Attention and prediction in human audition: a lesson from cognitive psychophysiology

Erich Schröger, Anna Marzecová and Iria SanMiguel
Institute for Psychology, BioCog - Cognitive and Biological Psychology, University of Leipzig, Neumarkt 9-19, D-04109 Leipzig, Germany

Abstract

Attention is a hypothetical mechanism in the service of perception that facilitates the processing of relevant information and inhibits the processing of irrelevant information. Prediction is a hypothetical mechanism in the service of perception that considers prior information when interpreting the sensorial input. Although both (attention and prediction) aid perception, they are rarely considered together. Auditory attention typically yields enhanced brain activity, whereas auditory prediction often results in attenuated brain responses. However, when strongly predicted sounds are omitted, brain responses to silence resemble those elicited by sounds. Studies jointly investigating attention and prediction revealed that these different mechanisms may interact, e.g. attention may magnify the processing differences between predicted and unpredicted sounds. Following the predictive coding theory, we suggest that prediction relates to predictions sent down from predictive models housed in higher levels of the processing hierarchy to lower levels and attention refers to gain modulation of the prediction error signal sent up to the higher level. As predictions encode contents and confidence in the sensory data, and as gain can be modulated by the intention of the listener and by the predictability of the input, various possibilities for interactions between attention and prediction can be unfolded. From this perspective, the traditional distinction between bottom-up/exogenous and top-down/endogenous driven attention can be revisited and the classic concepts of attentional gain and attentional trace can be integrated.

Introduction

Prediction and attention are theoretical constructs with a long tradition in psychology and biology. In 1867 the physiologist Hermann von Helmholtz described perception as relying on unconscious inferences (Helmholtz, 1867). Prediction still denotes such a hypothetical 'inferential' mechanism in the service of perception that considers prior information when interpreting the sensorial input. In 1890 the psychologist William James gave what is still a valid description of the functional role of attention for perception: "immediate effects of attention are to make us perceive … better than otherwise we could" (James, 1890, Chapter XI, Attention). Thus both (prediction and attention) have the joint function of aiding perception, yet they are rarely considered together. Questions such as whether these two constructs are the same, and, if not, whether and how they interact, are rarely addressed.

The present review provides an overview of what we learned from cognitive psychophysiology about prediction and attention in audition and discusses these two concepts under the umbrella of the predictive coding theory (Friston, 2009, 2010). For pragmatic reasons we have

to be selective with respect to the type of brain measures and experimental paradigms that we consider. This article will mainly refer to research measuring stimulus-evoked brain activity, although oscillatory activity also taps into prediction and attention (e.g. Arnal & Giraud, 2012). With respect to the paradigms we will mainly refer to variants of two classic paradigms studying prediction and attention with the event-related potential (ERP) technique, both published in 1973 in *Science*, i.e. the 'self-generation paradigm' tapping into auditory prediction (Schafer & Marcus, 1973) and the selective listening paradigm tapping into auditory selective attention (Hillyard *et al.*, 1973).

The joint comparison of prediction and attention will yield evidence that the predictive coding theory is able to integrate attention and prediction into a common framework. Whereas prediction relates to predictions sent down from predictive models housed in higher levels of the processing hierarchy to lower levels of the hierarchy, attention refers to gain modulation of the prediction error signal sent up to the higher level via the feedforward connections. According to predictive coding theory, predictive models generate predictions regarding both the contents of the input and their inferred precision (Feldman & Friston, 2010). The gain can be modulated by the intention of listener, but also by the predictability of the input. This enables various possibilities for interactions and can explain various patterns of results. From this perspective, the traditional distinction between bottom-up/exogenous and top-down/endogenous driven attention (James, 1890) can be revisited and the classic concepts of attentional gain (Hillyard *et al.*, 1973) and attentional trace (Näätänen *et al.*, 1978) can find their place.

Cognitive psychophysiology of auditory prediction

A widely accepted account of perception states that perception results from the interaction between the sensory input and an internal model of the world generating predictions about the ongoing sensory input. The internal model contains inferred causes of the sensorial input, i.e. the representation or belief of what in the outer (hidden) world has generated the input. The inference about hidden causes follows Bayesian rules. It is based on estimates of likelihood and prior probability (prior belief based on our experience), which constrain and bias our interpretation of the sensorial information (posterior probability). To verify the validity of the brain's inference and to decide which of the concurrently active models applies, the inferred causes are used to formulate predictions about the sensory input. The predictions are compared with sensory input and the difference expresses prediction error, which, in turn, is used to update the model and keep it faithful (Rao & Ballard, 1999). According to these models, perception can be understood as a result of hierarchical Bayesian inference, in which bottom-up sensory input and top-down predictions are integrated in recurrent feedforward/feedback loops across different levels of the cortical hierarchy (Lee & Mumford, 2003).

The most influential computational implementation of these principles is Friston's predictive coding theory (e.g. Friston, 2009; Friston & Kiebel, 2009). This model proposes that the hierarchical message passing takes place between representational units that encode the inferred causes of sensory input (the model of predicted states) and prediction error units. Prediction error units compare predictions received from the higher level via backward, top-down projections and inputs received from the lower level via feedforward, bottom-up projections. This dynamic system of feedback/feedforward recurrent loops aims at minimizing the prediction error. At each level of the cortical hierarchy, only the prediction error is passed onto the higher levels (Fig. 1). As a consequence, the amount of sensory data that are fed forward and need to be processed further is reduced to only those parts that are not already accounted for by the model. In fact, the predictive coding theory interprets electrophysiological measures of brain activity as an

expression mainly of prediction error (see, e.g. Garrido *et al.*, 2007, 2009; Feldman & Friston, 2010). From this perspective, sensory ERPs are understood as the transient expression of prediction error (Friston, 2005), which is going to be suppressed by increasingly improved predictions from higher areas. Once the input is explained by the prediction, the perceptual problem is solved and we perceive our internal model, i.e. the content that is specified by representational neurons. This solution of the perceptual problem is more difficult for novel stimuli and for stimuli that do not fit to the context in which they appear. At first glance, it may sound counter-intuitive to cognitive psychophysiologists that ERPs such as the N100 can be regarded as prediction error signals. However, it provides a coherent interpretation of many experimental results from prediction research, which we will address in this section, from attention research (Section Cognitive psychophysiology of auditory attention) and from research tapping into the interaction between prediction and attention (Section Studies on the relation between auditory attention and prediction).

