by dashed lines and the lowest levels are hidden for convenience. The locations of the averaged stationary levels from L1 to L4 are indicated on the right. The green dashed line separates data from the first and second functional run. The middle trace shows the group averaged mean loudness ( $L_m$ , in categorical units, red line) overlaid onto the group averaged context loudness ( $L_c$ , blue line). Panel C shows the z-values obtained from frame-wise Wilcoxon signed-ranked-tests between  $L_c$  and  $L_m$  across participants. Significant values (at  $p < .05$ , FDR-corrected) are displayed in blue, nonsignificant values are displayed in gray

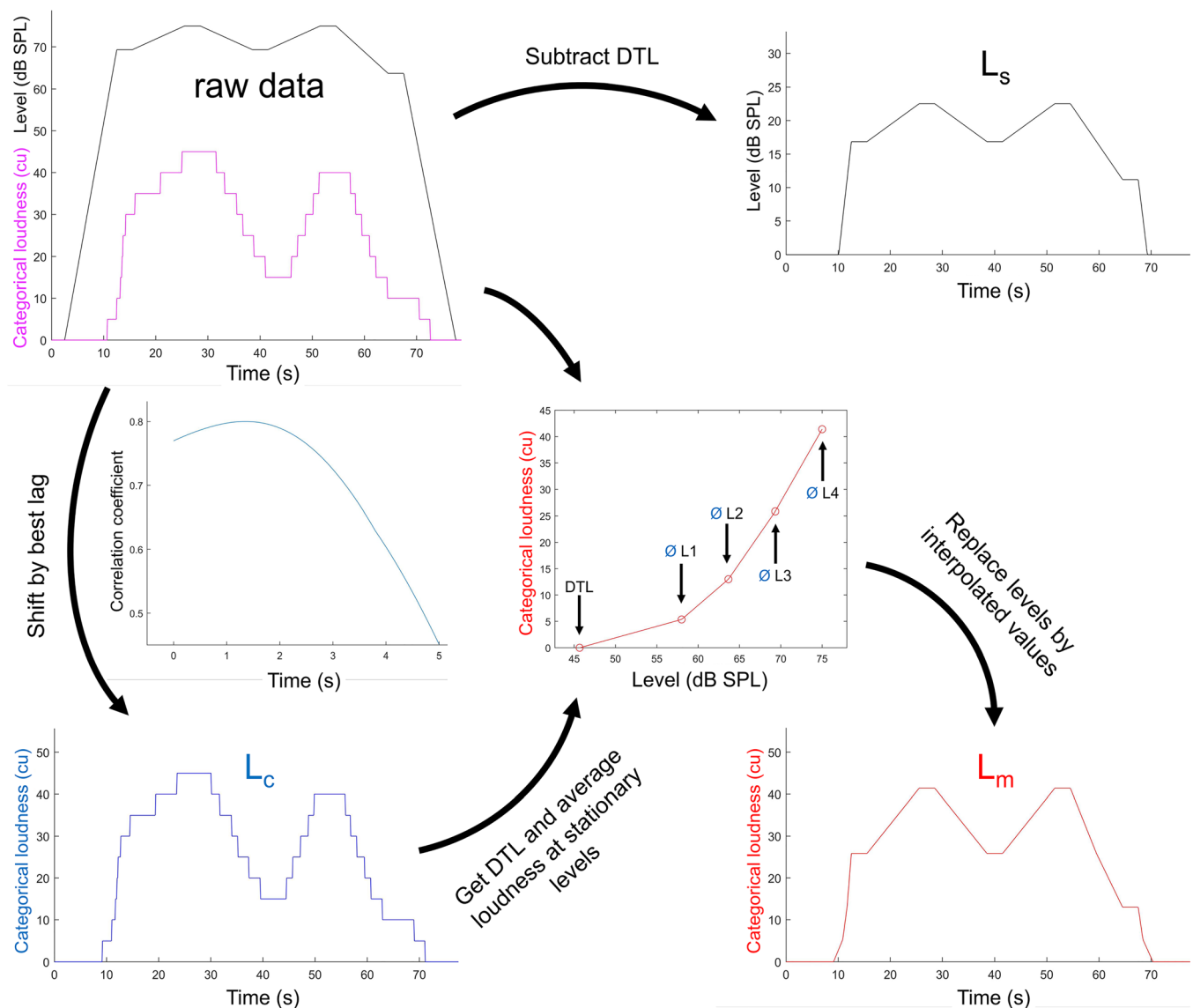
values below zero in the level time course were subsequently set to zero. The resulting time course represents the physical sound level, corrected for individual detection thresholds.

- Context loudness ( $L_c$ ): Within each stimulus block, cross-correlation was performed between the continuous loudness judgments and sound levels for all samples above 30 dB SPL (below which stimuli were virtually always judged inaudible). The loudness time course was then shifted back in time according to the highest correlation coefficient in the range from 0 to 5 s. Occasional loudness judgments above 0 cu corresponding to levels below 30 dB SPL (which were considered artifacts) were removed in this process. The resulting time course represents the individual and context-specific loudness judgments at every instance, corrected for response delay.
- Mean loudness ( $L_m$ ): First, the individual DTL was re-estimated by averaging across the levels corresponding to every first and last audible sample of each block ( $L_c > 0$ ) throughout both functional runs. Then, the average categorical loudness units corresponding to every stationary level (L1 to L4) were calculated, using all samples of  $L_c$  across the experiment corresponding to the respective levels. Linear interpolation was performed between the new DTL (with loudness defined as 0 cu) and the four averaged loudness values to produce a level-to-loudness function. Finally, all levels across both functional runs were replaced by the corresponding

categorical loudness units according to this function. The resulting time course represents the average loudness perception of the individual listener for the momentary level. It was designed to be devoid of contextual (i.e., stimulus history) effects, as the same loudness is assumed for a given level irrespective of the context it is presented in.

The third step was preceded by a visual inspection of the sensation level and context loudness time courses by the authors for the purpose of artifact detection and to assess the quality of the delay correction. Specifically, for every participant and each block of the experiment, they labeled the time frames in the context loudness time course corresponding to every stationary part in the level time course to the best of their abilities. When both authors independently concluded that the judgments were too ambiguous to complete this task, the respective stationary part and both adjacent ramps were marked for rejection and ignored in all further analyses, and in the calculation of the mean loudness. In most cases, this decision was based on the lack of change with respect to loudness judgments over a longer period despite large changes of level (e.g., L2  $\rightarrow$  L4). In total, less than 0.5% of all samples were rejected due to this criterion.

For the statistical analyses of fMRI data,  $L_s$ ,  $L_c$ , and  $L_m$  were convolved with the canonical hemodynamic response function and down-sampled to the fMRI sample rate of 1/1.5 Hz.



**FIGURE 3** Preprocessing of individual level and loudness data. The scheme illustrates the construction of the three explanatory variables used in the statistical analyses from the presented sound pressure level (upper left, black line) and continuous loudness judgments (upper left, purple line) collected during one stimulus block in the main experiment. First, the detection threshold level (DTL), as assessed before the main experiment, was subtracted from the sound pressure level. This produced the individual “sensation level” ( $L_s$ , upper right). Then, the raw loudness time course was shifted back in time to correct for the individual response delay according to cross-correlation between the loudness judgments and  $L_s$  (center left). The resulting time course is referred to as “context loudness” ( $L_c$ , lower left). Finally, the average categorical loudness corresponding to every stationary level (L1 to L4) in the main experiment was calculated and the DTL was re-estimated using  $L_c$ . Linear interpolation between the respective five levels was used to produce a level-to-loudness function (center-right). The “mean loudness” time course ( $L_m$ , lower right) was constructed by replacing all presented levels with the corresponding categorical loudness of this function

### 2.8 | Statistical analysis of psychoacoustic data

First, to characterize the overall similarity of the preprocessed level and loudness time courses, Pearson correlation coefficients were calculated for every pair of the three abovementioned variables as well as the sound pressure level. Differences between context loudness and mean loudness across participants were assessed by means of a two-tailed Wilcoxon signed-rank test at every time sample of the main experiment, using the “signrank” function in MATLAB 2019b

(Mathworks, Natick, MA) with the “approximate” method, which calculates the p-value using the z-statistic, given by

$$z = \frac{\left(W - \frac{n(n+1)}{4}\right)}{\sqrt{\frac{n(n+1)(2n+1) - tiedadj}{24}}}$$

where  $W$  is the sum of the ranks of positive differences between the observations in the two samples (i.e.,  $L_c - L_m$ ),  $n$  is the number of



differences, and *t*<sub>adj</sub> represents an adjustment value for ties. The results were thresholded with a significance level of  $p < .05$ , corrected for false discovery rate (FDR).

