ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR | WILEY

Cross-modal predictive processing depends on context rather than local contingencies

Tjerk T. Dercksen[1,2,3] | Maria V. Stuckenberg[1,4] (iD) | Erich Schröger[1] | Nicole Wetzel[2,3] | Andreas Widmann[1,2]

[1]Institute of Psychology, Leipzig University, Leipzig, Germany

[2]Leibniz Institute for Neurobiology, Magdeburg, Germany

[3]Center for Behavioral Brain Sciences, Magdeburg, Germany

[4]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Correspondence
Maria V. Stuckenberg, Institute of Psychology, Leipzig University, Neumarkt 9-19, D-04109 Leipzig, Germany.
Email: maria_viktoria.stuckenberg@uni-leipzig.de

Funding information
Leibniz-Gemeinschaft, Grant/Award Number: P58/2017; European Regional Development Fund, Grant/Award Number: ZS/2016/04/78120

Abstract

Visual symbols or events may provide predictive information on to-be-expected sound events. When the perceived sound does not confirm the visual prediction, the incongruency response (IR), a prediction error signal of the event-related brain potentials, is elicited. It is unclear whether predictions are derived from lower-level local contingencies (e.g., recent events or repetitions) or from higher-level global rules applied top-down. In a recent study, sound pitch was predicted by a preceding note symbol. IR elicitation was confined to the condition where one of two sounds was presented more frequently and was not present with equal probability of both sounds. These findings suggest that local repetitions support predictive cross-modal processing. On the other hand, IR has also been observed with equal stimulus probabilities, where visual patterns predicted the upcoming sound sequence. This suggests the application of global rules. Here, we investigated the influence of stimulus repetition on the elicitation of the IR by presenting identical trial trains of a particular visual note symbol cueing a particular sound resulting either in a congruent or an incongruent pair. Trains of four different lengths: 1, 2, 4, or 7 were presented. The IR was observed already after a single presentation of a congruent visual-cue-sound combination and did not change in amplitude as trial train length increased. We conclude that higher-level associations applied in a top-down manner are involved in elicitation of the prediction error signal reflected by the IR, independent from local contingencies.

KEYWORDS

auditory processing, cross-modal processing, ERPs, error processing, neuropsychology, priming

1 | INTRODUCTION

Visual information may influence the processing of auditory information, as illustrated by phenomena like the ventriloquist illusion (e.g., Alais & Burr, 2004), the McGurk effect (McGurk & MacDonald, 1976), or cross-modal spatial attention effects (e.g., Eimer & Driver, 2001). One way in which visual information can influence the processing of auditory information

---

Tjerk T. Dercksen and Maria V. Stuckenberg contributed equally to this work as co-first authors.

PSYCHOPHYSIOLOGY SPR

is to exploit contingencies between succeeding visual and auditory information so that the occurrence of a visual stimulus enables the prediction of a forthcoming sound. For instance, the sighting of an approaching dog might induce the expectation of a barking sound, and after seeing a flash in the sky the sound of thunder would be expected. The existence of such visually induced auditory predictions is often probed in experiments by the presentation of sounds being incongruent to the prediction, which results in the elicitation of prediction error signals in the event-related brain potential (ERP). Prediction error signals are a well-explained phenomenon in the framework of predictive coding. In this framework, predictions about sensory input—which are generated based on prior experience—are sent top-down in a cortical hierarchy. These predictions are compared to actual sensory input at the level of sensory information processing. Where predictions do not meet the expected input, a prediction error signal is fed forward bottom-up the hierarchy, informing higher levels and adjusting generative models to better predict input in the future (Friston, 2005). When the visual modality is predictive of upcoming auditory input, the visual modality is thought to generate predictions regarding the incoming auditory stream (e.g., Pieszek et al., 2013; Widmann et al., 2004). In line with this hypothesis, whenever the auditory input does not match the prediction generated by the visual modality, an auditory prediction error should be elicited.

The study of Widmann and colleagues (2004) revealed such a prediction error. They presented simple score-like patterns of rectangles above or below a fixation-cross that were highly predictive (probability = 0.9) of the occurrence of high- or low-pitched tones in a corresponding auditory pattern. When the tone's pitch was not congruent with the visual-based prediction the incongruency response (IR) was elicited: an enhanced, fronto-laterally distributed ERP component in the latency range of 105–130 ms after sound onset. However, to date it is unclear whether the IR is the result of a low-level, local association, or of higher-level associations. Wacongne et al. (2011) illustrate this distinction by differentiating between local and global violations. They presented series of five tones, whereby in one condition the last tone was always a different tone. On a local level, this tone can be considered a deviant, but in the global context, the repeating different last tone is highly predictable. Interestingly, Wacongne et al. (2011) show that while the earlier MMN is elicited even in response to the predictable local deviants, the later P3b is only elicited when a deviant tone violates the global context (see also Bekinschtein et al., 2009; Marti et al., 2014). Apparently, faster, presumably lower-level responses are sensitive to deviance on a local level, whereas slower, higher levels reflect violation of global rules.

In line with this are results from repetition suppression studies. For instance, Haenschel et al. (2005) use a roving paradigm to present uninterrupted trial trains (of 2, 6, and 36 trials) of the same sound, where the first trial functioned as

the deviant relative to the preceding trial train and the last as the standard sound. They show a suppression of the auditory N1 with increasing amount of standard repetitions. Predictive coding considers such repetition suppression effects of the auditory N1 to be an effect of increasingly successful prediction of the upcoming stimulus (Auksztulewicz & Friston, 2016; Winkler et al., 2009). This can be considered in line with the conclusions of Wacongne et al. (2011), as early components (N1) are sensitive to the local repetition of a stimulus.

In addition, several other studies (Bekinschtein et al., 2009; Costa-Faidella et al., 2011; El Karoui et al., 2015; Nourski et al., 2018) demonstrate that early components are sensitive to local repetition. In the current study, we are interested whether these mechanisms are also at play in a cross-modal setup.

In the context of these findings, the IR can also be regarded as an early response, which would suggest the IR to be driven by local effects. One study that might support this was conducted by Stuckenberg et al. (2019). They presented an eighth note symbol either above or below a fixation cross which was succeeded by the presentation of either a high- or low-pitched sound. The probability of congruent (high-high and low-low) versus incongruent (high-low and low-high) visual-cue-sound combinations was predefined such that overall, the congruent combinations predominated (90% vs. 10% of trials). Apart from this, the presentation probability of the two visual-cue-sound combinations (high and low) was manipulated. They presented the high- and low-pitched sounds (preceded by a visual cue) either with the same probability (50% vs. 50% of trials), or one sound was presented more frequently than the other (83% vs. 17% of trials). In the latter condition, the probability of, for instance, the high visual-cue-sound combination was increased relative to the low visual-cue-sound combination, leading to an overall presentation frequency of 83% high versus 17% low sounds. Stuckenberg et al. (2019) only observed an IR in this specific condition, when one visual-cue-sound combination was presented more frequently than the other (83/17 condition), whereas no IR was observed when high and low visual-cue-sound combinations were presented with equal probability (50/50 condition). They conclude that the increased global probability of one visual-cue-sound combination is a prerequisite for the formation of a strong bimodal association. It has been shown that there is a cross-modal correspondence between high-pitched sounds and small, bright objects that are located high up in space (Spence, 2011). But the 50/50 condition of Stuckenberg et al. (2019) revealed that this correspondence alone is not sufficient to trigger visual-based predictions. However, the change in global probability (from 50/50 to 83/17) is inherently accompanied by a change in local repetitions of one pair, as increasing the frequency of one visual-cue-sound combination over the other affects the local probability of that combination to occur. This in turn leads to longer uninterrupted trial trains of the dominant

congruent combination, possibly leading to a temporary, low-level association between the visual and auditory stimulus. If the IR is the result of such a low-level association, the increased occurrence of local repetitions might explain why an IR is only observed in the high-probability (83/17) condition. However, Widmann et al. (2004) observed an IR with equal distribution of high and low sounds, suggesting that visual–auditory associations are the result of learned higher-level associations that are predicted as soon as the visual stimulus is presented (i.e., without the need for local repetition).