Prediction effects in audition have often been explored with mismatch (oddball) paradigms, where the auditory stimulation that is presented follows a particular rule, enabling the system to formulate predictions. Once rules are established, the predictions that the system could have made are either matched or mismatched by the current input (Bendixen *et al.*, 2012). It has been found in thousands of studies that incoming sounds that violate an established regularity evoke mismatch responses that can be interpreted as prediction error signals occurring at different levels along the hierarchy of predictive models. One example of such error signals is the mismatch negativity (MMN) of the auditory ERP or its magnetoencephalographic equivalent, which occurs at around 150 ms after the onset of an irregular sound. The MMN has neural generators in the auditory cortex, and is elicited even when subjects do not attend to the auditory stimulation (e.g. Kujala *et al.*, 2007). Another example of a neurophysiological error signal that is elicited by salient irregular sounds is the P3a, a frontocentral positivity occurring at around 250–300 ms after the onset of the salient sound, which is proposed to reflect the involuntary orienting of attention towards the perturbating sound (Escera *et al.*, 2000; Friedman *et al.*, 2001; Polich, 2007). Thus, involuntary attention may come into play when an expectation is violated (Parmentier *et al.*, 2011; Wetzel & Schröger, 2014). Research on auditory error signals such as the MMN has been extensively reviewed from the predictive coding perspective in several reviews (e.g. Garrido *et al.*, 2009; Winkler *et al.*, 2009; Wacongne *et al.*, 2012; Schröger *et al.*, 2014). We will mainly review studies using psychophysiological indicators of prediction that have not yet been addressed frequently. First, we will describe studies investigating the effects of prediction on the processing of predicted sounds, which have shown that responses to predicted sounds are attenuated. A particular case of sensory predictions will be considered, namely those that can be formulated on the basis of one's own motor acts. Second, we will describe studies that have attempted to study the underlying prediction mechanisms that result in the above-mentioned modulations of the stimulus-evoked responses, investigating prediction effects in the absence of stimulation. These studies have investigated purely endogenous neurophysiological responses that can be observed when predicted stimuli are omitted.
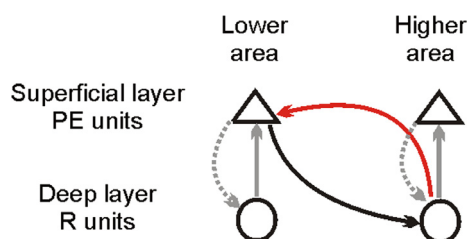

FIG. 1. Skeleton of the flow of information as postulated by the predictive coding theory. Generative models encoding the inferred causes of the sensorial input in representational units (R), located at a higher level, send predictions down to a lower level, where they are compared with the input arriving at the lower level from a still lower level (hierarchical system). The mismatch between the two is computed by prediction error (PE) units and sent forward to the higher level, so that the generative model can be improved and kept faithful. As only the PE is passed on to the higher levels, the amount of sensory data that need to be processed further is reduced to only those parts that are not already accounted for by the model. The system tries to minimize the PE, which is assumed to be generated by superficial pyramidal cells, where ERPs (and high-frequency oscillatory activity) are generated to a large extent. Deep pyramidal cells seem to be involved in the transmission of information backwards throughout the hierarchy, where more sustained ERPs (and low-frequency oscillatory activity) are generated. This is a simplified version of a figure published in Friston (2005) as Fig. 2.

Prediction effects on stimulus-evoked responses: self-generation studies

In a pioneering study, Schafer & Marcus (1973) showed that self-administered auditory stimuli elicit smaller amplitudes of the N1

component of the auditory ERP than the N1 amplitudes to identical but machine-delivered stimuli (Fig. 3C). This self-generation suppression effect has since been replicated in several studies (for a comprehensive review, see Hughes *et al.*, 2012) and it is usually believed to result from predictive processing. In short, it is proposed that the planning and execution of a motor act are accompanied by an estimation (prediction) of its foreseeable sensory consequences, which is delivered to the sensory cortices. If there is a match between the predicted and the received sensory consequences, sensory responses are suppressed. This idea finds its roots in classic animal physiology studies (von Holst & Mittelstaedt, 1950; Sperry, 1950), and was later embraced and revived by computational models of motor control (Miall & Wolpert, 1996). Classic studies addressed the conundrum that, although the movement of the eyes causes a displacement of the visual image in the retina, we do not perceive the world as moving, instead, stable visual perception is retained through saccades. The conundrum can be solved by postulating the reafference principle, which assumes that a predictive mechanism uses predictions of the sensory consequences of the motor act to compensate the motion effects. In line with these ideas, motor control researchers have inferred that the existence of internal forward and inverse models is necessary to explain our motor abilities (Miall & Wolpert, 1996). Forward models estimate, given the current state of the system and the motor commands sent, the future state of the system and the associated sensory consequences. Internal forward models are used, for example, to estimate motor errors during performance and to adjust motor commands online overcoming feedback delays, or to plan motor behavior in advance by testing several alternatives and estimating their outcomes before the execution of the motor behavior. Inverse models, however, estimate the motor commands needed to achieve a certain state. This view has thus elaborated on the physiological postulates of the reafference principle, and considers sensory prediction as a necessary and integral part of motor control. For the sake of completeness, it should be mentioned that increased rather than suppressed responses to self-generated sounds in the auditory cortex have also been reported (Reznik *et al.*, 2014).

In audition, the processing of self-generated sounds has been studied either using the subject's own speech as the self-generated auditory stimulation (e.g. Gunji *et al.*, 2000; Kudo *et al.*, 2004; Baess *et al.*, 2011; Mock *et al.*, 2011), or asking participants to deliver sounds via button presses (e.g. McCarthy & Donchin, 1976; Bourbon *et al.*, 1987; Bäss *et al.*, 2008; von Carlowitz-Ghori *et al.*, 2011; Lange, 2011; Knolle *et al.*, 2012; Sowman *et al.*, 2012; Hughes *et al.*, 2013; Ott & Jäncke, 2013; Timm *et al.*, 2013, 2014a). It is important to make an explicit distinction between these two types of paradigms. Whereas self-generated speech represents an over-learned natural situation, where the sensory stimulation is unavoidable and isomorphic with the motor act, in button-press studies, the action–sound relationship is an arbitrary association that needs to be learned within the context of the experimental session. Hence, self-generated speech studies are probably closer to the motor control literature, whereas button-press studies seem to share many characteristics with associative learning research. Also, proper controls for confounding factors, such as motor activity, are more difficult to find for speech studies in which one needs to consider many variables, e.g. differences in intensity and distortion of the sound due to bone conduction, than for button-press studies.

Nevertheless, in both paradigms the general experimental setup is the same: participants listen to self-produced sounds in one condition, and to the same sounds that were generated externally in another condition. In order to enable a comparison between the

brain activity elicited by self-generated and externally-generated sounds, the activity due to the motor act has to be considered. Some speech studies include conditions like silent articulation or mental rehearsal. In button-press studies usually an additional motor-control condition is included, in which participants perform the same motor act without producing a sound. It should be noted that this subtraction controls for main effects of the motor act (and therefore of forward models related to other information such as the somatosensory stimulation by pressing a button), but not of interactions between the factors motor act and self-generation vs. external generation. The different conditions can be administered in between-block and within-block designs. The main finding is that the auditory N1 component of the ERP elicited by self-generated sounds is strongly attenuated as compared with the N1 elicited by externally-generated sounds. The N1 amplitudes may differ by a factor of two. This result, in line with the existence of forward models, can be explained by a mechanism suppressing the neural responses when a sound matches the predicted sensorial consequences of the motor act. Such a mechanism may subserve different functions for the auditory and cognitive processes (cf. Schröger *et al.*, 2013). For example, it may aid in flexibly allocating attention, enabling quick reactions to unexpected events, tracking moving sounds, or deciding whether a sound was self-generated or generated by others.

The N1 suppression for self-generated sounds is often accompanied by a suppression of the P2 component (e.g. Sowman *et al.*, 2012). However, it seems that the N1-suppression effect and the P2-suppression effect tap into different predictive functions. In a vocalization study, Chen *et al.* (2012) instructed participants to produce sustained vowel phonation and to generate a pitch-shifted voice signal via a mouse click. They only observed a P2 suppression to self-generated vocals, but no N1 suppression. Another dissociation of N1 and P2 suppression was reported in two studies by Knolle *et al.* (2012, 2013a) investigating the N1-suppression effect in the button-press paradigm in patients with lesions in the cerebellum. Although they found that the N1 suppression was (almost) absent in participants with cerebellar lesion, the P2 suppression was equal to the P2 suppression obtained in control participants. This suggests that the cerebellum contributes to N1 suppression for self-generated sounds and shows a dissociation of N1 and P2 suppression. It can be speculated that the P2 suppression may be a correlate of the conscious predictability of the sounds, whereas the N1 suppression is related to the corollary discharge or efference copy transmitting the predicted sensory consequences of the motor command.