To further assess the effects of level and context on the loudness judgments, for each participant, we extracted the average  $L_c$  value during every presentation of the medium stationary levels L2 and L3. We then calculated the individual's average  $L_c$  for both levels separately when presented after a down-ramp and when presented after an up-ramp, excluding those instances at the end of stimulus onset ramps starting from silence. The resulting values were analyzed in relation to level (L2, L3) and ramp direction (up, down) by means of a linear mixed effects model with random intercept (participants as random variable). In addition, an interaction between level and ramp direction was included in the model. Model parameters were estimated by means of maximum likelihood using the “fitlme” function with default settings in MATLAB 2019b. The end-level loudness values for L1 and L4 were not included in this analysis, since both levels were always presented in the context of a down- or up-ramp, respectively.

Finally, for each participant, we also extracted the average loudness change associated with every up- and down-ramp in the experiment, as defined by the difference between the averaged  $L_c$  values at the samples corresponding to the stationary part preceding the respective ramp and those following it. Absolute loudness changes were then analyzed in relation to ramp direction (up, down), ramp size (small, large; i.e., changes over one vs. two stationary levels) and intensity region (1 to 5; increases with the average level of ramps, where 1 corresponds to L1↔L2 and 5 to L3↔L4) by means of a linear mixed effects model with random intercept. An interaction between ramp direction and intensity region was included in the model.

## 2.9 | Preprocessing of fMRI data

Preprocessing of the functional and structural imaging data was done using the SPM12 toolbox (FIL, Wellcome Trust Centre for Neuroimaging, University College London, London, UK, <http://www.fil.ion.ucl.ac.uk/spm>) and custom scripts in MATLAB.

Functional images were corrected for slice acquisition times and realigned to the first image of the first functional run by means of rigid body spatial transformation. The structural image was co-registered to the averaged functional image. It then underwent segmentation to produce (a) a set of individual (posterior) probability maps for different tissue types, (b) a structural image (bias-)corrected for intensity non-uniformity, and (c) a forward deformation field that encodes the information required for spatially normalizing the structural image and the statistical parametric maps described below to Montreal Neurological Institute (MNI) space.

Next, trials with excessive head movement were detected. For this purpose, we used the estimated realignment parameters to calculate a scalar sample-wise (image-to-image) displacement as described in Power, Barnes, Snyder, Schlaggar, and Petersen (2012). With respect to the contribution of rotational movement, an average voxel

distance of 50 mm from the center of the head was assumed. Samples with an absolute displacement above 0.5 mm, together the previous and subsequent samples ( $\pm 1$ ), were marked for rejection and ignored in all further analyses (1.1% of samples were rejected due to this on average per participant).

Low frequency drifts and other stimulus-unrelated signal fluctuations were attenuated by means of multiple linear regression. The regressors comprised a constant and linear term, a set of sine and cosine functions with one, two, and three periods per run (representing a high-pass filter with 1/256 Hz cut-off frequency), and the averaged signal from voxels located in the cerebrospinal fluid (as defined by an eroded mask obtained from the individual probability maps). All regressors were fit to the signal time course of every voxel as well as to the explanatory variables of interest described above ( $L_s$ ,  $L_c$ , and  $L_m$ ). The residuals of these regressions were then used in the statistical analyses.

## 2.10 | Statistical analyses of fMRI data

The following analyses were performed on the functional data in individual space, for every voxel within gray matter masks obtained from individual tissue probability maps. Using 10-fold cross-validation, we assessed and compared the robustness of the following models in terms of predicting fMRI activation using custom functions in MATLAB 2019b:

- Baseline: Only a constant term
- Sensation level: Baseline plus  $L_s$
- Context loudness: Baseline plus  $L_c$
- Mean loudness: Baseline plus  $L_m$

The data of each participant that were not rejected were split into 10 equally sized parts (the first 10% of samples, the following 10%, and so on). Then, a 10-fold cross-validation procedure was performed for each of the models described above: In each fold, the respective model was fitted by means of ordinary least squares regression to nine parts of the data (the “training data set”). The resulting beta estimates were used on the regressors of the remaining part (the “validation data set”) to predict the corresponding fMRI data. To avoid temporal dependence of training and validation data sets, 32 s worth of samples (i.e., the length of the informed basis set functions) directly adjacent to the test data set were discarded before fitting models to the training data set, a modification known as “hv-block” cross-validation (Racine, 2000). Collapsing over all folds, a cross-validated  $R^2$  was calculated.

The resulting cross-validated  $R^2$  maps for each model and the bias-corrected structural image were then normalized to MNI space based on the deformation field calculated in the process of structural segmentation. During the normalization procedure, functional images were resampled to  $3 \times 3 \times 3$  mm voxel size. Finally, the  $R^2$  maps were spatially smoothed with an isotropic Gaussian kernel of 8 mm full width at half maximum.

The differences between models in terms of prediction performance were then statistically assessed via two-tailed Wilcoxon signed-rank tests (as described above) across the cross-validated  $R^2$ s of all participants. These tests were performed separately in every voxel that was characterized by a gray matter probability above 50% for each participant and a group averaged cross-validated  $R^2$  above zero for at least one of the two contrasted models. The resulting statistical (z-)maps were thresholded at a significance level of  $p < .05$ , FDR-corrected, which was extended to a minimum cluster-size of at least 10 adjacent significant voxels. For the purpose of anatomical localization, thresholded maps were overlaid onto the group averaged normalized structural image, using the MRIcron software (Version 12015; Chris Rorden, <https://people.cas.sc.edu/rorden/mricron/>). Structures corresponding to significant clusters were determined by means of the aal (Tzourio-Mazoyer et al., 2002) labels database.

### 3 | RESULTS

#### 3.1 | Psychoacoustic results

The DTL estimates assessed in the first functional run ranged from 42 to 62 dB SPL (group average and SD  $53.3 \pm 4.5$  dB SPL). In comparison, the DTL estimates calculated from the loudness judgments in the main experiment were on average 5 dB lower ( $48.3 \pm 3.0$  dB SPL). Estimates of UCL (from the first run) ranged from 79 to 99 dB SPL ( $88.8 \pm 5.5$  dB SPL). The largest individual dynamic range between L4 and L1 was 36 dB, the smallest was 10 dB ( $23.7 \pm 7.7$  dB). In terms of the mean loudness ( $L_m$ ), the largest corresponding difference was 39 cu and the smallest 23 cu ( $32.9 \pm 4.3$  cu). The spread of context loudness ( $L_c$ ) was obviously larger, since it compares the highest and lowest momentary loudness judgments corresponding to L4 and L1, respectively, at any point in time over the course of both runs. Specifically, it averaged to  $41.7 \pm 4.2$  cu, ranging from 35 to 50 cu across participants (one participant judged L4 as “too loud” and L1 as “inaudible” at least once in the experiment). The average delay between level and corresponding loudness judgments, as inferred from cross-correlation and corrected for in  $L_c$ , was  $770 \pm 462$  ms. The shortest delay for an individual participant (averaged across all stimulus blocks) was 150 ms, the longest was 2.2 s.

Figure 2 shows the group averaged time courses of presented sound pressure levels (upper trace), as well as  $L_c$  and  $L_m$  (middle trace) throughout both functional runs of the main experiment. Clearly, all three variables were highly inter-correlated: Correlation coefficients (Pearson's  $r$ , all significant at  $p < .001$ ) averaged to  $.75 \pm .05$  between  $L_c$  and level,  $0.81 \pm 0.04$  between  $L_m$  and level, and  $0.92 \pm 0.03$  between  $L_c$  and  $L_m$ . Naturally, after conversion of presented levels to sensation level ( $L_s$ ), correlation increased to  $r = .90 \pm .04$  with  $L_c$  and  $r = .98 \pm .02$  with  $L_m$ .