Given these contradicting findings (Stuckenberg et al., 2019 vs. Widmann et al., 2004) and the literature on short-latency components (Wacongne et al., 2011), the present study aims to better understand the role of local repetition on the elicitation of the IR. A similar trial-by-trial design originally used by Pieszek et al. (2013) and later by Stuckenberg et al. (2019) was employed to study the effect of local repetition on the IR (Figure 1). Two different visual-cue-sound combinations were presented (high and low), with equal presentation probability (50% vs. 50%), whereby in 15% of the trials the visual-cue-sound combination was incongruent. To study the effect of local repetition, trial trains with different lengths of identical congruent trials were concluded by an incongruent trial with a misleading visual cue but the same pitch sound as the preceding congruent trial train. Therefore, any changes in the incongruent auditory ERP could be attributed to the preceding visual cue. The conditions tested were 1, 2, 4, and 7 consecutive identical congruent trials before an incongruent trial.

The current design has an equal global probability between high and low visual-cue-sound combinations. In Stuckenberg et al. (2019), this condition did not result in the elicitation of an IR. Therefore, if an IR is elicited with increasing local repetition, this would show that the IR is driven by local effects. Furthermore, if the IR is influenced by local repetition, amplitude differences would be expected as trial trains get longer.

Apart from the IR, subsequent N2 and P3 responses are elicited in the processing of incongruent stimuli. These higher-order prediction error-related processes seem to be related to deviance detection and attention reorienting (Folstein & Van Petten, 2008). Interestingly, the N2 and P3 responses were elicited in both conditions (50/50 and 83/17) of Stuckenberg et al. (2019), congruent with the notion that these higher-order processes are driven by global rule violations. For this reason, the N2-P3 response is expected for all conditions in the current experiment.

2 | METHOD

2.1 | Participants

EEG and behavioral data of 21 participants were acquired. One participant was excluded because of taking medication that affected the central nervous system. The remaining 20 participants (11 female; age range: 19–33; mean age: 23.45 ($SD = 4.61$); 19 right handed) all reported normal hearing, normal or corrected to normal vision, and not taking any medication affecting the central nervous system. All participants gave written consent in accordance with the declaration of Helsinki and received either credit points or modest financial compensation for their participation. The project was approved by the local medical ethical committee.

2.2 | Apparatus and stimuli

Before starting the experiment, all participants completed a general questionnaire and a shortened German version of the Edinburgh handedness inventory (Oldfield, 1971). The general questionnaire contained personal information such as age, gender, occupation, musical skills, health status (hearing and visual ability), and information about intake of caffeine, nicotine, medication, and alcohol prior to the experiment. The main experimental setup was identical to Stuckenberg et al. (2019). Participants were seated in an acoustically attenuated and electrically shielded chamber, while EEG was continuously recorded. The experiment was programed using Psychophysics Toolbox (version 3; Brainard, 1997) and presented on a Linux-based system using GNU Octave (version 4.0.0). Participants sat at a distance of 60 cm from a screen, looking at a white fixation cross on a black background ($0.3° × 0.3°$ of visual angle) that was present throughout the experiment. Starting a trial, a white eighth note symbol ($0.7° × 0.9°$ of visual angle) was presented for 100 ms that was positioned either above or below the fixation cross. The visual stimulus was followed by an auditory stimulus (tone with either 440 Hz or 352 Hz frequency), which was presented 600 ms after the onset of the visual stimulus. The auditory stimulus was presented for 100 ms (including 5 ms rise and fall times respectively; with an intensity of 73 dB SPL) via loudspeakers (Bose Companion 2 Series II, Bose Corporation, Framingham, MA) that were positioned left and right of the screen. Two visual (above vs. below fixation cross) and two auditory (high vs. low pitch) stimuli resulted in four combinations from which two were congruent (above-high and below-low) and two were incongruent pairs (above-low and below-high). We investigated the effect of preceding identical congruent trial repetitions (all presenting the same pitch sound) on the IR by contrasting conditions in which the number of congruent trials directly preceding an incongruent trial were manipulated. Within a trial train, the congruent pairs were concluded with an incongruent pair that had a different visual stimulus as the preceding trials but the same auditory stimulus, resulting in a violation of the visual-based prediction. The four conditions consisted of 1, 2, 4, and 7 congruent trials preceding an incongruent trial (i.e., n1, n2, n4, n7). To
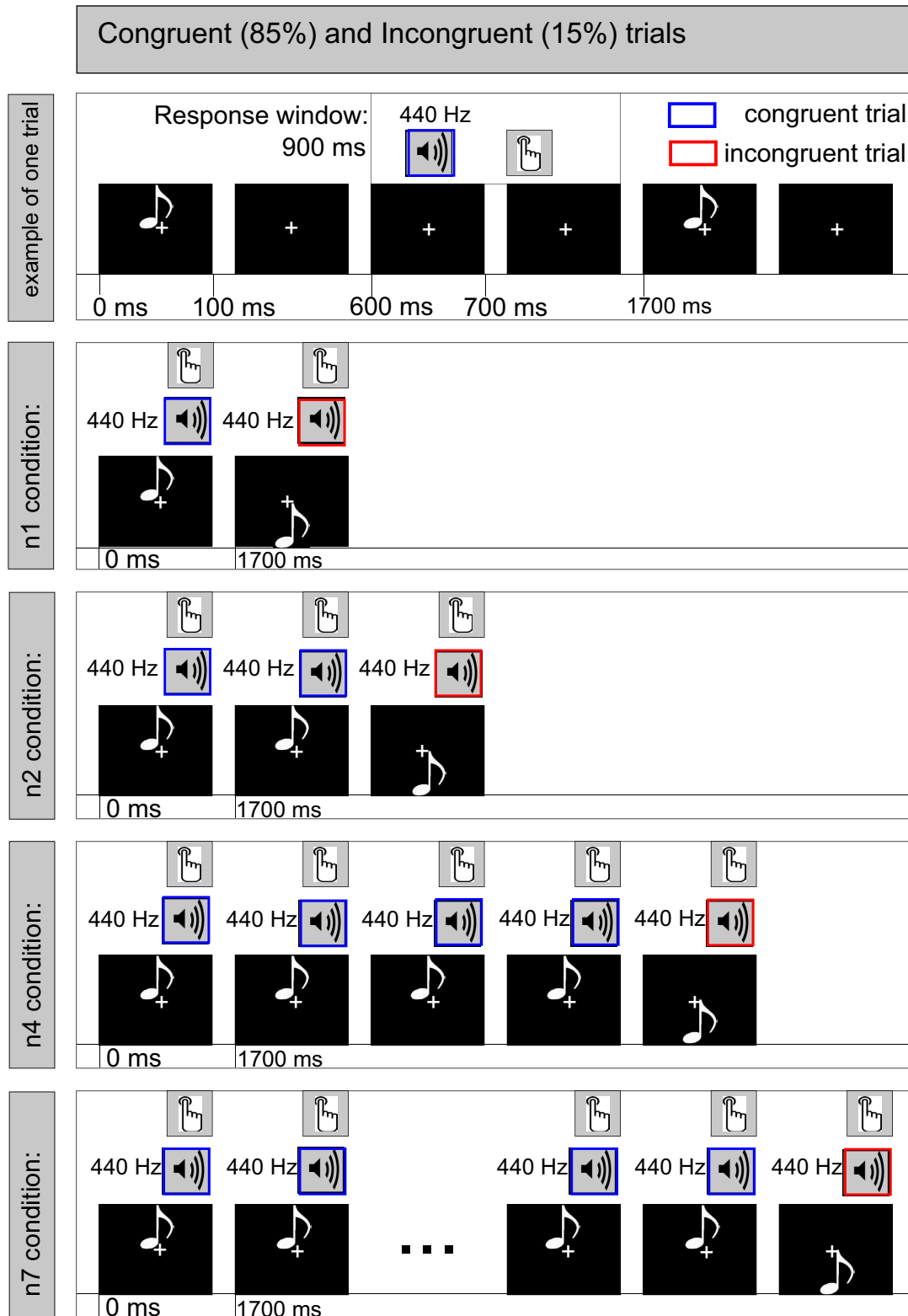
FIGURE 1 Schematic representation of the experimental paradigm. Exemplary presentation of a single trial and each trial train condition (n1, n2, n4, n7) of a high-tone train. In each trial, the visual stimulus (either eighth note symbol below or above fixation cross) preceded the auditory stimulus (either 352 Hz or 440 Hz pitched sound, 440 Hz is shown as an example here) by 600 ms. Overall, 85% of the trials were congruent (**blue:** visual symbol above fixation cross and high-pitched sound or visual symbol below fixation cross and low-pitched sound) and 15% incongruent (**red**). Depending on the trial train condition, a congruent trial was succeeded by a certain number of identical congruent trials (n1: no congruent trial repetition; n2: one congruent trial repetition; n4: three congruent trial repetitions, n7: six congruent trial repetitions) before the trial train was interrupted by an incongruent trial with the same pitched sound but a misleading visual stimulus. The participants' task was after each trial to indicate the pitch of the presented sound (via button press; association of left versus right button with low versus high tone was balanced across participants). The response window lasted 900 ms and started after the onset of the auditory stimulus