The locus of the attenuation in the processing of self-generated sounds

The comparison and matching of predicted and received stimulation are believed to take place in early sensory areas, where incoming sensory information that progresses up the processing hierarchy meets the sensory predictions estimated from motor commands. Converging evidence from different methodologies supports a sensory locus for the suppression effects. Functional imaging studies have reported reduced activity in the auditory cortices for self-generated sounds compared with passive listening (button press: Blakemore *et al.*, 1998; Melcher *et al.*, 2008; speech: Christoffels *et al.*, 2007; Paus *et al.*, 1996; Wise *et al.*, 1999). However, these imaging studies lack the temporal resolution to differentiate between the direct modulation of feedforward responses and modulations by feedback from higher areas at a later stage of processing.

As discussed above, the most replicated finding in ERP studies is the suppression of the auditory N1, which is known to have sources

in the auditory cortex (Näätänen & Picton, 1987). Magnetoencephalography (MEG) provides a good compromise between spatial and temporal resolution. Using this method, several studies have been able to assess the magnitude of the magnetic counterpart of the N1 directly at its auditory cortex sources, reporting that these respond with reduced magnitude to self-generated sounds compared with passive listening, both when using speech (Numminen & Curio, 1999; Numminen *et al.*, 1999; Curio *et al.*, 2000; Gunji *et al.*, 2001; Ventura *et al.*, 2009; Niziolek *et al.*, 2013; Tian & Poeppel, 2013) and sounds generated via button press (Martikainen *et al.*, 2005; Aliu *et al.*, 2009) as the self-generated stimulation. A few human electrocorticography studies have recorded directly from the auditory cortices with intracranial electrodes in patients undergoing brain surgery. All of these studies have investigated the patterns of activity elicited in auditory areas during overt speech compared with listening to speech stimuli, observing various differences between the two (Creutzfeldt *et al.*, 1989; Crone *et al.*, 2001; Towle *et al.*, 2008; Flinker *et al.*, 2010; Greenlee *et al.*, 2011). For example, Flinker *et al.* (2010) found that multiunit activity recorded from the superior temporal gyrus (STG) was attenuated for spoken compared with heard words, the suppression effect peaking at a latency of around 150 ms. It is worth noting, however, that Flinker *et al.* (2010) found a wide spectrum of responses with varying degrees of suppression across STG spatial locations, including a few sites in the posterior STG showing enhanced rather than suppressed responses to self-generated speech. This finding is consistent with previous animal studies (Müller-Preuss & Ploog, 1981; Eliades & Wang, 2003). In human patients, the pattern of suppressed, non-suppressed, or enhanced responses was consistent across trials in each subject, demonstrating a specific topography of self-speech suppression on the surface of the cortex, which is presumably usually recorded as an averaged suppressed response from scalp electrodes. In any case, human electrocorticography studies have been able to unequivocally demonstrate that at least some auditory-responsive units in sensory auditory areas along the temporal lobe in humans are suppressed during overt speech.

In the study by Flinker *et al.* (2010), the suppression effects on stimulus-evoked responses to self-generated speech were found in the time range of the auditory N1. At least one scalp ERP study, however, has reported earlier suppression effects of stimulus-evoked responses. Baess *et al.* (2009) found an amplitude attenuation of the middle-latency Pa (23–33 ms) and Nb (40–46 ms) components, as well as of the 40 Hz evoked response, for sounds triggered by the participant's button presses. Such early modulation provides additional evidence that the initial feedforward processing of sensory information is attenuated when it matches a prediction. Moreover, middle latency responses constitute the earliest response of the human auditory cortex with involvement of the ascending thalamo-cortical pathways. The results of this study therefore point to the suppression effect being present at very early processing stages; in particular, it might involve the first and second transverse of Heschl's gyrus (Pa) and the STG (Nb).

Is the attenuation of self-generated sounds a specific prediction effect?

The studies comparing brain responses to self-generated and externally-generated sounds have provided converging evidence for reduced responses to self-generated sounds. The effects have often been interpreted in the context of forward models or the predictive coding theory. As the suppression effects take place in early sensory processing stages, they can hardly be (fully) explained as a 'retro'-

spective evaluation of whether or not a sound has been generated by me, which is initiated after the sound has been processed to its full extent. It has to be assumed that the information causing the attenuation rather needs to be provided in a 'pro'-spective or 'pre'-dictive manner.

However, despite the indication that suppression effects should be prospective, there is still the open question of whether the suppression is due to a specific prediction of the expected sensory consequences, or whether the processing of sounds is unspecifically gated during movement. The unspecific gating of sensory responses has been thoroughly described in the physiological literature, e.g, the phenomenon of saccadic suppression (Ross *et al.*, 2001), or the dampened sensitivity on moving body parts (Chapman *et al.*, 1987; Cohen & Starr, 1987; Williams *et al.*, 1998; Williams & Chapman, 2000, 2002). Although this may be a prospective phenomenon, it does not reflect a specific prediction, but an unspecific modulatory mechanism. Before showing evidence that the suppression effects are (at least partly) specific, we will review the state of the art with respect to the unspecificity of self-generation effects.

Several findings seem to indicate that responses to sounds may indeed be unspecifically gated during movement. In particular, some speech studies did not directly compare self-generated speech with the same auditory stimulation during passive listening conditions, but have rather inspected the responsiveness to other (not self-generated) sounds during speech, finding dampened responses to those sounds (e.g. Houde *et al.*, 2002; Kudo *et al.*, 2004). In this case, the dampening could be simply the result of simultaneously receiving two overlapping sound signals (the self-generated speech and the external sounds). However, other studies have corroborated that auditory responses are dampened during other types of (silent) voluntary movements (Hazemann *et al.*, 1975; Tapia *et al.*, 1987; Makeig *et al.*, 1996; Horváth, 2013a,b; Horváth & Burgyán, 2013; Horváth *et al.*, 2012; Horváth, 2014). In particular, Horváth *et al.* (2012) designed an N1-suppression paradigm in which the series of button presses and the series of sounds were uncorrelated. Despite the lack of contingent associations between the two, random coincidences between button presses and sounds resulted in an attenuation of the auditory N1. This suggests that the temporal proximity between motor behavior and sound could be responsible for the attenuation effect, rather than a predictive mechanism anticipating self-generated stimulation.

The unspecific gating of auditory responses during movement may be the result of peripheral or central mechanisms. Regarding peripheral mechanisms, it is known that voluntary movements may cause the coactivation of the stapedius muscle with the task-relevant effector (Galambos & Rupert, 1959; Carmel & Starr, 1963; Salomon & Starr, 1963). This would reduce the signal transmission efficiency in the middle ear, and, as a consequence, result in reduced amplitudes of the ERPs elicited by concurrently presented tones. However, Horváth (2013b) could rule out this alternative explanation, as the stapedius muscle contraction does not apply for frequencies higher than 2 kHz, whereas the N1-suppression effect is not modulated as a function of tone frequency. Moreover, a peripheral locus for the suppression effects would result in a generalized dampening of all auditory responses along the whole auditory processing hierarchy (brainstem, middle latency response, short-latency and long-latency ERPs), but dissociations have been found in the suppression effects across different processing stages. Hence, it seems that central mechanisms must be responsible for the suppression effects observed in neurophysiological responses to self-generated sounds.