Comparing  $L_c$  to  $L_r$ , it is apparent that the mean loudness and the contextual loudness judgments often diverge, especially at and around the intermediate levels (L2 and L3). At a closer look,  $L_m$  typically “underestimates”  $L_c$  following an intensity up-ramp, whereas it

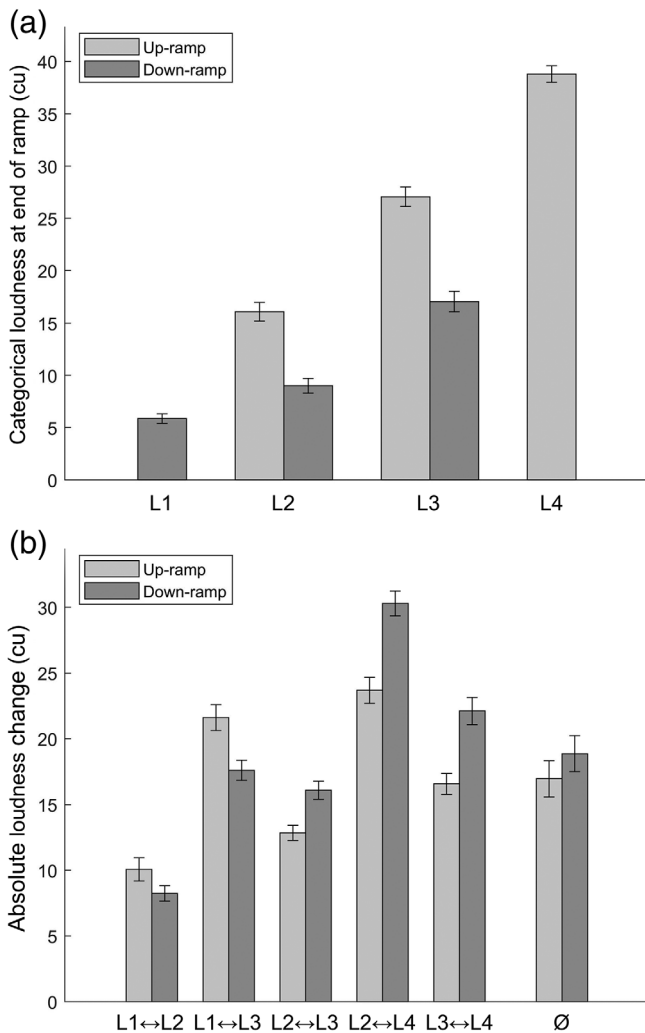
typically “overestimates”  $L_c$  following a down-ramp. As shown in the lower trace of Figure 2, the difference between both variables is significant (blue bars) a considerable amount of time (in 33% of all audible samples). When considering only the samples at stationary parts following an up-ramp,  $L_c$  is significantly higher than  $L_m$  29.2% of the time, yet lower only in 0.4% of all samples. Following a down-ramp, the situation reverses. Here,  $L_c$  is significantly lower than  $L_m$  in 52.6% and higher in only 0.9% of the respective samples.

Group averaged context loudness values at every individually adjusted stationary level in relation to the preceding ramp direction are shown in panel A of Figure 4. The linear mixed effects model analysis revealed that all investigated fixed effects were highly significant. Specifically, loudness was judged higher for L3 than L2 (coefficient estimate ( $\beta$ ) = 8.1,  $t(134) = 12.1$ ,  $p < .001$ ), higher at the end of an up-ramp versus down-ramp ( $\beta = 7.1$ ,  $t(134) = 10.6$ ,  $p < .001$ ), and the effect of ramp direction was greater at L3 than L2 ( $\beta = 2.9$ ,  $t(134) = 3.1$ ,  $p = .002$ ). Absolute context loudness changes associated with each type of intensity ramp between the stationary levels are presented in panel b of Figure 4. The linear mixed effects model analysis again yielded highly significant results for all investigated fixed effects: Loudness change was overall greater for down-ramps than for up-ramps ( $\beta = 1.0$ ,  $t[225] = 4.3$ ,  $p < .001$ ), greater for large than for small ramps ( $\beta = 4.5$ ,  $t[225] = 19.9$ ,  $p < .001$ ), and greater at high as compared low levels ( $\beta = 2.8$ ,  $t[225] = 17.8$ ,  $p < .001$ ). Moreover, the difference in judgments between up- and down-ramps was dependent on the intensity region: Greater loudness change for up-ramps than for down-ramps was observed in the two low intensity regions, whereas greater change was found for down-ramps in the higher intensity regions ( $\beta = 1.3$ ,  $t[225] = 8.1$ ,  $p < .001$ ).

#### 3.2 | Functional MRI results

As expected, cross-validated  $R^2$ s for the baseline model were virtually zero across all voxels (absolute averaged values across participants smaller than .001). By contrast, the cross-validated  $R^2$ s of the three other regression models that either included the sensation level, context loudness or mean loudness as explanatory variables were significantly greater in a large number of voxels across several areas of the brain (highest averaged values:  $L_s$ : .040,  $L_r$ : .051,  $L_m$ : .041; lowest values:  $-.001$  for all three models). The respective activation maps are presented in Figure 5. They clearly show highly similar patterns of significant clusters, characterized by more accurate predictions of fMRI activation as compared with baseline, across all three models. In the temporal lobe, these clusters include bilaterally the superior, middle, and inferior temporal gyri, Heschl's gyrus, fusiform gyrus, the hippocampus, the amygdala and the temporal pole. In the frontal lobe, they include bilaterally the precentral gyrus and supplementary motor area, the dorsolateral and medial superior frontal gyrus, middle frontal gyrus, the opercular and triangular parts of inferior frontal gyrus, the Rolandic operculum, and the orbitofrontal cortex (the orbital parts of superior, middle and inferior frontal gyrus as well as gyrus rectus). In the parietal lobe, they cover parts of the postcentral gyrus, the





**FIGURE 4** Summary statistics of psychoacoustic results: Loudness and loudness change in relation to intensity ramp direction. (Panel a) Group averaged context loudness during each stationary level (L1 to L4) that followed an up-ramp (light gray bars) or a down-ramp (darker gray bars). (Panel b) Group averaged absolute loudness changes for all possible up-ramps and down-ramps between the stationary levels. Loudness changes associated with every ramp were calculated as the difference between context loudness values at the stationary parts preceding and following the level transition. The two rightmost bars represent the averaged changes across all up- and down-ramp conditions. In both panels, error bars represent standard errors of the mean across participants

superior and inferior parietal gyrus, the precuneus, supramarginal gyrus, and angular gyrus in both hemispheres. In the occipital lobe, they include bilaterally the calcarine fissure and surrounding cortex, the cuneus, lingual gyrus, and parts of superior, middle and inferior occipital gyrus. Lastly, significant clusters were detected in the anterior, middle and posterior cingulate gyrus, bilaterally in the cerebellum, and in the right insula.