determine incongruency effects, incongruent trials were compared with congruent trials at the same place in the trial train. So, if an incongruent trial was preceded by one congruent trial (n1 condition), these incongruent trials were compared with congruent trials that were also preceded by one congruent trial. The overall ratio of congruent versus incongruent trials was 85/15, and the ratio of high versus low pitched sounds was 50/50. To obtain the 85/15 ratio, the necessary trial trains were supplemented with trial trains that were congruent to congruent (changing from one pair to the other, that is pitch change but no visual-based prediction error), or congruent to incongruent trial trains that changed in pitch (visual information stays constant). Trial trains were always preceded by a trial train with a different pitch sound (similar to a roving-standard paradigm). The trial train order was pseudo-randomized for every block, so that all possible combinations were present in every block. During the experiment, participants had to indicate whether the tone they heard was high- or low-pitched by pressing the corresponding button (association of left vs. right button with low vs. high tone was balanced across participants). Hence, the task was independent of the trial trains. Additionally, the participants were instructed to pay attention to the visual stimulus and to use the informational content to be able to respond fast and accurately. The response window started after the onset of the auditory stimulus and lasted 900 ms. The stimulus onset asynchrony (SOA) was jittered, namely the next trial (onset of the visual stimulus) started either 1,550, 1,700, or 1,850 ms (on average 1,700 ms) after the onset of the previous trial (onset of the visual stimulus). For each of the four conditions, 80 incongruent trials (only those where the change was in the visual stimulus, same pitch in congruent and incongruent trials) were acquired in 16 blocks that were later used for the analyses. Before the experimental blocks, the participants performed a training block to familiarize themselves with the stimulation (same length and probability distribution as experimental blocks). Within the whole experiment, the participants received oral feedback regarding accuracy after approximately every second block to keep them motivated. The experimental task itself had a duration of about 60 min excluding breaks (Figure 1).

2.3 | Data recording and analysis

EEG was recorded from overall 64 active Ag/AgCl electrodes. A BrainAmp amplifier and the Vision Recorder 1.20 software (Brain Products) were used. An EEG cap (actiCAP, Brain Products) was used to place the active electrodes in accordance with the extended international 10–20 system (Chatrian et al., 1985). Additional electrodes were placed at the tip of the nose (reference electrode), on the center of the forehead (ground electrode), and on the left and right mastoid sites. The horizontal and vertical electrooculogram

(EOG) was recorded from two electrodes placed on the outer canthi of the left and right eye, and one electrode below the left eye. EEG was continuously recorded (500 Hz sampling rate) and amplified with 64 active electrodes and a BrainProducts ActiCap EEG system, using a nose-tip reference. For data analyses, the EEGLAB toolbox (v13; Delorme & Makeig, 2004) for MATLAB was used. The data were processed analog to the "50/50 condition" in the data recording and analysis section of Stuckenberg et al. (2019). The data were high-pass filtered with a 0.1-Hz cutoff (finite impulse response [FIR] filter; Kaiser-windowed; Kaiser beta = 5.65; filter length = 9,056 points) and low-pass filtered with a 48-Hz cutoff (FIR filter; Kaiser-windowed; Kaiser beta = 5.65; filter length = 1,812). Epochs were generated from 100 ms before sound onset until 500 ms after the onset of the auditory stimulus. The data were baseline corrected from 0 to +50 ms after onset of the auditory stimulus analog to the preprocessing used by Pieszek et al. (2013) and Stuckenberg et al. (2019). Both studies used this baseline correction window because of contingent negative variation (CNV) potentials that were ongoing within the typical baseline window (due to increased presentation frequency of one visual stimulus). In the present study, we presented the visual stimuli with equal probability and did not observe CNV potentials. For comparison reasons, we used the same baseline correction from 0 to +50 ms (baseline correction from −100 to 0 ms revealed the same main findings). Further data preprocessing involved the identification of bad channels based on a deviation criterion (Bigdely-Shamlo et al., 2015) that detects channels with unusually high amplitude deviation. The robust z score of the robust standard deviation is calculated for each channel. Values greater than 3 identify bad channels which were removed from analysis and interpolated after the independent component analysis (ICA). An extended ICA was trained on 1-Hz filtered data and non-overlapping epochs with maximal length (−600 to 800 ms). The faster and adjust criteria (Chaumon et al., 2015; Mognon et al., 2011; Nolan et al., 2010) were used to identify bad independent components (ICs) that were removed from the data. Furthermore, the first two epochs of each block were removed. Epochs with signal changes exceeding thresholds of 150 μV after ICA artifact removal were excluded and bad channels that were identified before the ICA were interpolated.

Finally, grand averages were calculated for incongruent and congruent epochs for each trial train condition (n1, n2, n4, n7). Note that as trial trains are presented separately, congruent epochs do not overlap. That is, congruent epochs in one condition (e.g., n1) do not include congruent epochs from any other condition. In more detail, for each condition (congruent and incongruent) a maximum of 80 epochs per condition (limited by the number of incongruent presentations) was selected (see Table 1 for a detailed overview of how many epochs finally were included). The IR is encountered as an enhancement in

	n1 INC	n1 CON	n2 INC	n2 CON	n4 INC	n4 CON	n7 INC	n7 CON
Mean	74.05	74.5	78.5	76.7	78.6	78.1	78.9	78.8
In %	92.6	93.1	98.1	95.9	98.2	97.6	98.6	98.5
SD	2.1	3.9	2.7	4.6	2.6	2.6	1.8	1.7
Min	69	65	70	67	72	72	74	74
Max	77	79	80	80	80	80	80	80
Median	74	76	80	78.5	80	79	80	79.5

TABLE 1 For each condition (n1, n2, n4, n7) 80 trials were acquired. This table shows the finally included trials per condition (mean, in %, standard deviation [*SD*], min, max, median)

the difference wave of incongruent-minus-congruent ERP elicitation; hence it is important which kind of congruent epochs are selected for the calculation of the grand averages and the difference waves. To remove any confound from the signal that could have been present due to adaptation or different number of preceding congruent presentations, we contrasted incongruent epochs with congruent epochs that had the same trial train history. In other words, we made sure that the congruent trial that we analyzed was in the same position of the trial train as the incongruent trial. For instance, for condition n2 we considered incongruent epochs that were preceded by two identical congruent epochs and congruent epochs that were as well preceded by two identical congruent epochs (i.e., this one being the third presentation of the specific congruent trial). For the n7 condition, due to a minor error in the randomization, we did not have congruent epochs that were preceded by seven identical congruent trials. Our analyses showed that there were only very minor differences in the processing of the congruent trials with different history (independent of the number of identical congruent trial presentations; cf. Figure 6), hence we used congruent epochs that were preceded by six identical congruent trials in comparison to incongruent epochs that were preceded by seven identical congruent trials (analog to the typical analyses of auditory roving paradigms). Furthermore, the corresponding difference waves were calculated, resulting into four difference waves: n1: INC-CON, n2: INC-CON, n4: INC-CON, n7: INC-CON.