A recent *in vivo* intracellular recording study in behaving mice by Schneider *et al.* (2014) has provided direct evidence for a central

mechanism leading to unspecific gating effects in the auditory cortex. They reported that excitatory neurons in the auditory cortex were suppressed by secondary motor cortex neurons innervating the auditory cortex when the mice were moving. Movements reduced the responsiveness of the auditory cortex to sounds overall, supporting the view that the predictions underlying the N1-suppression effect are unspecific, i.e. the system seems to expect some consequence of the motor act, but does not generate a prediction on the specific effect of the motor behavior. This idea is in line with a human ERP study (SanMiguel et al., 2013b), showing that suppression effects for self-generated sounds mainly affect the unspecific component of the auditory N1, one of several subcomponents of the N1 (Näätänen & Picton, 1987). It is believed that the unspecific N1 mainly reflects the orienting response elicited by sound presentation, rather than the specific features of the sounds. Thus, it seems that the attenuation effect is partly driven by the fact that we are not (that much) startled by sounds produced by ourselves. From this perspective, the attenuation reflects an unspecific effect of reducing attentional orientation towards a new sound.

These findings underline the importance of differentiating the specific suppression of predicted sounds from unspecific gating mechanisms. Predictive mechanisms should be reflected in a highly specific suppression. Despite the compelling evidence that the suppression of responses to self-generated sounds is partially accounted for by unspecific gating effects associated with the movement, there is also sufficient evidence that the N1 suppression for self-generated sounds is (at least partly) specific. In general, these studies have assessed to what degree the received input needs to precisely match the prediction to be attenuated, finding that more precise matching results in greater attenuation. For example, in vocalization studies, the N1-suppression effect was found to be reduced, or even reversed, when the auditory feedback was altered (Heinks-Maldonado et al., 2005; Behroozmand et al., 2011). In general, speech studies have found evidence for non-specific attenuation, but the effect also shows specificity for the expected unaltered feedback (McGuire et al., 1996; Hirano et al., 1997; Houde et al., 2002; Hashimoto & Sakai, 2003; Heinks-Maldonado et al., 2005, 2006; Fu et al., 2006). Further evidence can be found in button-press studies. For example, Knolle et al. (2013b) found the N1 suppression to be reduced when a rare pitch-deviating sound rather than a frequent pitch standard sound was generated by the button press. However, in this case, the deviancy per se may have elicited deviance-related effects that overlap the N1 suppression and it is thus difficult to disentangle whether these results reflect a genuine modulation of the N1-suppression effect. More evidence for the hypothesis that the N1-suppression effect includes some specificity of the prediction has been provided for example by Hughes et al. (2013). They found that the N1 (and P2) elicited by tones whose pitch was congruent with the hand-specific prediction was suppressed as compared with tones whose pitch was incongruent with the hand-specific prediction. This suggests that the N1 suppression reflects the effect of a specific effect of the own action. A similar conclusion can be drawn from a study by Ott & Jäncke (2013). They found regular N1-suppression effects for self-generated syllables (but not for piano tones) in musicians and non-musicians. However, when the pitch of the generated tone was not predictable, the N1 suppression turned into an N1 enhancement, again suggesting that the N1-suppression effect reflects a specific prediction about the identity of the sound.

The studies reviewed in the preceding paragraph have addressed to what degree the suppression effects depend on a match of the identity of the sound and the prediction. A related question is whether the timing of the sound presentation also needs to be precisely predicted in order to obtain suppression effects. In a study by Bäss et al. (2008), N1 for self-generated sounds delivered at predictable or unpredictable timing and of predictable or unpredictable frequency were compared with the same sounds when they were not self-generated. The N1 amplitude was attenuated for both predictable and unpredictable frequency and onset. However, the attenuation was largest when both the frequency and onset were predictable. This supports the notion that internal forward model mechanisms can tolerate uncertainties with respect to the frequency of the sound and its onset. As self-generated sounds can interact with the environment and may not always result in exactly the same feedback, we must still be able to recognize them as self-generated. For example, when pressing a piano key we might not know exactly what absolute frequency to expect, or when throwing a stone into a dark well we might not know when it will hit the bottom. In fact, intracranial recording studies have shown that suppression effects can begin before the stimulation starts (Creutzfeldt et al., 1989; Eliades & Wang, 2003; Flinker et al., 2010). The question that is yet to be answered is whether this prestimulus suppression pattern is specific to the units with receptive fields matching the predicted stimulus only (this would be irrefutable proof of a predictive template being active in the auditory cortex in a prospective way), or whether it represents a relatively unspecific gating.

In sum, there appears to be evidence for both relatively unspecific gating of sensation during certain motor acts (most likely of central origin, possibly with a small peripheral contribution), and highly specific prediction-match-dependent attenuation of sensory responses.

Other sources of prediction: auditory-to-auditory and visual-to-auditory attenuation

Are motor predictions a special case of predictive mechanisms, or do they operate via the same mechanisms and therefore share effects with other prediction sources? In the previously mentioned auditory oddball studies, it has not only been shown that prediction mismatch elicits MMN, but also that matching the prediction has been found to elicit a so-called repetition positivity. The repetition positivity is a frontally distributed positive deflection to sounds at around 100–200 ms after sound onset that increases with increasing repetition of standard tones (Haenschel et al., 2005; Cooper et al., 2013). It overlaps in time with the auditory N1 and, given their opposing polarities, it effectively results in a reduction of the N1 to predictable sounds. The repetition positivity is affected by predictability (Costa-Faidella et al., 2011) and has been related to predictive coding theory (Baldeweg, 2007). Predictions coming from other non-auditory sources can also result in suppressed responses, particularly auditory N1. For example, Vroomen & Stekelenburg (2010) report attenuation of the auditory N1 for a crashing sound emitted when a visual stimulus of a ball was seen to move towards and crash into a central square, compared with the same visual–auditory event occurring unpredictably. Moreover, in crossmodal studies that employ dynamic faces producing utterances, it was found that the auditory N1 and P2 is attenuated compared with auditory-only conditions (e.g. van Wassenhove et al., 2005; Ho et al., 2014). This has been interpreted as an effect of prediction, which is supported by the finding that it is (partly) specific to the congruency between the preceding visual content and the auditory content (Jessen & Kotz, 2013; Kokinous et al., 2014).

All in all, predictions stemming from non-motor sources seem to have effects on the stimulus-evoked auditory responses that are similar to those found for motor–sensory associations. Nevertheless,