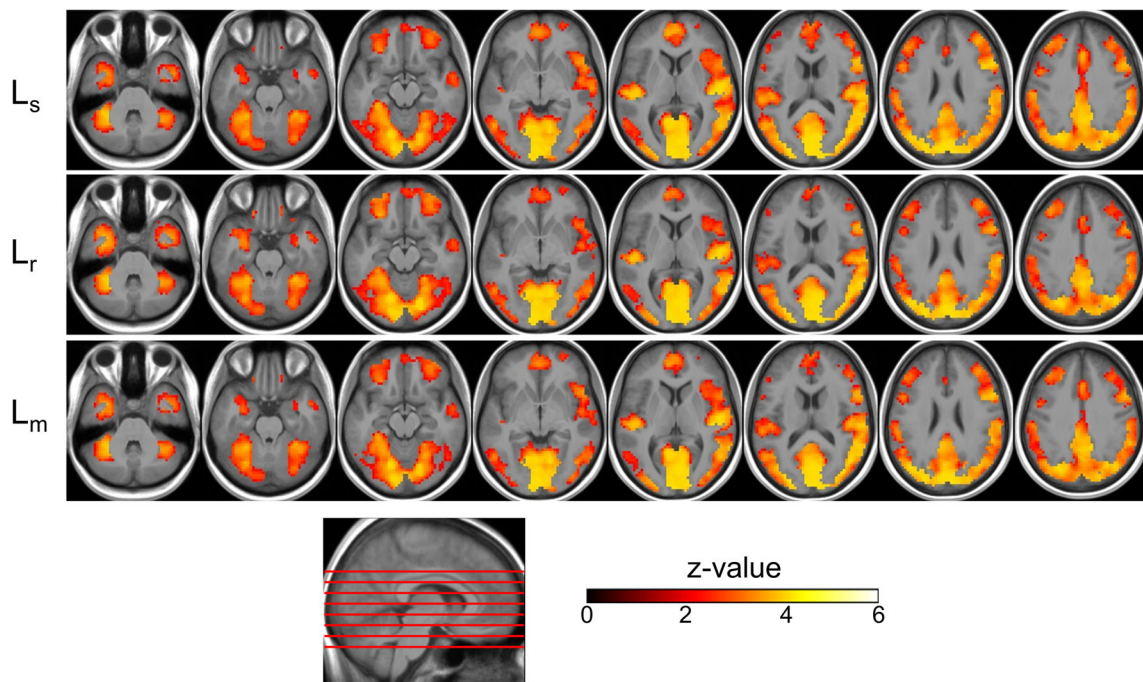
Despite these similarities between the activation maps of the three models with respect to  $L_s$ ,  $L_c$ , and  $L_m$  when contrasted with the

baseline, contrasting them against each other revealed significant differences, as shown in Figure 6.

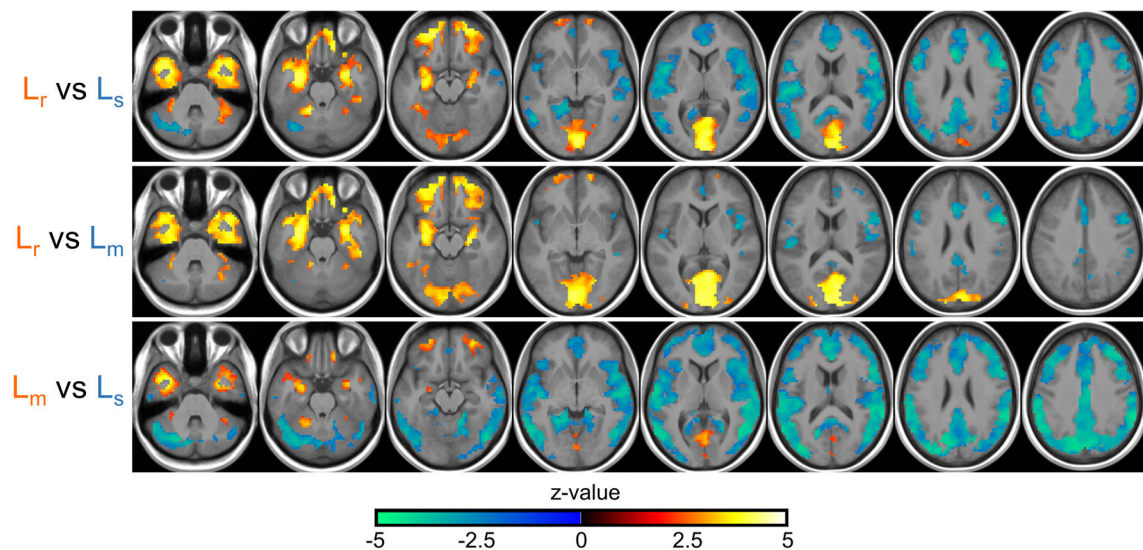
Clusters of voxels characterized by significantly higher prediction accuracy for  $L_r$  as compared with both  $L_s$  and  $L_m$  are found (always bilaterally from here, if not stated otherwise) in the calcarine area, lingual gyrus and cuneus in the occipital lobe as well as in the inferior temporal lobe in both hemispheres, where they span from the anterior part of fusiform gyrus at their caudal end to the amygdala and the temporal pole at their rostral end via the parahippocampal gyrus, hippocampus and inferior temporal gyrus. Moreover, they cover large parts of the orbitofrontal cortex, comprising orbital parts of the superior, middle, and inferior frontal gyrus, as well as the gyrus rectus (Brodmann areas 10, 11, and 47). Further clusters were detected in the cerebellum. Minor differences between the two contrasts,  $L_r-L_s$  and  $L_r-L_m$ , were mostly limited to the size of some of the aforementioned clusters, especially in occipital areas, where they also reached into the edges of superior, middle, and inferior occipital gyrus for  $L_r-L_m$ .

Conversely, activation in a large network of regions encompassing more superior areas of the temporal, frontal, and occipital lobe, as well as parietal areas and the cingulate cortex, was most accurately predicted by  $L_s$ . Specifically, significant clusters for both contrasts against the other models,  $L_s-L_m$  and  $L_s-L_r$ , included the Heschl's gyrus, superior and middle temporal gyrus, and the most rostral part of inferior temporal gyrus. In the frontal lobe, they comprised the precentral gyrus and supplementary motor area, the superior frontal gyrus (the right premotor cortex and most rostral part of BA 10), middle frontal gyrus, the Rolandic operculum, and the opercular and triangular parts of inferior frontal gyrus. In the parietal and occipital lobe, they comprised the postcentral gyrus, superior and inferior parietal gyrus, the precuneus, supramarginal, and angular gyrus. In the occipital lobe, they included the anterior cuneus, the left anterior calcarine area and lingual gyrus, and the superior and middle occipital gyrus. Furthermore, significant clusters covered the anterior, middle, and posterior cingulate gyrus, and the insula. Another cluster was detected in the left cerebellum. In general, significant clusters were noticeably larger for the  $L_s-L_m$  contrast as compared with  $L_s-L_r$ . This was especially true in occipital and cerebellar areas, where they additionally included the right calcarine area and lingual gyrus, much more of middle and inferior occipital gyrus, and several regions in the left and right cerebellum. There were also considerably larger clusters in the temporal lobes, where they extended more inferiorly and covered large parts of middle and inferior temporal gyrus, yet mostly the posterior fractions. Similarly, clusters reached into more inferior areas of the frontal lobe, particularly in middle frontal gyrus and the triangular part of inferior frontal gyrus, and the superior edge of medial orbitofrontal cortex.

Lastly, there were also clusters, albeit smaller in number and extent, in which fMRI activation was more accurately predicted by  $L_m$  than either by  $L_s$  or by  $L_c$ . For the contrast  $L_m-L_s$ , significant clusters were found in the temporal pole and the adjacent, most anterior parts of middle and inferior temporal gyrus, in the fusiform gyrus,



**FIGURE 5** Activation maps: Sensation level, context loudness and mean loudness versus baseline. Second-level z-statistic maps of signed-rank-tests comparing cross-validated  $R^2$ s between the models indicated on the left and the baseline model are thresholded at  $p < .05$ , FDR-corrected, with minimum cluster-size of 10+ voxels, and overlaid onto the group mean structural image. The maps are color-coded by z-values as indicated by the colorbar and show clusters with higher prediction accuracy for the respective models as compared with the baseline. The six axial slices are located at  $z = -35, -25, -15, -5, 5, 15, 25,$  and  $35$  mm in MNI space (from left to right), as illustrated by the red lines on the sagittal slice below. Note that not all the significant clusters that are reported in the text are visible on this slice selection (e.g., superior parietal and frontal areas may be missing).  $L_s$ , sensation level;  $L_m$ , mean loudness;  $L_c$ , context loudness



**FIGURE 6** Activation maps: Comparison of sensation level, context loudness and mean loudness. Second-level z-statistic maps of the differences in cross-validated predicted  $R^2$  between models were thresholded at  $p < .05$ , FDR-corrected, with minimum cluster-size of 10+ voxels. Significant clusters are color-coded according to the model with higher prediction accuracy in the respective clusters, as indicated on the left, and overlaid onto the group mean structural image. The six axial slices are located at the same z-coordinates as in Figure 4.  $L_s$ , sensation level;  $L_m$ , mean loudness;  $L_c$ , context loudness

parahippocampal gyrus, the hippocampus and the amygdala. They were also detected in the orbital parts of superior, middle and inferior frontal gyrus, the lingual gyrus and the cerebellum. By contrast, when

compared with  $L_r$ ,  $L_m$  performed significantly better bilaterally in the central part of Heschl's gyrus, in the left superior temporal gyrus, and in a small cluster in the right middle temporal gyrus. Moreover,

significant clusters included bilaterally the supplemental motor area, the Rolandic operculum, the opercular and triangular parts of inferior frontal gyrus, the inferior parietal gyrus, supramarginal gyrus, and the precuneus, the cingulate cortex (anterior, middle and posterior gyrus) and the insula. The frontal and parietal activation was noticeably more pronounced in the right hemisphere, where it extended from the triangular part of inferior frontal gyrus caudally into middle frontal gyrus and superiorly into precentral gyrus, and from inferior parietal gyrus into the superior parietal and the angular gyrus.