In addition, we calculated grand averages of the congruent trials dependent on their position in the trial train for the longest trial train sequence (n1_CON, n2_CON, n3_CON, n4_CON, n5_CON, n7_CON). This allows the analysis of the grand averages for the different congruent trials, dependent on the amount of preceding identical congruent trials (varying from one to six identical preceding congruent trials whereby for all conditions 80 trials were considered). This was done to investigate repetition suppression effects.

Behavioral data and ERP mean amplitudes within canonical region and time window of interests (ROIs; Pieszek et al., 2013; Stuckenberg et al., 2019) were tested with frequentist repeated-measures ANOVAs and with Bayesian repeated-measures ANOVAs estimating Bayes factors (BF_{10}) with identical designs. All statistical analyses were computed in JASP (version 0.11.1, JASP Team, 2018).

For all reported frequentist repeated-measures ANOVAs, an alpha level of .05 was defined. Statistically significant effects were reported, including the η^2 effect size measure. Significant interactions were investigated by computing follow-up two-tailed *t*-tests whereby Bonferroni correction was applied to control for multiple comparisons.

For the behavioral data, the first two trials of each block were removed, and the responses were grouped in the different categories according with the ERP data. To investigate the behavioral effects, a 2×4 frequentist repeated-measures ANOVA was performed including the factors Congruency (congruent vs. incongruent) and Train Length (n1 vs. n2 vs. n4 vs. n7).

For the ERP mean amplitudes, a 2×4 frequentist repeated-measures ANOVA was conducted. The ANOVA included the factors Congruency (congruent vs. incongruent) and Train Length (n1 vs. n2 vs. n4 vs. n7). The computation of the ERP mean amplitudes was based on canonical region and time window of interests (ROIs), as previously defined by Pieszek et al. (2013) and Stuckenberg et al. (2019). Hence, for the IR, an analysis time window from 105 to 130 ms after sound onset was used. For the calculation of the ERP mean amplitudes, the mean of the electrodes FC5, FC6, C3, and C4 was considered.

To confirm the significant elicitation of components in the N2-(185–225 ms) and P3-(280–340 ms) range, the ERP mean amplitudes of the electrodes Fz, FCz, and Cz (ROI midline) were considered and a 2×4 repeated-measures ANOVA with the factors Congruency (congruent versus. incongruent) and Train Length (n1 vs. n2 vs. n4 vs. n7) was performed.

To investigate repetition suppression effects, we calculated ERPs for each congruent trial. This resulted in six separate congruent conditions. The congruent condition that was preceded by one congruent visual–auditory pair (n1_CON) up to the congruent condition that was preceded by six identical congruent visual–auditory pairs (n7_CON). A one-way ANOVA of the mean amplitudes within the ROIs (FC5, FC6, C3, and C4) and the IR time window (105 to 130 ms) was performed.

For the Bayesian rANOVAs, participants' variation was included as random factor (prior scaling $r = 1$) whereby their variance was considered as nuisance. The calculation of BF_{10} was performed using 50,000 Monte–Carlo sampling iterations

and a prior scaling factor $r = .5$ (equivalent to the default "medium" effect size prior in the R BayesFactor package; Morey & Rouder, 2018) for fixed effects. The null hypothesis corresponded to a standardized effect size $\delta = 0$. We compared all models with the null model (BF_{10}). Additionally, BF_{Incl} was calculated across models including a main effect or interaction compared with equivalent models that do not include this effect (see Mathôt, 2017). Data were interpreted as moderate (or strong) evidence in favor of the alternative hypothesis if BF_{10} was ≥ 3 (or ≥ 10). If BF_{10} was ≤ 0.33 (or ≤ 0.1) this was interpreted as moderate (or strong) evidence in favor of the null hypothesis. Values close to 1 would be only weakly informative and were considered anecdotal evidence (Lee & Wagenmakers, 2013).

3 | RESULTS

3.1 | Behavioral data

3.1.1 | Response times

Response times (RTs) were faster in the congruent compared with the incongruent trials.

The 2×4 Bayesian ANOVA favored the model including the Congruency and Train Length main effects ($BF_{10} = 7.316 \times 10^{53}$). The Bayesian analysis provided strong evidence for the Congruency main effect ($BF_{Incl} = 6.901 \times 10^{53}$; $F(1, 19) = 162.497$, $p < .001$, $\eta^2 = 0.426$), the Train Length main effect ($BF_{Incl} = 32.301$; $F(1, 19) = 13.420$, $p < .001$, $\eta^2 = 0.010$), and moderate evidence against the Congruency x Train Length interaction ($BF_{Incl} = 0.113$, $F(1, 19) = 1.633$, $p = .192$, $\eta^2 = 0.001$).

As follow-up, we performed sequential testing and hence, only looked at differences between successive conditions. The follow-up Bayesian t tests and the follow-up two-tailed t tests provided strong evidence for an effect contrasting conditions n1 and n2 (n1 vs. n2: $BF_{10} = 94,173.163$, $t(19) = 8.011$, $p_{bonf} < .001$) but only anecdotal evidence for an effect contrasting the other conditions (n2 vs. n4: $BF_{10} = 2.760$, $t(19) = -2.512$, $p_{bonf} = .127$; n4 vs. n7: $BF_{10} = 0.867$, $t(19) = 1.771$, $p_{bonf} = .555$).

3.1.2 | Accuracy

Response accuracy was higher in response to congruent compared with incongruent trials.

The 2×4 Bayesian ANOVA favored the model including the Congruency, Train Length main effects, and the interaction Congruency $\times$ Train Length ($BF_{10} = 3.072 \times 10^{10}$). The Bayesian analysis provided strong evidence for the Congruency main effect ($BF_{Incl} = 3.685 \times 10^9$,

$F(1,19) = 18.595$, $p < .001$, $\eta^2 = 0.187$), and the Train Length main effect ($BF_{Incl} = 17.082$, $F(1,19) = 6.960$, $p < .001$, $\eta^2 = 0.050$) but only anecdotal evidence for the Congruency x Train Length interaction ($BF_{Incl} = 2.334$, $F(1,19) = 8.356$, $p < .001$, $\eta^2 = 0.030$).

The follow-up Bayesian t tests and the follow-up two-tailed t tests provided strong evidence for an effect contrasting congruent and incongruent conditions of the different train length conditions (n1_CON vs. n1_INC: $BF_{10} = 89.231$, $t(19) = 5.760$, $p_{bonf} < .001$; n2_CON vs. n2_INC: $BF_{10} = 26.255$, $t(19) = 4.307$, $p_{bonf} = .004$; n4_CON vs. n4_INC: $BF_{10} = 31.461$, $t(19) = 2.942$, $p_{bonf} = .165$; n7_CON vs. n7_INC: $BF_{10} = 13.857$, $t(19) = 1.825$, $p_{bonf} = 1.000$) (Figure 2).