some findings indicate that some aspects might still be exclusive to motor–sensory predictions. For example, sensory attenuation for self-generated sounds is reduced (but not abolished) when one controls for temporal prediction (Schafer & Marcus, 1973; Aliu *et al.*, 2009; Lange, 2011; Hughes *et al.*, 2012). These studies show that responses are suppressed when the stimulation has a predictable timing, either because it is delivered at a stable rate, or because other cues predict the stimulus onset time. Thus, (at least part of) the suppression effect can be due to the fact that the motor act precisely predicts the timing of the stimulus. However, when one compares self-generated stimuli with stimuli with predictable timing, the responses to self-generated stimuli are still further suppressed. Thus, it seems that suppression effects are usually the largest when stimuli are predicted by a motor act. We can think of at least three reasons why this could be the case. First, as we have seen, motor acts provide unspecific gating effects in addition to the specific prediction effects that other sensory–sensory associations also seem to provide. Second, motor acts might be better predictors, as they provide better information about, e.g. timing. Third, Hughes *et al.* (2012) have suggested that there might be something special about having control over the stimulation. In fact, the suppression effect has been linked to the sense of agency, denoting the experience of causing our own actions and their sensory consequences (Timm *et al.*, 2014b). Similarly to the dissociation between specific and unspecific effects that we have described, some models of agency perception also propose a double-step model for agency (Gallagher, 2005; Synofzik *et al.*, 2008). In the initial step, the motor act would create a temporal window of special processing, which could possibly be related to the unspecific gating mechanisms observed. This initial step is proposed to be directly responsible for the basic, prereflective, feeling of agency. In the second step, the comparison between prediction and input is performed, and only a specific match results in a positive judgment of agency. This second step could thus be linked to the prediction-specific part of the self-generation effects. We could stretch this double-step model even further and speculate that the initial step, accompanied by the unspecific gating, could possibly be the basis to establish motor–sensory associations in the first place (Horváth, 2013a), with the second step helping in the establishment of causality, which would be a rather more reflective process, and which is shared with other sensory–sensory associations. The relationship between self-generation suppression effects and the sense of agency has been empirically supported, e.g. by the findings of Timm *et al.* (2014a). In this study, N1-/P2-suppression effects for self-generated sounds were only found when the button press delivering the sound was voluntarily performed by the participant, and not when the finger movement was elicited involuntarily applying single-pulse transcranial magnetic stimulation over the motor cortex. Additional support for the idea that suppression effects are related to intention is provided by the finding that attenuation to self-generated speech is reduced for spontaneously occurring less prototypical utterances, suggesting that the suppression does not reflect a sensory prediction of the executed motor act, but of the sensory goal (Niziolek *et al.*, 2013).

Mechanisms underlying the attenuation effect

Thus far, we have generally considered that suppression effects result from a prospective mechanism that anticipates the sensory consequences of motor acts, or in general predicts sensory stimulation based on other sensory association rules. There are, however, different ways in which the prediction might result in an attenuation of sensory responses.

In their preactivation account, Waszak *et al.* (2012) claim that the brain preactivates the perceptual representation of the predicted action effect when an action is selected. This preactivation, in turn, may hamper the differentiation between the actual action effect and the preactivation of the predicted effect, resulting in a smaller brain activity (i.e. smaller prediction error). This idea is in general consistent with predictive coding models of perception. Waszak *et al.* (2012), however, consider several models to account for the suppression effect. The 'fatigue' model relates various forms of attenuation to the repetition suppression effect (Grill-Spector *et al.*, 2006), according to which neural responses are reduced when a stimulus is repeated. According to the preactivation account, the anticipation of the predicted sound preactivates neural networks very similar to those that are activated by the real sounds. This results in the respective neural networks being less responsive at the time when the real sound is presented (cf. Colder, 2011; for a general theory of emulation as a principle for cognition). Alternatively, the 'sharpening' model claims that a repeated sound or a sound that has been predicted involving the emulation of the respective networks can be represented more sparsely, because only those neurons encoding the characteristic features of the sounds are maintained in the representation, whereas those that are not so characteristic of the sound are not maintained in the representation (Giard *et al.*, 2000; Fritz *et al.*, 2007). Hence, the activation of the sound representation elicits less neural activity, because it is sharper. This idea has also been proposed in regard to visual object representation (Desimone, 1996; Gruber & Müller, 2005; Kok *et al.*, 2012a). Finally, it has often been proposed that the suppression effects for self-generated sounds might be related to attentional mechanisms (Brown *et al.*, 2013; Saupe *et al.*, 2013; Timm *et al.*, 2013). We will discuss these attentional interpretations in Section Studies on the relation between auditory attention and prediction, where interactions between prediction and attention mechanisms will be considered.

Prediction effects uncovered in the absence of stimulation: omission paradigms

Omission responses as a tool to uncover prediction templates

The fact that neurophysiological responses to prediction-matching and prediction-mismatching stimulation differ from each other clearly suggests that prediction is taking place. However, this is only indirect evidence for the generation of predictions. One can only assume that these responses differ because a hypothetical prediction must have been generated. However, the formulation of the prediction itself cannot be accessed in these types of paradigms. This issue in fact leaves the door open for other potential explanations that do not require the explicit formulation of a prediction to explain such effects. For example, one proposal has explained a great deal of match and mismatch effects based solely on the differential refractoriness of the neural populations responding to predicted and unpredicted sounds (May & Tiitinen, 2010). Definite proof for prospective prediction mechanisms can only be found in studies that are able to uncover the purely endogenous neural activity associated with the generation of predictions. This can be accomplished with omission paradigms.

The same models that can explain the attenuation of sensory responses to predicted stimuli as a result of a match between predicted and received input also make a somewhat counter-intuitive prediction: that the omission of a predicted stimulus should elicit a prediction error (sensory) response (Friston & Kiebel, 2009; Arnal & Giraud, 2012). In the particular case of self-generated stimuli that

occupies us here, the reafference principle (von Holst & Mittelstaedt, 1950) and ideomotor theories (Hommel *et al.*, 2001) suggest that, together with the motor act, the associated sensory consequences of this act are activated in the sensory cortices. When the sensory input consequence of the motor act arrives at the sensory cortices, a comparison is made between the predicted and received stimulation. As we have seen thus far, if a match arises, the prediction error, and thus the stimulus-evoked responses, is minimized. If a mismatch occurs, a larger prediction error response ensues, which is transmitted up the sensory hierarchy. In this case, the informational content of the prediction error is in fact composed of two parts: that part of the sensory input that is encountered but was not predicted, and that part of the sensory input that was predicted but is in fact absent (Fig. 2). This second part is what makes omission paradigms particularly useful. In the case of an omission, the prediction error consists solely of the prediction that was not met by the (lack of) input, i.e. the prediction error should amount to an exact copy of the efference copy delivered with the motor act. Omitting the expected consequences of a motor act thus provides us with a useful tool to observe the efference copy, and in consequence to shed some light into the neural code of sensory predictions.

Omissions of self-generated stimuli

Following the logic that omissions of expected stimuli reflect the neural code of sensory predictions, SanMiguel *et al.* (2013c) asked participants to produce sounds by pressing a button every 800 ms. ERPs to the self-generated sounds and to the rare omissions of those sounds were recorded. When self-generated sounds were omitted with a low probability (12%), omissions elicited an electrophysio-
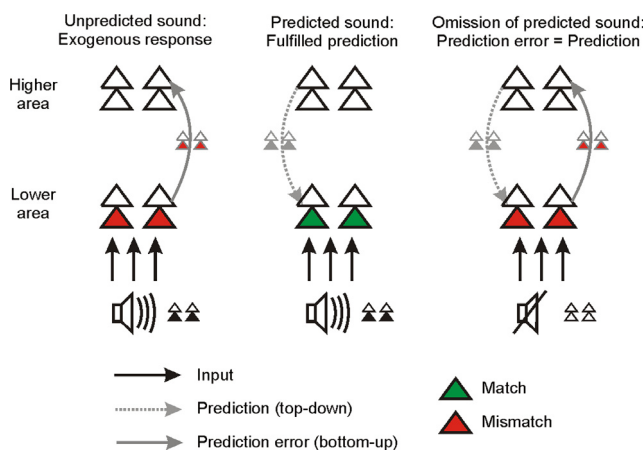

FIG. 2. Illustration of the generation of auditory ERPs/ERP effects in the case of the occurrence of an unpredicted sound (left), a fully predicted sound (middle), and the omission of a strongly predicted sound (right). Left: the sensorial input arriving in the lower area is compared with the prediction sent down from the higher area. As there is no prediction for the arriving sound, the prediction error contains the complete input to the lower area. Middle: the sensorial input is met by a perfect prediction sent down from the higher area (e.g. when the sound is self-generated). As the input and prediction match perfectly, the prediction error is zero and no prediction error is sent to the higher area. The sensory input can be fully accounted for by the inferred causes (the model of predicted states) and the sound is perceived. Right: a prediction is sent down, but no input arrives at the lower area (as in the omission paradigm). The prediction error is then identical to the prediction. Thus, omission responses are a means to make the prediction (which is not easily accessible to the ERP technique) visible. This is a simplified version of figures published by Arnal & Giraud (2012) and SanMiguel *et al.* (2013c).