## 4 | DISCUSSION

In this study, we explored neural correlates, as reflected by the BOLD response in functional MRI, of individual loudness perception for time varying sounds. Normal hearing listeners continuously judged their perceived loudness of acoustic stimuli that slowly changed in level over time while auditory fMRI was performed. As anticipated, the presented stimuli elicited contextual effects with respect to the listeners' loudness judgments. Specifically, loudness was generally judged as higher following an intensity up-ramp and lower following a down-ramp, despite similar sound levels at the respective points in time. Fluctuations of fMRI activation in several auditory and non-auditory regions were significantly related to individual sensation levels and loudness judgments. In contrast to our initial hypothesis, activation in the auditory cortex (AC) was not more accurately predicted by loudness estimates relative to sound levels when the latter were corrected for detection thresholds. Instead, our data indicate that neural responses in areas not typically involved in auditory processing most closely and reliably reflected the subjective loudness judgments. Below, we discuss these findings in detail.

### 4.1 | Context effects on loudness judgments

The high correlation between individual loudness judgments and sensation levels confirms that participants could follow the temporal evolution of level fairly accurately and were able to report their perception by means of the employed response device and categorical scale. A similar method was used by Kuwano and Namba (1985) to study continuous loudness judgments in response to road traffic noise with slowly fluctuating levels. In several respects, their data align well with ours. For instance, their highest coefficients of correlation between instantaneous judgments and the level time course were close to 0.9, with response delays varying from 0.4 to 2.3 s across listeners. In other respects, however, our findings differ considerably. Specifically, Kuwano and Namba found that instantaneous loudness judgments correlated best with the average across sound levels over a period of 2.5 s preceding them. If this was equally true in the present study, the same level should be judged higher when levels were previously falling and lower when they were rising. Conversely, we observed the opposite result. At least with respect to the medium levels (L2 and L3), loudness was judged higher following an up-ramp

and lower "in the context of" a down-ramp relative to the average judgment across conditions. It is likely that judgments at the more extreme levels, L1 and L4, were similarly affected by the direction of level change. Yet, evidence for this assumption remains elusive, since both stationary levels were always presented in the same context—L4 at the end of an up-ramp and L1 at the end of a down-ramp.

The discrepancy between both studies could be related to the intensity profiles of the acoustic stimuli, which increased and decreased monotonically in level at a constant rate and ramp duration in the present experiment, whereas they fluctuated more irregularly in the recordings used by Kuwano and Namba. On top of that, the stimuli differed with respect to their spectral characteristics. Previous investigations have demonstrated that ramp direction significantly affects judgments of loudness *change* for complex tones, as used here, but not for white noise (Neuhoff, 1998, 2001), which is more akin to the broadband traffic noise presented by Kuwano and Namba. This observation has been discussed from an evolutionary perspective in the context of looming and receding auditory motion (Neuhoff, 1998, 2001). Accordingly, harmonic sounds are somewhat special since they are commonly produced by single biological and potentially relevant sources in our environment. Continuous broadband noise, on the other hand, is more likely the result of multiple sources or dispersed phenomena such as wind or rain (Neuhoff, 1998). A perceptual bias for rising intensity harmonic sound may provide an organism with an advantage in preparing for the arrival of an object or potential predator, while loudness decruitment with decreasing levels may signal decreasing environmental importance due its departure (Neuhoff, 2001).

Similarly, our results only partially conform to previous investigations studying ramp-specific effects on judgments of loudness. In line with previous studies that asked participants to judge loudness repeatedly or continuously throughout stimuli with continuous level changes (e.g., Canévet et al., 2003; Canévet & Scharf, 1990; Olsen et al., 2014; Schlauch, 1992; Susini et al., 2007; Teghtsoonian et al., 2000), loudness change was on average greater in response to down-ramps relative to up-ramps. Yet, our data also revealed an interaction between the ramp direction and the intensity region (see Figure 4, panel b): While down-ramps elicited greater loudness change at higher levels, the asymmetry was reversed at the lowest levels. This finding is rather surprising in the light of previous reports, which instead suggest growing overestimation of up-ramps with increasing intensity (Neuhoff, 1998, 2001; Olsen et al., 2010) and more rapid loudness decruitment at decreasing levels below  $\sim 40$  dB SPL (Canévet & Scharf, 1990).

However, the comparison of our data to these findings is not straightforward. Aside from distinct methods to assess loudness, the acoustic environment in the MRI scanner was vastly different from the conditions in previous experiments, which were typically conducted in a shielded sound booth or even anechoic rooms. As a consequence, detection thresholds were markedly elevated in the present study, which ultimately led to a rather small range of levels that could be presented. In fact, the range covered by ramps in this study was often much smaller than the typical range used in previous studies,

which ranged from 15 to 45 dB across the studies cited above. Another probably important distinction is that previous studies primarily presented single ramps or pairs of opposing ramps (in Susini et al., 2007; Olsen et al., 2010, 2014) separated by silent intervals, as opposed to the sequences of several interconnected ramps in our experiment. The impact of these factors on the effects of intensity dynamics on judgments of loudness and perceived change is a topic probably worth exploring in future studies, but out of the scope of the present work.

## 4.2 | Common activation in relation to level and loudness

Significant activation in relation to sensation levels and loudness was found in a large network of regions throughout the whole brain. These included primary and parts of secondary AC in both hemispheres, which is consistent with previous neuroimaging studies that have repeatedly demonstrated intensity dependent activation in these areas for different kinds of synthetic, stationary sounds (e.g., Behler & Uppenkamp, 2016; Hall et al., 2001; Langers et al., 2007; Sigalovsky & Melcher, 2006; Woods et al., 2009), and more natural time varying sounds (piano music in Lehne et al., 2013; and speech in Thwaites et al., 2015, 2016). Conversely, our finding of intensity-related activation across various regions that are not typically associated with auditory processing may be surprising at first glance. It however appears reasonable when considering that participants were constantly involved in a rather complex task, which obviously involved early auditory processing of the acoustic stimuli (mediated by AC), but also responding via button presses (mediated by motor cortex), and the processing of visual information on the response scale (mediated by visual cortex). Moreover, it is conceivable that the continuous loudness judgment task demanded a plethora of higher cognitive processes required for allocating and sustaining attention toward the stimuli, their evaluation in terms of the available response alternatives, monitoring of current and previous responses, adaptive selection of appropriate responses, and inhibition of inappropriate ones to changing conditions. Most of these “executive functions” have traditionally been linked to the prefrontal cortex (Alvarez & Emory, 2006), but there is evidence for an involvement of other non-frontal structures such as the cingulate cortex (Allman, Hakeem, Erwin, Nimchinsky, & Hof, 2001), the insula (Clark et al., 2008), and the cerebellum (Noroozian, 2014), which were significantly activated here, as well. Although it is convenient to assume that activity in “nonauditory” regions was more related to other aspects of the task as opposed to the acoustic stimulation itself, we cannot rule out a significant contribution of the latter. As a matter of fact, Neuner et al. (2014) found that increasing the intensity to high sound pressure levels activated many additional regions other than auditory cortex in response to short tone bursts, even in the absence of a stimulus-related task. Several of these regions were also related to stimulus intensity in the present study, including the angular gyrus, the frontal operculum, pre- and postcentral gyrus, middle temporal gyrus and the temporal pole,

the insula, OFC, and visual cortex. Lastly, the OFC (Brodmann area 47, in particular) could be more generally involved in processing (acoustic) stimuli that evolve over time, as speculated by Levitin and Menon (2003) based on studies of language processing and their own results. In their auditory fMRI study, they found increased activation in OFC and the anterior insula when participants listened to classical music as compared with “scrambled” versions of that same music, in which its temporal structure was disrupted.