3.2 | ERP data

3.2.1 | Incongruency response

An IR component (time window 105–130 ms) was observed in all Train Length conditions (n1, n2, n4, n7; see Figures 3 and 4).

The 2×4 Bayesian ANOVA favored the model including the Congruency and Train Length main effects ($BF_{10} = 1.324 \times 10^{14}$). The Bayesian analysis provided strong evidence for the Congruency main effect ($BF_{Incl} = 1.525 \times 10^{14}$, $F(1,19) = 50.077$, $p < .001$, $\eta^2 = 0.137$) and anecdotal evidence for the Train Length main effect ($BF_{Incl} = 1.790$, $F(1,19) = 4.542$, $p = .006$, $\eta^2 = 0.015$). Furthermore, the Bayesian analysis provided strong evidence against the Congruency $\times$ Train Length interaction ($BF_{Incl} = 0.086$, $F(1,19) = 0.241$, $p = .867$, $\eta^2 = 0.001$).

3.2.2 | N2

An N2 component was elicited in response to incongruent sounds. ERP mean amplitudes in the N2 range (185–225 ms) were more negative in incongruent compared with congruent trials in all Train Length conditions (n1, n2, n4, and n7; see Figure 5).

The 2×4 Bayesian ANOVA favored the model including the Congruency main effect ($BF_{10} = 7.974 \times 10^{32}$). The Bayesian analysis provided strong evidence for the Congruency main effect ($BF_{Incl} = 8.098 \times 10^{32}$, $F(1,19) = 60.529$, $p < .001$, $\eta^2 = 0.479$) and strong evidence against the Train Length main effect ($BF_{Incl} = 0.055$, $F(1,19) = 1.168$, $p = .330$, $\eta^2 = 0.002$). Furthermore, the Bayesian analysis provided moderate evidence against the Congruency $\times$ Train Length interaction ($BF_{Incl} = 0.114$, $F(1,19) = 1.384$, $p = .257$, $\eta^2 = 0.002$).
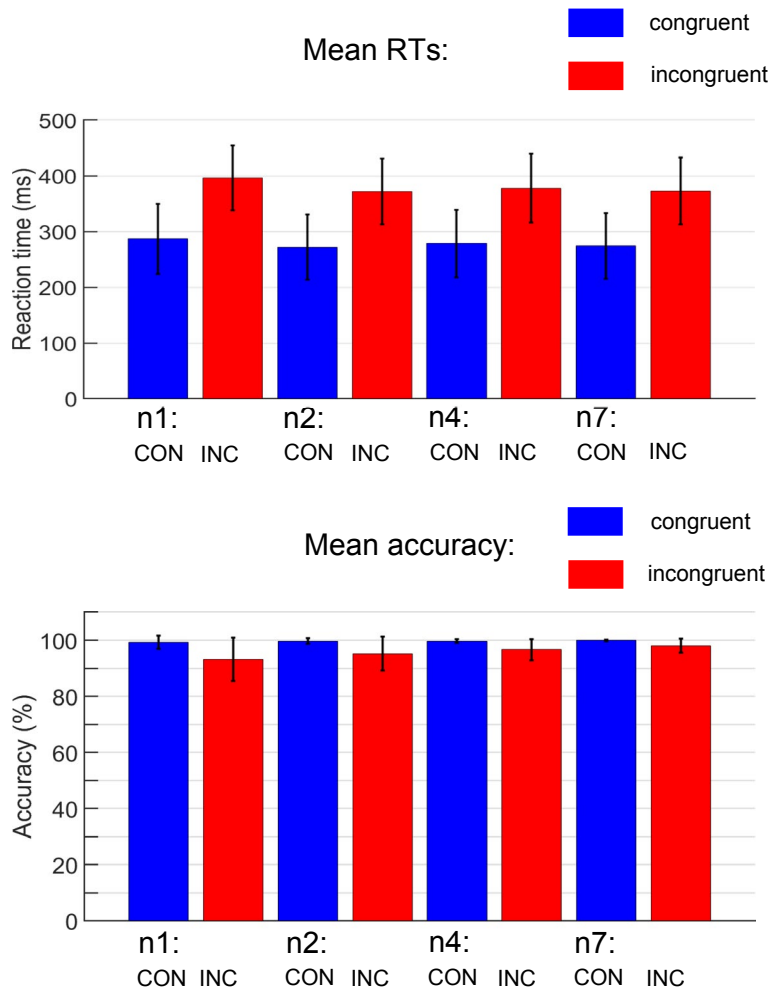
FIGURE 2 Mean RTs and mean accuracy data and their standard deviations

3.2.3 | P3

A P3 component was elicited in response to incongruent sounds. ERP mean amplitudes in the P3 range (280–340 ms) were more positive in incongruent compared with congruent trials in all Train Length conditions (n1, n2, n4, and n7; see Figure 5).

The 2 × 4 Bayesian ANOVA favored the model including the Congruency main effect ($BF_{10} = 7.770 \times 10^{10}$). The Bayesian analysis provided strong evidence for the Congruency main effect ($BF_{Incl} = 8.560 \times 10^{10}$, $F(1,19) = 17.023, p < .001, \eta^2 = 0.112$). The frequentist analysis revealed a significant Train Length main effect, but the Bayesian analysis revealed moderate evidence against an effect ($BF_{Incl} = 0.232, F(1,19) = 3.653, p = .018, \eta^2 = 0.008$). Furthermore, the Bayesian analysis provided anecdotal evidence against the Congruency × Train Length interaction, but the frequentist analysis revealed a significant interaction ($BF_{Incl} = 0.513, F(1,19) = 4.268, p = .009, \eta^2 = 0.009$).

3.2.4 | Repetition suppression effects

A one-way ANOVA was performed unveiling a non-significant main effect of the factor Number of Congruent Repetitions ($F(5,95) = 0.404, p = .795$, GG $\varepsilon = .755$, $\eta^2 = 0.004$; see Figure 6). The Bayesian ANOVA revealed very strong evidence against an effect of the factor Number of Congruent Repetitions ($BF_{10} = 0.043$).

As repetition suppression effects were potentially to be expected also outside our IR mean amplitude time window at midline electrode sites (Haenschel et al., 2005), we tested for effects of number of repetitions separately at each latency in a running Bayesian ANOVA at Cz electrode site where repetition suppression effects are more typically observed. All observed BF_{10} were below 1 (and all GG corrected p values > .05 in a corresponding frequentist ANOVA). That is, we did not find any evidence for, rather evidence against repetition suppression effects at any latency.