logical response resembling, in morphology, time-course and sources, the auditory N1 elicited by the sounds themselves (Fig. 3D). However, when button presses only generated sounds on 50% of the trials, no omission response was elicited. This result provides compelling evidence in favor of prospective prediction mechanisms. Additionally, it supports the idea that prediction operates through the top-down modulation of stimulus-specific templates in the sensory cortex. To further test this idea, SanMiguel *et al.* (2013a) compared a condition in which the button presses always produced the same sound with a condition in which the identity of the sound produced by the button press was random, and thus unpredictable. The results reinforced the idea that the neural mechanism behind prediction effects relies on the modulation of stimulus-specific sensory templates. Omission responses were only obtained when the specific physical characteristics of the sound could be predicted, and not when only the expectation of a sound was possible, but its identity was unknown. To our knowledge, ERP responses elicited by omissions of the expected sensory consequences of a button press have been reported in another two studies, also when the missing stimulus was in the visual modality (Nittono, 2005; Nittono & Sakata, 2009). Unfortunately these studies focused their analysis only on late responses (~200 ms and later).

However, the findings of SanMiguel *et al.* (2013a,c) are consistent with the findings of two imaging studies showing that, after a particular motor act has been associated with the delivery of a sensory stimulus, the execution of the motor act alone is later sufficient to induce activation in brain areas responsive to the associated sensory consequence (Kühn & Brass, 2010; Kühn *et al.*, 2010). In the first of these studies (Kühn *et al.*, 2010), left and right button presses were respectively associated with the presentation of faces or houses. The authors showed that, after this association had been established, the button presses alone triggered the activation of either the face-selective (FFA, Fusiform Face Area) or place-selective (PPA, Parahippocampal Place Area) brain area, consistent with the action–effect mapping acquired previously. In the second study (Kühn & Brass, 2010), the authors replicated similar effects when button presses were associated with tones; after an initial acquisition phase, the action alone was able to produce activation in the auditory cortex. Further, the effects were also present when the decision to withhold an action was associated with the presentation of the sounds. Hence, not only motor acts but also voluntary non-actions can trigger predictive activation of their associated sensory consequences.

In sum, self-generation paradigms using stimulus omissions have been particularly successful in providing empirical evidence for the brain responses that could be predicted based on predictive coding models, the reafference principle as well as ideomotor theory. The omission response should be an exact copy of the sensory prediction, and thus resemble an evoked sensory response to the predicted stimulus. However, the approach of using motor–sensory associations to uncover prediction-related activity with omissions is relatively rare. As we will see in the next section, the possibility of obtaining omission responses has been explored before with different types of paradigms in the sensory–sensory domain, and these studies have yielded more inconsistent results, showing different types of omission responses. Essentially, only a few studies using sensory–sensory rules have shown results that are similar to those of the self-generation omission studies. Omitting tokens from a sequence of predictable sounds has more frequently resulted in MMN-like or other N2-like responses, and/or P3 responses similar to those obtained for unpredicted sounds in oddball-type paradigms, i.e. most omission studies that use sensory–sensory associations have obtained slightly later error responses time-locked to the
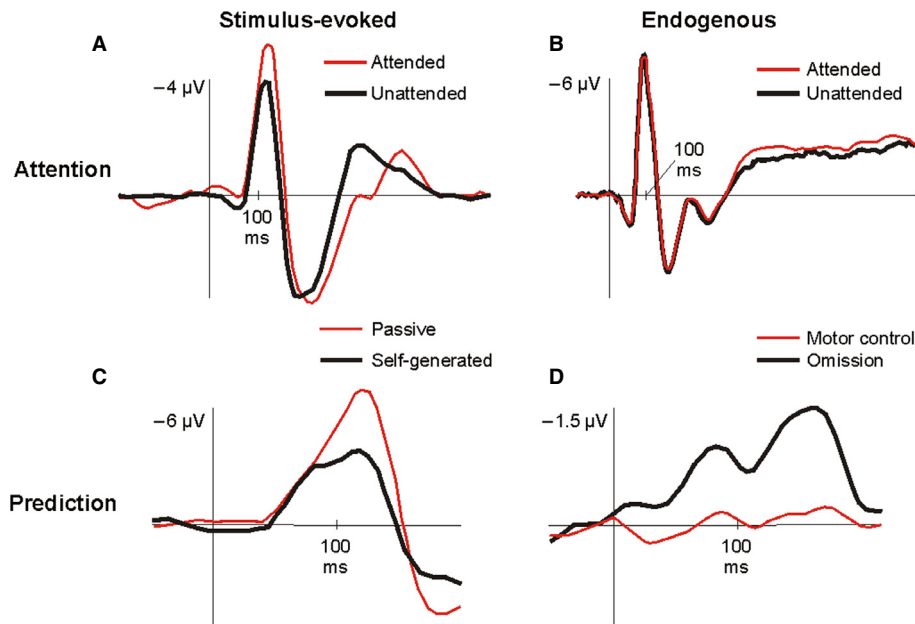
FIG. 3. The main effects of attention and prediction on the auditory N1. (A) Attended sounds result in an increased N1 as compared with the same sounds when unattended (modified from Saupe *et al.*, 2013). It is proposed that attention increases the gain of the prediction error, reflected in the N1. (B) In the silent interval between two sounds or between the cue and the sound, attention can elicit (small) sustained activity (modified from Seiss *et al.*, 2007). This may reflect the attentional control/attentional trace. (C) Self-generated sounds reveal smaller N1 than the same sounds generated by the computer (modified from Timm *et al.*, 2013). Due to a reasonably good model, the prediction accounts for part of the sensorial input, resulting in smaller N1 (prediction error). (D) When pressing a button that usually generates a sound, the unpredicted omission of the sound generates N1-like ERPs, which is not the case for the same button press when no sound is predicted (modified from SanMiguel *et al.*, 2013c). The omission response directly reflects the prediction error, but it indirectly yields information about the sensory template represented by the prediction. Please note that the scale on the y-axes and x-axes varies between the four plots.

omissions that do not resemble (or resemble less) the predicted stimulus.

Omission responses in sensory–sensory rules

A few studies have reported omission responses that are similar to the results obtained in the self-generation studies described above. These studies have been able to show omission responses that are in their morphology and sources similar to the responses to the omitted sound, when it was actually presented.

Some early ERP studies, mostly using oddball-type paradigms, described what they termed 'emitted potentials', which shared some characteristics with evoked potentials related to the absent stimuli that were expected (Weinberg *et al.*, 1974; Simson *et al.*, 1976; McCallum, 1980). For example, Simson *et al.* (1976) presented visual or auditory stimuli in different blocks at a rate of one stimulus per second. For both modalities, when a stimulus was omitted, they found a negative potential whose onset was virtually synchronous with the missing stimulus and peaked at around 275 (visual) or 230 (auditory) ms. The topography of the negative omission response was dependent on the modality, showing some sensory specificity. Simson *et al.* (1976) suggested that this omission response could in part reflect anticipatory events activated by an internal representation of the temporal rhythm of stimulus presentation. Using tone pairs, Weinberg *et al.* (1974) also showed an ERP response to the omission of the second tone that was similar to the ERP response to the first tone in the pair. Furthermore, McCallum (1980) reported an initial negative peak in the omitted stimulus potential that did not differ significantly in latency or amplitude from the auditory N1 to the auditory stimuli presented in the sequence.