## 4.3 | The neural representation of contextual loudness judgments

The direct comparison of individual sensation levels, context loudness and mean loudness in terms of predicting fMRI activation revealed partly surprising findings, as well. Contrary to our first initial hypothesis, sensation level yielded significantly better predictions than both loudness variables in primary AC and the surrounding areas (among other regions). This result seems to be at odds with previous studies that have investigated neural correlates of loudness by means of fMRI (e.g., Behler & Uppenkamp, 2016; Hall et al., 2001; Langers et al., 2007; Röhl & Uppenkamp, 2012) and EEG/MEG (Thwaites et al., 2015, 2016), which suggest that activation in AC is more closely related to loudness rather than sound level. In the present study, variables representing the individual loudness of participants instead performed significantly better than level mainly in the OFC, the inferior medial temporal lobe including the parahippocampal gyrus, fusiform gyrus, hippocampus and amygdala, the temporal pole, and the visual cortex. More precisely, activation in all of these regions was best predicted by (i.e., most closely related to) context loudness, followed by mean loudness, with sensation level showing the comparatively weakest performance.

However, our results need not necessarily be in conflict with the existing literature. Instead, we argue that the discrepancies described above may reflect important distinctions between studies. Two major differences between the present and previous investigations pertain to the task performed by the participants during the neuroimaging procedure and to the specific loudness estimates that were used in statistical analyses. In the present experiment, listeners continuously judged the loudness of acoustic stimuli while activation data were collected simultaneously, and individual loudness estimates were derived from these judgments. This is vastly different from previous neuroimaging studies that typically employed simple tasks to ensure attention of their participants toward the acoustic stimuli, but which were otherwise not relevant to the main analysis. Loudness estimates, on the other hand, were either derived from functions fitted to individual data obtained in a different task and environment (e.g., Behler & Uppenkamp, 2016; Langers et al., 2007; Röhl & Uppenkamp, 2012), or calculated equally for all participants by means of loudness models (e.g., Hall et al., 2001; Thwaites et al., 2015, 2016). Thus, previous estimates did not capture possible context effects on the evaluation of acoustic stimuli and might in fact have been more similar to the sensation level as opposed to context loudness in this study. As



elaborated below, we argue based on the identified brain regions associated with contextual judgments, that these judgments may more reflect the outcome of a cognitive decision-making process related to episodic memory rather than the early perceptual processing of acoustic stimuli only.

With respect to visual cortex, systematic changes of activation with context loudness are probably best explained by the visual feedback provided to participants in the judgment task. Changes of loudness were always accompanied by a change of the highlighted response category and, importantly, a systematic shift of the response scale's position relative to the visual focus. Visual cortex is well-known for its retinotopic organization (see however Murray, Boyaci, & Kersten, 2006; Fang, Boyaci, Kersten, & Murray, 2008; and Joo, Boynton, & Murray, 2012 for context- and attention-dependent effects). It is therefore possible that activation in visual areas was closely related to the contextual judgments based on the visual feedback alone.

With respect to the OFC, the position of the response categories is much less likely to play a decisive role. While activity in primate OFC has been shown to be indicative of the subjective decision-making process when two or more visual response alternatives are presented, it appears that information encoding in this region is independent of the spatial location of available options (e.g., Grattan & Glimcher, 2014; Rich & Wallis, 2016). Instead, a large body of evidence suggests that activation in OFC represents the subjective affective (or reward) value of stimuli on a continuous scale in a spatially-independent, and modality-unspecific manner, as supported by significant correlations between BOLD signals and subjective pleasantness ratings of taste, olfactory, flavor, temperature, and face beauty stimuli, as well as monetary reward value (for a review, see: Rolls & Grabenhorst, 2008). In this endeavor, the OFC is thought to closely interact with the amygdala (Holland & Gallagher, 2004). Yet, to a much greater extent than the amygdala, the OFC appears to be critically involved in guiding decision-making based on the expected value or outcome (e.g., Holland & Gallagher, 2004; Wallis, 2007). It is consistently activated when an explicit judgment or choice is required, especially under conditions of insufficient information to determine the appropriate or "right" response, and activation increases with the difficulty of the decision (e.g., Arana et al., 2003; Elliott, Dolan, & Frith, 2000). Finally, lateral OFC, in particular, seems to play a crucial role in the suppression of previously rewarded responses or appealing alternatives and in reversal learning under changing conditions (e.g., Arana et al., 2003; Elliott et al., 2000; Ghods-Sharifi, Haluk, & Floresco, 2008; Hampshire, Chaudhry, Owen, & Roberts, 2012).

Finally, the medial temporal lobe is most renowned for its role in memory encoding and retrieval (Squire & Zola-Morgan, 1991). Specifically, the amygdala has been assigned an essential role in (implicit) emotional learning (e.g., Maren, 1999; Phelps & LeDoux, 2005), whereas the hippocampus and parahippocampal gyrus are recognized as central structures with respect to declarative memory. Neuroimaging studies suggest that the hippocampus combines item-related features and concepts input from the perirhinal cortex (located in the anterior parahippocampal gyrus) with contextual information represented in the parahippocampal cortex (located in the posterior

parahippocampal gyrus and medial fusiform gyrus) to support episodic memory encoding and recollection (e.g., Davachi, 2006; Diana, Yonelinas, & Ranganath, 2007; Eichenbaum, Yonelinas, & Ranganath, 2007). Anatomically, the OFC is densely interconnected with the medial temporal lobe (Carmichael & Price, 1995). It is therefore not surprising that it has been implicated in memory formation as well (e.g., Frey & Petrides, 2002; Petrides, 2007; Ramus, Davis, Donahue, Discenza, & Waite, 2007). Moreover, computing the expected reward value of available choices likely profits from access to information about previous actions and their respective outcomes.

But how does activation in the aforementioned regions, especially the OFC, relate to the continuous loudness judgments of participants in the present study, which were obviously neither rewarded nor punished? A possible explanation is that participants might have experienced a high degree of uncertainty with respect to the most appropriate response at every instance in time. Consequently, they may have resorted to devising strategies based on previous choices and the context in which they were made. This might have led to the recruitment of brain regions involved in episodic memory and the computation of more abstract values that guided their decisions, in addition to the currently available sensory information.

This hypothesis supports the interpretation issued by Schlauch (1992) that the phenomenon of loudness decreitment for sounds with continuous decreases of acoustic intensity may have a substantial cognitive contribution (despite having a sensory origin). His psychoacoustic experiments revealed substantially greater decreitment when participants constantly focused their attention on the acoustic stimuli as compared with a diverted attention condition. The idea that the OFC, which is apparently modality-unspecific, could be the main driver of the observed context effects is consistent with the findings of Teghtsoonian et al. (2000). They reported a phenomenon analogous to loudness decreitment, but in the visual domain, when they asked observers to judge the apparent size of a changing disk. Still, even within the hearing modality, OFC recruitment does not seem to be specific to loudness judgments. In the auditory fMRI study by Lehne et al. (2013), for example, listeners were asked to continuously rate their subjective experience of musical tension for piano pieces. In their analysis, Lehne and colleagues also included estimates of loudness for the music stimuli, as calculated by means of a loudness model. They found neural correlates of loudness exclusively in AC, and correlates of subjective tension ratings exclusively in OFC (BA 47). This is in line with the idea described above that activation in OFC may be more related to the evaluation of stimuli in the context of a given task rather than to general features in the perception of sound. Whether this hypothesis can account for the present findings certainly requires further empirical investigation.