4 | DISCUSSION

The present study investigated the influence of local repetition on the elicitation of the IR. Two different visual-cue-sound combinations were presented (high and low) with equal global presentation probability. Local repetitions were manipulated by presenting identical congruent trial trains
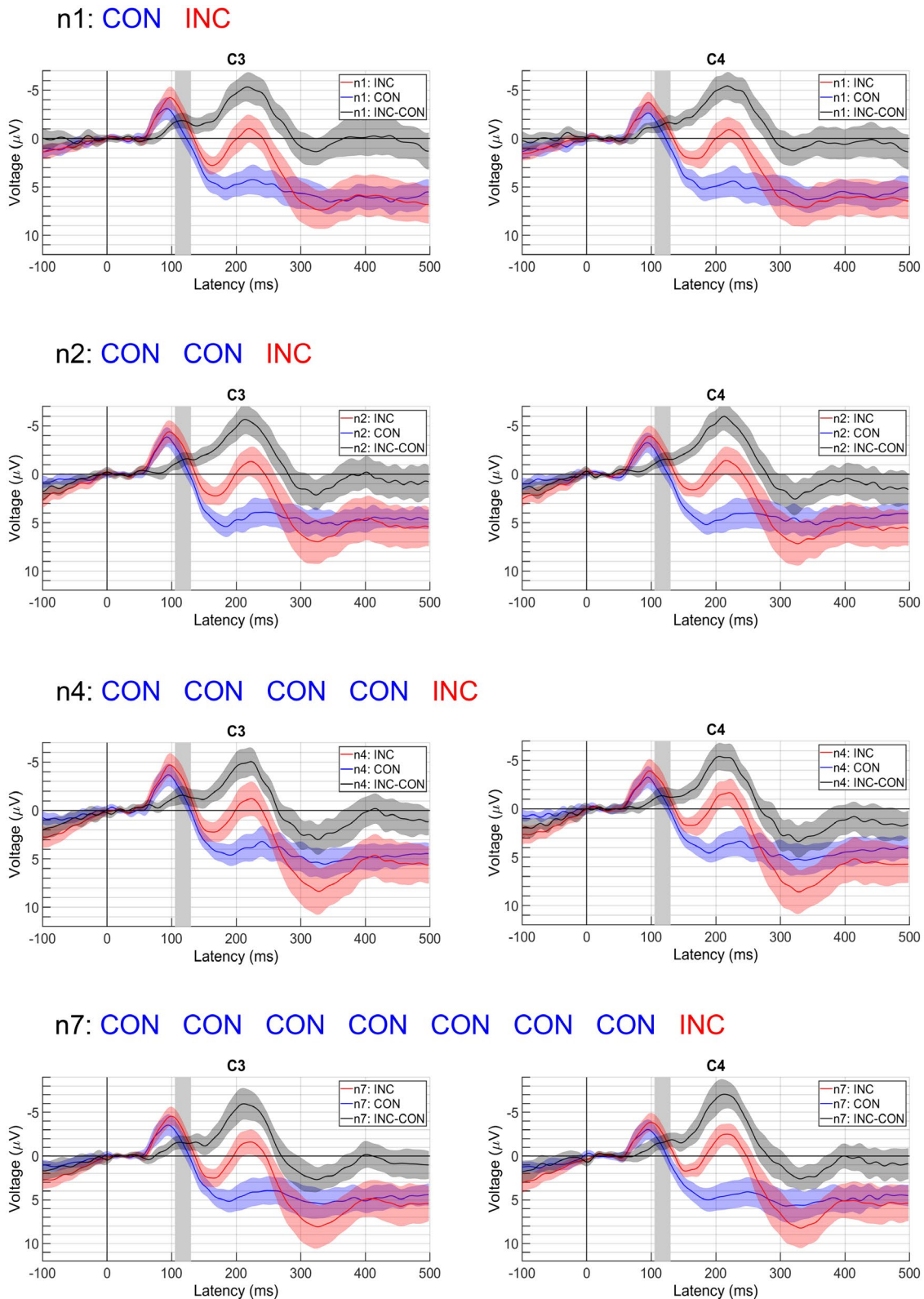
FIGURE 3 Auditory ERPs and difference waveforms for each condition (Train Length: n1, n2, n4 and n7). The grey bar indicates the statistical testing window for the IR (105–130 ms). Shaded areas surrounding ERP curves reflect corresponding 95% confidence intervals (CIs)

(either high or low trial trains) of four different lengths, which were concluded by an incongruent trial. The incongruent trial presented a different visual cue as the trial train but the same sound, leaving the visual cue responsible for any changes in the incongruent auditory ERP. We assumed that if the IR is driven by local repetition, the elicitation of the

PSYCHOPHYSIOLOGY SPR
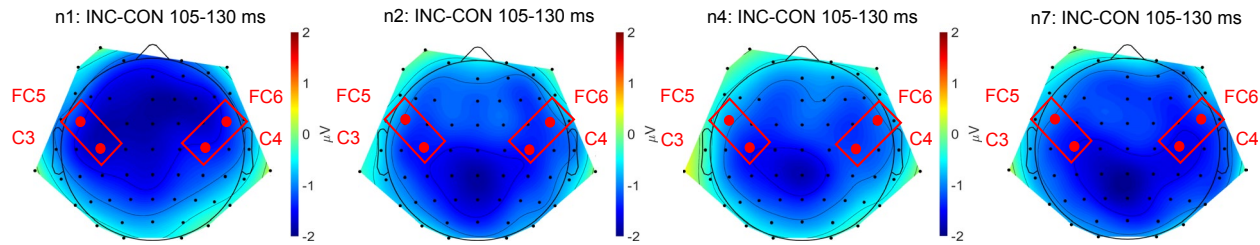
FIGURE 4 Potential maps (nose referenced) show the scalp distribution of the difference data (incongruent-minus-congruent) for each Train Length condition (n1, n2, n4, n7) within the IR time window (105–130 ms). ROIs are indicated in red
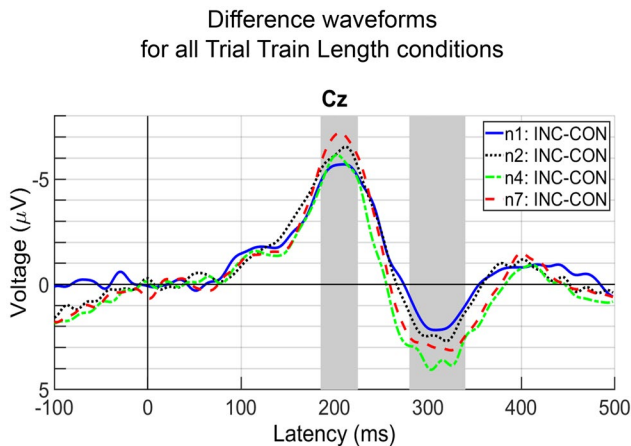

FIGURE 5 Difference waveforms (incongruent-minus-congruent) for each Train Length condition (n1, n2, n4 and n7). Exemplary plot at Cz. Statistical testing window for the investigation of the N2 (185–225 ms) and P3 (280–340 ms) elicitation is indicated with a grey bar

IR would be more likely to occur (and possibly increase in amplitude) with increasing trial train length.

4.1 | Influence of local repetition on IR elicitation

Surprisingly, we observed an IR in all train length conditions (n1, n2, n4, and n7) when comparing the mean ERP amplitudes elicited by congruent and incongruent visual-cue-sound combinations. Perhaps even more surprising is that the observed IR amplitude did not change as the number of preceding congruent trials increased.

Predictive coding considers repetition suppression effects of the auditory N1 with increasing repetitions to be an effect of increasingly successful prediction of the upcoming stimulus (Auksztulewicz & Friston, 2016; Winkler et al., 2009). Hence, a priori, our hypothesis regarding local repetition was that an increasing local suppression effect might build up as the trial train of identical congruent trial presentations gets longer. In the concluding incongruent trial of the trial train, the visual-based prediction changes

as another visual stimulus is presented. When in this incongruent trial the sound unexpectedly stays the same, the violation of the visual-based prediction elicits an auditory ERP in which the N1 is not suppressed. When computing the difference wave between congruent and incongruent trials, this effect would show up as the IR. This hypothesized model of the IR suggests that as the trial train gets longer, repetition suppression increases, and so the IR would be more likely to occur (and increase in amplitude) with increasing trial train length. However, our study finds an IR that is already elicited after one preceding congruent trial (n1 condition), where the IR does not increase in amplitude as the trial train gets longer (see Figure 3). The typical IR topography, with a central negativity and a small postero-lateral positivity (see e.g., Pieszek et al., 2013, 2014), was observed in all trial train conditions (Figure 4). The presence of the IR in the absence of a local rule demonstrates that the IR—which can be considered an early, low-level prediction error signal—can be elicited on the basis of longer-term associations. Therefore, a different model of the IR seems more fitting, in which visual–auditory associations are learned at higher levels, and the auditory stimulus is predicted top-down as soon as the visual stimulus appears.