In the more recent literature, an MEG study by Raij *et al.* (1997) found that omitting sounds from a regularly repeating sound

sequence evoked responses over temporal and frontal areas, with the peak amplitude at 145–195 ms. These could be accurately explained by the N100 dipole source estimated for sound responses. Not many current ERP studies have been able to find similar responses. In a notable exception, Bendixen *et al.* (2009) presented tones of varying sound frequencies. Crucially, tones were presented in pairs, with the second tone always being a repetition of the previous sound frequency. Thus, the first tone of each pair was unpredictable, but the second tone was always fully predictable. The authors occasionally omitted either the first or the second tone out of some pairs. When a predictable tone was omitted, an ERP response time-locked to the absence of sound was obtained. The omission response was indistinguishable from the ERP response elicited by the tones until up to 50 ms after the omission. This was not true for omissions of the first tone of the pairs (unpredictable ones), which did not elicit such a response.

Studies that used musical material have been particularly successful in finding N1-like omission responses. The implicit structure of music provides a particular situation in which very strong auditory expectations arise. Therefore, if a sequence is interrupted early, clear omission responses can be found. In an exemplar ERP study, Janata (2001) asked participants to mentally complete melodic ascending and descending fragments in the appropriate tempo, and found that, after the sequence was interrupted, the imagined note elicited an auditory N1-like component that resembled the N1 to heard notes. In an MEG study, Otsuka *et al.* (2008) found a similar response, and reported that it shared sources in the auditory cortex with the N1 elicited in response to the real auditory stimuli. Nevertheless, in these last two studies, omission responses are likely to reflect effortful auditory imagery.

The origin of N1-like omission responses in the auditory cortex has been corroborated in a study by Hughes *et al.* (2001) with intracerebral

recordings from the temporoparietal junction. Hughes *et al.* (2001) presented pairs of tones separated by 100 ms and omitted the second tone of the pair in 20% of the trials. They were able to record omission responses peaking at 100 ms from electrodes that were also responsive to tones. The omission responses had a topographic distribution indicative of source generators in or near the posterior Sylvian fissure (auditory association cortex), but not in the A1. It is of note that this study was able to find two different response patterns from separate electrodes, i.e. those that responded to both tones and omissions, and those that responded only to omissions; perhaps providing a rare piece of direct evidence for the existence of representational and error units, as assumed in predictive coding models.

Most ERP studies on omitted sounds come from the MMN field. Using typical oddball paradigms, two early studies found that, when stimulus omissions are employed as the deviant events, an MMN is also elicited (Nordby *et al.*, 1994; Tervaniemi *et al.*, 1994). Similarly to the classic MMN, the omission MMN peaks at 120 ms and shows a frontocentral maximum with polarity inversion in mastoids (Yabe *et al.*, 1997). Yabe *et al.* (1998) localized the omission MMN using MEG to the supratemporal areas, and specified that its generator is different from that of the sound-elicited auditory N1 response. Sources for the omission MMN in the right planum temporale and temporoparietal junction have been corroborated using functional magnetic resonance imaging (fMRI) (Mustovic *et al.*, 2003). However, there is one critical constraint for the elicitation of the omission MMN that differentiates it from the classic MMN elicited by deviant sounds. In order for a sound omission to elicit an MMN response, the time delay between the sounds composing the predictable sequence needs to be less than around 150 ms. Yabe *et al.* (1997) termed this the 'time window of integration', and interpreted that separate sounds that are presented within this time window are in fact encoded by the system as a single object. From this point of view, the omission MMN is a response that denotes the detection of a missing part of an object.

Partial stimulus omissions have been used in a recent study using linguistic material. In two experiments, Bendixen *et al.* (2014) have shown omission MMN effects when omitting a part of a word. In the separate experiments, predictions regarding the word ending were induced either via a simple repetition rule, or using the semantic context of the sentence. In both experiments, they found an omission MMN in the 125–165 ms time window. Moreover, when the expectations were semantic, and participants were actively engaged in listening to the sentences, the partial word omission also elicited a later omission N2 response (210–250 ms). Similar omission N2 responses have been obtained in other studies, usually when employing active tasks, or when the omissions are used as targets (e.g. Busse & Woldorff, 2003; Otsuka *et al.*, 2008).

Finally, several ERP studies, particularly the very early studies, have investigated P3 responses and other related late positive components elicited by stimulus omissions. Early studies were often concerned with investigating whether these omission responses reflected the same or different processes to the P3 elicited by rare deviants and by target stimuli (Sutton *et al.*, 1967; Picton & Hillyard, 1974; Ruchkin *et al.*, 1981; Stapleton *et al.*, 1987). For example, Sutton *et al.* (1967) showed that the P3 response was elicited not only when a click was present, but also when it was absent, showing that P3 is concerned with the content of the information. In general, P3-like responses related to stimulus omissions have also been more often investigated in the context of active tasks (see, e.g. Janata, 2001; Busse & Woldorff, 2003; Penney, 2004).

A hierarchy of predictions and omission responses

In sum, omission studies using auditory rules have shown a variety of different neurophysiological responses associated with the missing stimulus, including N1-like responses, omission MMN and other N2-like responses, and P3 responses and other similar late positive components. However, in fact, in many omission studies, not only one response, but rather a sequence including several of these responses has been observed (e.g. Raij *et al.*, 1997). This is also true for self-generation omission studies. For example, in a study by SanMiguel *et al.* (2013a), the initial N1-like omission responses were followed by an N2 and a P3 response. Moreover, all of these responses were cancelled when the button presses were not predictive of the sounds. These various omission responses have been recently interpreted in terms of a hierarchy of predictions (Wacongne *et al.*, 2011), with each successive response representing prediction error at each successive processing level along the hierarchy. Later responses, which would correspond to prediction error in higher levels of the processing hierarchy, would also incrementally incorporate activity related to the updating or correction of the internal model to account for the prediction error that is being received from the lower levels.

To test this idea of a hierarchy of predictions and corresponding prediction error signals, Wacongne *et al.* (2011) presented auditory sequences that contained two types of rules, local and global. The assumption was that global rules should be encoded at higher levels of the hierarchy than local rules, as they require a certain level of abstraction and integration of more stimuli across time. In particular, they presented five-tone sequences in which the first four tones were always exact repetitions, and the last tone could either be another repetition, or a different tone (xxxxx or xxxxy). They inspected responses to the last tone of the sequence. The local rule is based on the repetition of the tone in the first four sequence positions, and thus this rule would always predict another presentation of the same tone. The global rule was manipulated across blocks, adjusting the relative probability of xxxxx and xxxxy sequences within each block. In blocks in which the xxxxx sequence was more probable, the global rule would predict the sequence to end in x, agreeing with the local rule. However, in blocks in which the xxxxy sequence was more probable, the global rule would predict the sequence to end in y, conflicting with the prediction derived from the local rule. Violations of the local rule (sequences ending in y) elicited a classic MMN response. The comparison between rare sequences and frequent sequences yielded differences on a later time window, corresponding to a P3 response. Crucially, local MMN effects were modulated by the higher-order expectation of the deviance, i.e. MMN responses were reduced in blocks in which xxxxy sequences were more probable. Importantly for the present discussion, in 10% of the sequences, the fifth tone was omitted. Unexpected omissions in five-tone sequence blocks evoked a response in both electroencephalography and MEG recordings peaking at 100 ms after the onset of the omitted tone, followed by a later P3 effect. The sources of the early effect were similar to those of the tone-evoked responses, in the vicinity of Heschl. The early effects were also modulated by the global rule, being larger in xxxxy blocks than in xxxxx blocks. All of these findings can be explained from the perspective of a hierarchy of representations, with higher-order representations generating predictions about lower level representations, attempting to explain away the prediction error that is calculated at each level and transmitted through forward connections. The interaction between local and global deviance effects in the MMN time window is consistent with the view that, in the xxxxy blocks, a

second level prediction can be used to partially cancel out the first-order error response. If we consider that omission responses uncover the underlying predictions, this second level prediction would also be responsible for larger omission responses in the same time window in xxxxy blocks, as this prediction would only be present in these blocks, increasing the magnitude of prediction-related activity.