Finally, one might have expected that at least the mean loudness estimates should have predicted activation in the AC better than sensation levels. After all, this variable was included to be more directly comparable to loudness estimates used in previous neuroimaging studies. We would certainly agree with this assumption, if the mean loudness time course were in fact free of contextual effects on judgments, as initially intended. Unfortunately, this was probably not the

case. As stated earlier, the two extreme stationary levels were always presented in the same contexts, and possible effects would therefore remain even in the averaged values. This concern impedes a straightforward comparison between the present and previous findings.

## 5 | CONCLUSION

In the present study, sounds with continuous increases and decreases of intensity produced robust context effects with respect to continuous judgments of loudness while auditory fMRI was performed. Specifically, sounds were judged comparatively softer following intensity down-ramps and louder in the context of up-ramps. Neural correlates of these context effects were mainly found in bilateral orbitofrontal cortex, medial temporal lobe, and in the visual cortex. While activation in the latter was likely more related to the visual feedback provided in the task, the identified activation in regions associated with decision-making and memory points toward cognitive processes involved in the generation of the observed context effects. This conforms to the interpretation issued by other researchers that these effects may be of sensory origin, but entail a strong cognitive contribution. In fact, they may to a large degree be contingent on the task itself, which has important implications for future investigations on this topic. Still, further research is needed to substantiate these findings and to disentangle the contributions of cognitive and perceptual aspects to loudness judgments for sounds with dynamic intensity.

## CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

## DATA AVAILABILITY STATEMENT

All data used in the statistical analyses of this work is available via the Harvard Dataverse (<https://doi.org/10.7910/DVN/5BBDU5>).

## ORCID

Oliver Behler  <https://orcid.org/0000-0001-9263-2045>

## REFERENCES

- Allman, J. M., Hakeem, A., Erwin, J. M., Nimchinsky, E., & Hof, P. (2001). The anterior cingulate cortex: The evolution of an interface between emotion and cognition. *Annals of the New York Academy of Sciences*, 935(1), 107–117. <https://doi.org/10.1111/j.1749-6632.2001.tb03476.x>
- Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review*, 16, 17–42. <https://doi.org/10.1007/s11065-006-9002-x>
- Arana, F. S., Parkinson, J. A., Hinton, E., Holland, A. J., Owen, A. M., & Roberts, A. C. (2003). Dissociable contributions of the human amygdala and orbitofrontal cortex to incentive motivation and goal selection. *Journal of Neuroscience*, 23, 9632–9638. <https://doi.org/10.1523/JNEUROSCI.23-29-09632.2003>
- Bach, D. R., Schächinger, H., Neuhoff, J.G., Esposito, F., Salle, F.D., Lehmann, C., & Seifritz, E. (2008). Rising sound intensity: an intrinsic warning cue activating the amygdala. *Cerebral Cortex*, 18, 145–150.
- Bach, D. R., Neuhoff, J. G., Perrig, W., & Seifritz, E. (2009). Looming sounds as warning signals: The function of motion cues. *International Journal of Psychophysiology*, 74, 28–33. <https://doi.org/10.1016/j.ijpsycho.2009.06.004>
- Behler, O., & Uppenkamp, S. (2016). The representation of level and loudness in the central auditory system for unilateral stimulation. *NeuroImage*, 139, 176–188. <https://doi.org/10.1016/j.NeuroImage2016.06.025>
- Brand, T., & Hohmann, V. (2002). An adaptive procedure for categorical loudness scaling. *The Journal of the Acoustical Society of America*, 112, 1597–1604. <https://doi.org/10.1121/1.1502902>
- Canévet, G., & Scharf, B. (1990). The loudness of sounds that increase and decrease continuously in level. *The Journal of the Acoustical Society of America*, 88, 2136–2142. <https://doi.org/10.1121/1.400110>
- Canévet, G., Teghtsoonian, R., & Teghtsoonian, M. (2003). A comparison of loudness change in signals that continuously rise or fall in amplitude. *Acta Acustica united with Acustica*, 89, 339–345.
- Carmichael, S. T., & Price, J. L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal of Comparative Neurology*, 363, 615–641. <https://doi.org/10.1002/cne.903630408>
- Chalupper, J., & Fastl, H. (2002). Dynamic loudness model (DLM) for normal and hearing-impaired listeners. *Acta Acustica united with Acustica*, 88, 378–386.
- Chambers, J., Bullock, D., Kahana, Y., Kots, A., & Palmer, A. (2007). Developments in active noise control sound systems for magnetic resonance imaging. *Applied Acoustics*, 68, 281–295. <https://doi.org/10.1016/j.apacoust.2005.10.008>
- Clark, L., Bechara, A., Damasio, H., Aitken, M. R. F., Sahakian, B. J., & Robbins, T. W. (2008). Differential effects of insular and ventromedial prefrontal cortex lesions on risky decision-making. *Brain*, 131, 1311–1322. <https://doi.org/10.1093/brain/awn066>
- Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, 16, 693–700. <https://doi.org/10.1016/j.conb.2006.10.012>
- Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences*, 11, 379–386. <https://doi.org/10.1016/j.tics.2007.08.001>
- Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, 30, 123–152. <https://doi.org/10.1146/annurev.neuro.30.051606.094328>
- Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex*, 10, 308–317. <https://doi.org/10.1093/cercor/10.3.308>
- Fang, F., Boyaci, H., Kersten, D., & Murray, S. O. (2008). Attention-dependent representation of a size illusion in human V1. *Current Biology*, 18, 1707–1712. <https://doi.org/10.1016/j.cub.2008.09.025>
- Frey, S., & Petrides, M. (2002). Orbitofrontal cortex and memory formation. *Neuron*, 36, 171–176. [https://doi.org/10.1016/s0896-6273\(02\)00901-7](https://doi.org/10.1016/s0896-6273(02)00901-7)
- Ghods-Sharifi, S., Haluk, D. M., & Floresco, S. B. (2008). Differential effects of inactivation of the orbitofrontal cortex on strategy set-shifting and reversal learning. *Neurobiology of Learning and Memory*, 89, 567–573. <https://doi.org/10.1016/j.nlm.2007.10.007>
- Glasberg, B. R., & Moore, B. C. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50, 331–342.
- Grattan, L. E., & Glimcher, P. W. (2014). Absence of spatial tuning in the orbitofrontal cortex. *PLoS One*, 9, e112750. <https://doi.org/10.1371/journal.pone.0112750>
- Hall, D. A., Haggard, M. P., Summerfield, A. Q., Akeroyd, M. A., Palmer, A. R., & Bowtell, R. W. (2001). Functional magnetic resonance imaging measurements of sound-level encoding in the absence of background scanner noise. *The Journal of the Acoustical Society of America*, 109, 1559–1570. <https://doi.org/10.1121/1.1345697>