Support for this model can also be found in the ERPs of the congruent trials. To further investigate the expected repetition suppression effects in the congruent trials, we analyzed congruent ERPs for each position in the longest trial train (n7). This resulted in six separate congruent conditions (see Figure 6), ranging from congruent trials being preceded by one identical congruent trial (n1_CON) to being preceded by six identical congruent trials (n7_CON). ERPs in the time range of 105–130 ms showed no indicators of repetition suppression as the trial train length increased. This is surprising given the gradual suppression that is observed in purely auditory paradigms (e.g., Haenschel et al., 2005), the well-documented phenomena of stimulus-specific adaptation (Escera & Malmierca, 2014), and the N1-suppression effect for expected sounds (Lange, 2013). However, in contrast to normal roving-standard paradigms, the current paradigm presents only two different tones, and a visual cue predicts the upcoming tone. This allows the generative model to learn

Congruent trial processing
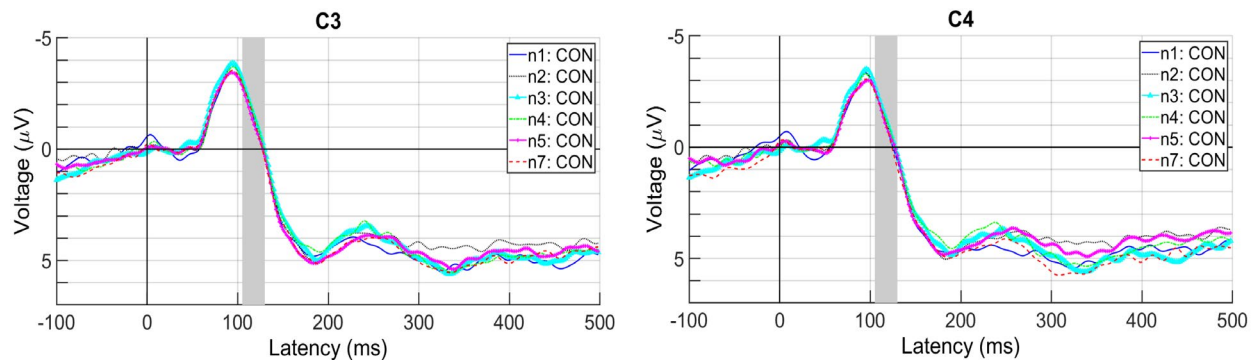different number of preceding congruent trials
Auditory ERPs

FIGURE 6 Auditory ERPs elicited by congruent trials. The number of congruent repetitions, that is identical trial repetitions, varies from one preceding congruent trial presentation (n1_CON) up to six preceding identical congruent trials (n7_CON). Exemplary for electrodes C3 and C4. Statistical testing window is indicated with a grey bar (105–130 ms)

a higher order association, making it unnecessary for a prediction to build up over multiple trials. In other words, the lack of a gradual repetition suppression might be the result of floor effects as the tone is predicted optimally already from n1_CON onwards. This fits the model of instantly effective top-down, visual-based prediction effects on auditory processing.

Wacongne et al. (2011), Haenschel et al. (2005), and other studies (Bekinschtein et al., 2009; Costa-Faidella et al., 2011; El Karoui et al., 2015; Nourski et al., 2018) demonstrate that early components are sensitive to local repetition/violation. However, Wacongne et al. (2011) also demonstrate that global rules can affect earlier components. In their study, the MMN elicited by expected deviants was significantly reduced compared with the MMN elicited by unexpected deviants. Wacongne et al. (2011) state that this can be interpreted as an influence of higher-level prediction on first-order error response, which is consistent with hierarchical predictive coding. The current study shows that similar effects might be at play on a cross-modal level, where visual predictions influence auditory perception through a top-down pathway.

The visual–auditory relation in regard of predictive processing mechanisms is especially important for speech comprehension and human communication (see e.g., Kokinous et al., 2015, 2017; von Kriegstein et al., 2008). Nevertheless, there might be differences between unimodal (auditory-auditory or visual–visual) and cross-modal (visual–auditory) paradigms. For instance, it is unknown whether the mapping of visual symbols to sounds would result in similar effects when applied to a unimodal setting (e.g., visual symbol predicts other visual symbol). Therefore, more research is required to fully understand the underlying mechanisms.

4.2 | Subsequent processing

In all trial train length conditions, N2 and P3 components were elicited. Similar to the IR findings, we did not observe increasing amplitudes with increasing trial train length for later processing stages (N2: 185–225 ms and P3: 280–340 ms). N2 (i.e., N2b) and P3 (i.e., P3a) have been reported to reflect the response to task-relevant or highly salient infrequent sounds (e.g., Patel & Azzam, 2005; Polich, 2003). This finding shows that also higher-order prediction error-related processes are independent of trial train length.

4.3 | Stuckenberg et al. (2019) and previous IR studies

If the IR is indeed the result of a learned, higher-order association that is projected down to lower levels, the question remains why no IR was observed in the equal probability condition of Stuckenberg et al. (2019).

One notable difference between the current study and Stuckenberg et al. (2019) is that in the latter study there was no control of pitch change before congruent and incongruent trials. By contrast, the current study compared incongruent and congruent trials that were always preceded by a congruent trial with the same pitch sound. In the 83/17 condition of Stuckenberg et al. (2019) preceding same pitched sounds must have happened repeatedly since the amount of congruent visual-cue-sound combinations with the same pitch was higher and hence, the likelihood that such a trial would precede an incongruent trial was higher as well. In other words, the 50/50 condition of Stuckenberg et al. (2019) might have been perceived as somewhat

chaotic, with frequent pitch changes plus occasional incongruent trials. This might have required considerably more mental resources, possibly preventing the establishment of a stable mental model and the generation of a respective prediction. In contrast, the 83/17 condition provides a more stable context in which one sound is dominant. Similarly, in the current study the stable context is provided by the pseudo-randomization constraints, resulting in a predictable roving structure that alternately presents strings of identical trials. The IR elicitation might have been facilitated by these contextual aspects despite a 50/50 sound probability. Finally, Widmann et al. (2004) also observed an IR elicitation using a 50/50 high/low sound probability. They presented a sequence of score-like visual symbols before a corresponding sound sequence. Again, this kind of context (visual pattern) might provide more stability and facilitate prediction processes.

4.4 | Behavioral measures

Response time was faster in the congruent compared with the incongruent trials. Incongruent responses were slowest and more often wrong in the n1 condition compared with the other trial train length conditions (n2, n4, n7). This implies that the participants needed more resources in the incongruent n1 condition to accurately respond to the auditory stimulus. We do not think that these differences are necessarily attributed to differences in perception of the sound. It could be that participants relied more on the assumed predictive value of the visual information. Whereas in the other conditions, participants could have encountered that the same tone (i.e., visual-cue-sound combination) is presented repeatedly and hence, are less distracted by a deviating visual stimulus. In other words, due to the repeated identical tones, participants are probably more prone to press the same button again even if the visual cue suggests a pitch change (e.g., Horner & Henson, 2008; Koch et al., 2011; Mawase et al., 2018). We therefore assume that the observed behavioral effects reflect effects of increasing number of response repetitions rather than differences in sound perception.

5 | CONCLUSIONS

Based on the existing literature, we assumed that preceding visual information of the upcoming auditory stimulus would lead to visual-based sensory predictions. Violating the visual prediction would result in an increased response in the N1 range of the auditory evoked potential in comparison to the processing of identical sounds that were expected based on the visual cue (i.e., the incongruency response, IR).

First, results show that the IR is not the direct result of local, low-level associations between visual and auditory information. Instead, it seems more likely that it is the result of learned visuo–auditory associations (i.e., cross-modal generative models) that trigger top-down predictions instantly as soon as the visual stimulus is presented. Second, the conditions under which the visual–auditory associations are learned and applied might be connected to the context of the trial presentation. This has been shown through dominance of one visual–auditory pair over the other, a visual overview of the sequence presentation, or, in the current study, through local repetition of the same visual–auditory pair. From this point of view, local repetition might be considered to have indirectly facilitated the elicitation of the IR in the current experiment, by creating a structured presentation of trials.

CONFLICT OF INTEREST
The authors declare no competing financial interests.

AUTHOR CONTRIBUTION
Tjerk Tian Dercksen: Conceptualization; Investigation; Writing-original draft. **Maria Viktoria Stuckenberg:** Conceptualization; Data curation; Formal analysis; Investigation; Project administration; Visualization; Writing-original draft. **Erich Schröger:** Resources; Supervision; Writing-review & editing. **Nicole Wetzel:** Funding acquisition; Supervision; Writing-review & editing. **Andreas Widmann:** Conceptualization; Data curation; Project administration; Supervision; Writing-original draft.

ORCID
Maria V. Stuckenberg (iD) https://orcid.org/0000-0002-2391-0221

REFERENCES
Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, *80*, 125–140. https://doi.org/10.1016/j.cortex.2015.11.024

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing

of auditory regularities. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(5), 1672–1677. https://doi.org/10.1073/pnas.0809667106

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*(16), 1–20. https://doi.org/10.3389/fninf.2015.00016

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *American Journal of EEG Technology*, *25*(2), 83–92. https://doi.org/10.1080/00029238.1985.11080163

Chaumon, M., Bishop, D. V., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, *250*, 47–63. https://doi.org/10.1016/j.jneumeth.2015.02.025

Costa-Faidella, J., Baldeweg, T., Grimm, S., & Escera, C. (2011). Interactions between "what" and "when" in the auditory system: Temporal predictability enhances repetition suppression. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(50), 18590–18597. https://doi.org/10.1523/JNEUROSCI.2599-11.2011

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Eimer, M., & Driver, J. (2001). Crossmodal links in endogenous and exogenous spatial attention: Evidence from event-related brain potential studies. *Neuroscience & Biobehavioral Reviews*, *25*(6), 497–511. https://doi.org/10.1016/S0149-7634(01)00029-X

El Karoui, I., King, J. R., Sitt, J., Meyniel, F., Van Gaal, S., Hasboun, D., Adam, C., Navarro, V., Baulac, M., Dehaene, S., Cohen, L., & Naccache, L. (2015). Event-related potential, time-frequency, and functional connectivity facets of local and global auditory novelty processing: An intracranial study in humans. *Cerebral Cortex*, *25*(11), 4203–4212. https://doi.org/10.1093/cercor/bhu143

Escera, C., & Malmierca, M. S. (2014). The auditory novelty system: An attempt to integrate human and animal research. *Psychophysiology*, *51*(2), 111–123. https://doi.org/10.1111/psyp.12156

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(45), 10494–10501. https://doi.org/10.1523/JNEUROSCI.1227-05.2005

Horner, A. J., & Henson, R. N. (2008). Priming, response learning and repetition suppression. *Neuropsychologia*, *46*(7), 1979–1991. https://doi.org/10.1016/j.neuropsychologia.2008.01.018

JASP Team. (2018). *JASP [Computer software]* (Version 0.11.1). https://jasp-stats.org/download

Koch, I., Schuch, S., Vu, K.-P.-L., & Proctor, R. W. (2011). Response-repetition effects in task switching—Dissociating effects of anatomical and spatial response discriminability. *Acta Psychologica*, *136*(3), 399–404. https://doi.org/10.1016/j.actpsy.2011.01.006

Kokinous, J., Kotz, S. A., Tavano, A., & Schröger, E. (2015). The role of emotion in dynamic audiovisual integration of faces and voices. *Social Cognitive and Affective Neuroscience*, *10*(5), 713–720. https://doi.org/10.1093/scan/nsu105

Kokinous, J., Tavano, A., Kotz, S. A., & Schröger, E. (2017). Perceptual integration of faces and voices depends on the interaction of emotional content and spatial frequency. *Biological Psychology*, *123*, 155–165. https://doi.org/10.1016/j.biopsycho.2016.12.007

Lange, K. (2013). The ups and downs of temporal orienting: A review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. *Frontiers in Human Neuroscience*, *7*(263), 1–14. https://doi.org/10.3389/fnhum.2013.00263

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. https://bayesmodels.com/

Marti, S., Thibault, L., & Dehaene, S. (2014). How does the extraction of local and global auditory regularities vary with context? *PLoS ONE*, *9*(9), e107227. https://doi.org/10.1371/journal.pone.0107227

Mathôt, S. (2017). *Bayes like a Baws: Interpreting Bayesian repeated measures in JASP*. https://www.cogsci.nl/blog/interpreting-bayesian-repeated-measures-in-jasp

Mawase, F., Lopez, D., Celnik, P. A., & Haith, A. M. (2018). Movement repetition facilitates response preparation. *Cell Reports*, *24*(4), 801–808. https://doi.org/10.1016/j.celrep.2018.06.097

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229–240. https://doi.org/10.1111/j.1469-8986.2010.01061.x

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes factors for common designs*. R Package Version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor

Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods*, *192*(1), 152–162. https://doi.org/10.1016/j.jneumeth.2010.07.015

Nourski, K. V., Steinschneider, M., Rhone, A. E., Kawasaki, H., Howard, M. A., & Banks, M. I. (2018). Processing of auditory novelty across the cortical hierarchy: An intracranial electrophysiology study. *NeuroImage*, *183*, 412–424. https://doi.org/10.1016/j.neuroimage.2018.08.027

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected studies of the event-related potential. *International Journal of Medical Sciences*, *2*(4), 147–154. https://doi.org/10.7150/ijms.2.147

Pieszek, M., Schröger, E., & Widmann, A. (2014). Separate and concurrent symbolic predictions of sound features are processed differently. *Frontiers in Psychology*, *5*, 1295. https://doi.org/10.3389/fpsyg.2014.01295

Pieszek, M., Widmann, A., Gruber, T., & Schröger, E. (2013). The human brain maintains contradictory and redundant auditory sensory predictions. *PLoS ONE*, *8*(1), e53634. https://doi.org/10.1371/journal.pone.0053634

Polich, J. (2003). Theoretical overview of P3a and P3b. In J. Polich (Ed.), *Detection of change: Event-related potential and fMRI findings* (pp. 83–98). Springer US. https://doi.org/10.1007/978-1-4615-0294-4_5

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Stuckenberg, M. V., Schröger, E., & Widmann, A. (2019). Presentation probability of visual-auditory pairs modulates visually induced auditory predictions. *Journal of Cognitive Neuroscience*, *31*(8), 1110–1125. https://doi.org/10.1162/jocn_a_01398

von Kriegstein, K., Dogan, Ö., Grüter, M., Giraud, A.-L., Kell, C. A., Grüter, T., Kleinschmidt, A., & Kiebel, S. J. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(18), 6747–6752. https://doi.org/10.1073/pnas.0710826105

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20754–20759. https://doi.org/10.1073/pnas.1117807108

Widmann, A., Kujala, T., Tervaniemi, M., Kujala, A., & Schröger, E. (2004). From symbols to sounds: Visual symbolic information activates sound representations. *Psychophysiology*, *41*(5), 709–715. https://doi.org/10.1111/j.1469-8986.2004.00208.x

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*(12), 532–540. https://doi.org/10.1016/j.tics.2009.09.003