- Hampshire, A., Chaudhry, A. M., Owen, A. M., & Roberts, A. C. (2012). Dissociable roles for lateral orbitofrontal cortex and lateral prefrontal cortex during preference driven reversal learning. *NeuroImage*, *59*, 4102–4112. <https://doi.org/10.1016/j.neuroimage.2011.10.072>
- Holland, P. C., & Gallagher, M. (2004). Amygdala-frontal interactions and reward expectancy. *Current Opinion in Neurobiology*, *14*, 148–155. <https://doi.org/10.1016/j.conb.2004.03.007>
- Joo, S. J., Boynton, G. M., & Murray, S. O. (2012). Long-range, pattern-dependent contextual effects in early human visual cortex. *Current Biology*, *22*, 781–786. <https://doi.org/10.1016/j.cub.2012.02.067>
- Kuwano, S., & Namba, S. (1985). Continuous judgment of level-fluctuating sounds and the relationship between overall loudness and instantaneous loudness. *Psychological Research*, *47*, 27–37. <https://doi.org/10.1007/BF00309216>
- Langers, D. R., van Dijk, P., Schoenmaker, E. S., & Backes, W. H. (2007). fMRI activation in relation to sound intensity and loudness. *NeuroImage*, *35*, 709–718. <https://doi.org/10.1016/j.NeuroImage2006.12.013>
- Lehne, M., Rohrmeier, M., & Koelsch, S. (2013). Tension-related activity in the orbitofrontal cortex and amygdala: An fMRI study with music. *Social Cognitive and Affective Neuroscience*, *9*, 1515–1523. <https://doi.org/10.1093/scan/nst141>
- Levitin, D. J., & Menon, V. (2003). Musical structure is processed in “language” areas of the brain: A possible role for Brodmann area 47 in temporal coherence. *NeuroImage*, *20*, 2142–2152. <https://doi.org/10.1016/j.NeuroImage.2003.08.016>
- Maren, S. (1999). Long-term potentiation in the amygdala: A mechanism for emotional learning and memory. *Trends in Neurosciences*, *22*, 561–567. [https://doi.org/10.1016/S0166-2236\(99\)01465-4](https://doi.org/10.1016/S0166-2236(99)01465-4)
- Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience*, *9*, 429–434. <https://doi.org/10.1038/nn1641>
- Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature*, *395*(6698), 123–124. <http://doi.org/10.1038/25862>
- Neuhoff, J. G. (1999). Perception of changes in loudness: Reply. *Nature*, *398*(6729), 673–674. <https://doi.org/10.1038/19446>
- Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology*, *13*, 87–110. [https://doi.org/10.1207/S15326969ECO1302\\_2](https://doi.org/10.1207/S15326969ECO1302_2)
- Neuner, I., Kawohl, W., Arrubla, J., Warbrick, T., Hitz, K., Wyss, C., ... Shah, N. J. (2014). Cortical response variation with different sound pressure levels: A combined event-related potentials and fMRI study. *PLoS One*, *9*, e109216. <https://doi.org/10.1371/journal.pone.0109216>
- Noroozian, M. (2014). The role of the cerebellum in cognition: Beyond coordination in the central nervous system. *Neurologic Clinics*, *32*, 1081–1104. <https://doi.org/10.1016/j.ncl.2014.07.005>
- Oberfeld, D., Heeren, W., Rannies, J., & Verhey, J. (2012). Spectro-temporal weighting of loudness. *PLoS One*, *7*, e50184. <https://doi.org/10.1371/journal.pone.0050184>
- Oberfeld, D., & Plank, T. (2011). The temporal weighting of loudness: Effects of the level profile. *Attention, Perception, & Psychophysics*, *73*, 189–208. <https://doi.org/10.3758/s13414-010-0011-8>
- Olsen, K. N. (2014). Intensity dynamics and loudness change: A review of methods and perceptual processes. *Acoustics Australia*, *42*(3), 159–165.
- Olsen, K. N., & Stevens, C. J. (2012). Forward masking of dynamic acoustic intensity: Effects of intensity region and end-level. *Perception*, *41*, 594–605. <http://doi.org/10.1068/p7128>
- Olsen, K. N., Stevens, C. J., Dean, R. T., & Bailes, F. (2014). Continuous loudness response to acoustic intensity dynamics in melodies: Effects of melodic contour, tempo, and tonality. *Acta Psychologica*, *149*, 117–128. <https://doi.org/10.1016/j.actpsy.2014.03.007>
- Olsen, K. N., Stevens, C. J., & Tardieu, J. (2010). Loudness change in response to dynamic acoustic intensity. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1631–1644. <https://doi.org/10.1037/a0018389>
- Petrides, M. (2007). The orbitofrontal cortex: Novelty, deviation from expectation, and memory. *Annals of the New York Academy of Sciences*, *1121*, 33–53. <https://doi.org/10.1196/annals.1401.035>
- Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron*, *48*(2), 175–187. <https://doi.org/10.1016/j.neuron.2005.09.025>
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*, 2142–2154. <https://doi.org/10.1016/j.NeuroImage2011.10.018>
- Racine, J. (2000). Consistent cross-validated model-selection for dependent data: Hv-block cross-validation. *Journal of Econometrics*, *99*, 39–61. [https://doi.org/10.1016/S0304-4076\(00\)00030-0](https://doi.org/10.1016/S0304-4076(00)00030-0)
- Ramus, S. J., Davis, J. B., Donahue, R. J., Discenza, C. B., & Waite, A. A. (2007). Interactions between the orbitofrontal cortex and the hippocampal memory system during the storage of long-term memory. *Annals of the New York Academy of Sciences*, *1121*, 216–231. <https://doi.org/10.1196/annals.1401.038>
- Rannies, J., Verhey, J. L., & Fastl, H. (2010). Comparison of loudness models for time-varying sounds. *Acta Acustica united with Acustica*, *96*, 383–396. <https://doi.org/10.3813/AAA.918287>
- Rich, E. L., & Wallis, J. D. (2016). Decoding subjective decisions from orbitofrontal cortex. *Nature Neuroscience*, *19*, 973–980. <https://doi.org/10.1038/nn.4320>
- Röhl, M., & Uppenkamp, S. (2012). Neural coding of sound intensity and loudness in the human auditory system. *Journal of the Association for Research in Otolaryngology*, *13*, 369–379. <https://doi.org/10.1007/s10162-012-0315-6>
- Rolls, E. T., & Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: From affect to decision-making. *Progress in Neurobiology*, *86*, 216–244. <https://doi.org/10.1016/j.pneurobio.2008.09.001>
- Schlauch, R. S. (1992). A cognitive influence on the loudness of tones that change continuously in level. *The Journal of the Acoustical Society of America*, *92*, 758–765. <https://doi.org/10.1121/1.403999>
- Seifritz, E., Neuhoff, J. G., Bilecen, D., Scheffler, K., Mustovic, H., Schächinger, H., ... Di Salle, F. (2002). Neural processing of auditory looming in the human brain. *Current Biology*, *12*, 2147–2151. [https://doi.org/10.1016/S0960-9822\(02\)01356-8](https://doi.org/10.1016/S0960-9822(02)01356-8)
- Sigalovsky, I. S., & Melcher, J. R. (2006). Effects of sound level on fMRI activation in human brainstem, thalamic and cortical centers. *Hearing Research*, *215*, 67–76. <https://doi.org/10.1016/j.heares.2006.03.002>
- Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*, 1380–1386. <https://doi.org/10.1126/science.1896849>
- Susini, P., McAdams, S., & Smith, B. K. (2007). Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings. *Acta Acustica united with Acustica*, *93*, 623–631.
- Teghtsoonian, R., Teghtsoonian, M., & Canévet, G. (2000). The perception of waning signals: Decruitment in loudness and perceived size. *Perception & Psychophysics*, *62*, 637–646. <http://doi.org/10.3758/BF03212115>
- Thwaites, A., Glasberg, B. R., Nimmo-Smith, I., Marslen-Wilson, W. D., & Moore, B. C. (2016). Representation of instantaneous and short-term loudness in the human cortex. *Frontiers in Neuroscience*, *10*, 183. <https://doi.org/10.3389/fnins.2016.00183>
- Thwaites, A., Nimmo-Smith, I., Fonteneau, E., Patterson, R. D., Buttery, P., & Marslen-Wilson, W. D. (2015). Tracking cortical entrainment in neural activity: Auditory processes in human temporal cortex. *Frontiers in Computational Neuroscience*, *9*, 5. <https://doi.org/10.3389/fncom.2015.00005>
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*, 273–289. <https://doi.org/10.1006/nimg.2001.0978>

- Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annual Review of Neuroscience*, 30, 31–56. <https://doi.org/10.1146/annurev.neuro.30.051606.094334>
- Woods, D. L., Stecker, G. C., Rinne, T., Herron, T. J., Cate, A. D., Yund, E. W., ... Kang, X. (2009). Functional maps of human auditory cortex: Effects of acoustic features and attention. *PLoS One*, 4, e5183. <https://doi.org/10.1371/journal.pone.0005183>

**How to cite this article:** Behler O, Uppenkamp S. Contextual effects on loudness judgments for sounds with continuous changes of intensity are reflected in nonauditory areas. *Hum Brain Mapp*. 2021;42:1742–1757. <https://doi.org/10.1002/hbm.25325>