These findings suggest that stimulus omissions can in fact be used to uncover the current internal model that is maintained at each of the levels along the hierarchy, with each subsequent response in the sequence of omission responses that have been reported in different studies reflecting the internal model at each subsequent level of the hierarchy. Thus, whether or not each of these responses is elicited by the omission in a particular type of paradigm can also inform about the complexity of the models that can be maintained at each of the levels reflected by each of the responses. Some tentative conclusions can be extracted by inspecting the behavior of omission responses across the studies reviewed. Early omission responses that more directly resemble the exogenous sensory responses have proven quite difficult to find. Probably, in order to observe these responses, a very clear time-locking reference for the prediction is needed. Additionally, they seem to reflect highly specific and explicit predictions, which require a very high confidence in the model to be generated, and require that the model can formulate a prediction regarding the exact physical characteristics of the predicted stimulus. Hence, these responses seem to reflect the activity of a very literal model. The subsequent step, reflected in the MMN, may be slightly more 'liberal' but still has some very specific time constraints (i.e. it is only elicited within the time window of integration). These limitations can be explained taking into consideration the type of internal model that it reflects. The MMN appears to reflect a model of local auditory rules, which are most likely encoded in transition probabilities. Hence the time window of integration most likely reflects the time constraints within which these transitions are considered. Subsequent N2 and P3 responses appear to reflect violations of higher level, more global rules. These responses are also usually associated with the conscious perception of the omission, or with active tasks. Hence, they probably reflect models of increasing complexity, which incorporate more aspects of the situation, e.g. like the current goals.

Relationship between N1-suppression effects and omission responses: two sides of a coin?

To sum up, in this section on prediction, we have reviewed on the one hand evidence that responses to prediction-matching stimuli are attenuated, and on the other hand that the underlying prediction-related activity can be uncovered with omission paradigms. These two neurophysiological effects have been mostly studied in separate studies. However, according to the theoretical models to which we have been referring to explain these effects, suppression effects and omission responses should in fact be two sides of the same coin, i.e. theoretically, it is exactly the prediction-related activity uncovered with omission responses that is responsible for the suppression of responses to prediction-matching stimuli. Is there any direct evidence contrasting these two effects showing that they are indeed linearly dependent? In self-generation studies, a glance at the literature quickly reveals that classic self-generation effects are typically much larger than omission responses. However, as we have reviewed, these suppression effects have typically included specific and unspecific contributions. It is more likely that omission responses might be equivalent to the specific part of these suppression effects. To our knowledge, no study has attempted to directly compare these two effects. In studies using sensory–sensory rules, the question of

whether MMN responses are the inverse of the N1-attenuation effects has been raised before (repetition positivity studies, e.g. Haenschel *et al.*, 2005). However, these studies usually considered the stimulus-evoked MMN, and not the omission MMN.

At least one study has found a correlation between suppression effects and omission responses. Todorovic *et al.* (2011) studied prediction effects with tone pairs. In every trial either a single tone, or a tone pair, could be presented with a time delay of 500 ms between tones. In different blocks, the probability of the tone being repeated was either 75% or 25%. The event-related fields and time–frequency responses elicited by expected and unexpected repetitions of the tones, as well as expected and unexpected omissions of the second tone, were compared. Responses to expected tones were in general attenuated. In particular, the evoked activity in the 100–150 ms following tone onset was attenuated, accompanied by reduced power earlier in low-frequency bands (0–350 ms; 5–9 Hz) and later in the gamma band (200–300 ms; 80–95 Hz). When compared with expected tone omissions, unexpected tone omissions elicited an evoked field in the 100–150 ms time window, as well as a later gamma band response (60–75 Hz) in the 200–400 ms time window. Importantly, the authors were able to show that these two effects were correlated, i.e. individuals who presented larger attenuation effects for predicted repetitions also presented larger responses to unexpected omissions.

Disentangling prediction from prediction error

One final issue to be addressed in this section is whether prediction and prediction error responses can be fully disentangled at all. According to predictive coding, omission responses provide access to the representation of the prediction, as in this case the prediction error is an exact copy of the prediction. However, omission responses are still prediction error responses propagating in a feedforward manner along the hierarchy and not the neurophysiological signals encoding the prediction, which propagate backwards along the hierarchy. Although it is acknowledged that top-down signals may make a small contribution to scalp-recorded event-related responses, ERPs are believed to be, to a large extent, generated in superficial pyramidal cells that transmit prediction error in a feedforward fashion (Bastos *et al.*, 2012; Jackson & Bolger, 2014). Hence, looking for the activity directly reflecting the backwards transmission of predictions in ERPs might possibly be a lost cause.

However, a distinction between prediction error and prediction responses may be feasible in time–frequency measures (cf. Arnal *et al.*, 2011; Arnal & Giraud, 2012). It is proposed that gamma band responses may directly reflect prediction error, as gamma is supposedly generated in superficial neural layers that transmit information from lower to higher cortical areas. Deep pyramidal cells, however, appear to be involved in the transmission of information backwards throughout the hierarchy (i.e. from higher to lower levels), and it is suggested that they also operate in slower frequency rhythms. Therefore, the differentiation between prediction and prediction error can be perhaps more easily studied with time–frequency analyses.

Cognitive psychophysiology of auditory attention

One important function of attention is the selection of information at various levels of the information-processing system. Selective attention is involved in the control of the flow of information that establishes and modifies (conscious and unconscious) mental representations. Although attentional selection can take place for any type of information encoding contents of action, emotion, motivation,

memory, perception and the like, the majority of psychophysiological studies focused on selection for perception, the idea being that some hypothetical agency has to be postulated that controls which parts of the manifold of informational input are processed into the few perceptual objects that are available for consciousness at a given moment in time. A classic distinction of how auditory information can be selected is whether the selection happens voluntarily (endogeneously, top-down driven) by the intention of the listener or whether it happens involuntarily (exogenously, bottom-up driven) without the explicit intention of the subject (James, 1890; Näätänen *et al.*, 2002). We have introduced involuntary attention in Section Cognitive psychophysiology of auditory prediction as a special case of a sound that mismatches a prediction so strongly that it calls for our attention even when it has been ignored before. The present section will therefore deal with voluntary attention.

Effects of attention on brain activity

The majority of selective attention studies from cognitive psychophysiology are concerned about the effect that attention exerts on the processing of the attended as compared with the unattended sounds. In these studies, attention was manipulated to the to-be-selected feature (a particular spatial location or a particular pitch) either for a longer period ('sustained' attention) or for a shorter period in which attention is manipulated via a cue on a trial-by-trial basis ('transient' attention). Based on behavioral research, in some studies attention was oriented towards the to-be-selected feature by increasing the probability that the forthcoming target will also be of that feature, and vice versa, attention was directed away (or reduced) from the to-be-selected feature by decreasing the probability that the forthcoming target will also be of that feature value. However, a behavioral response was still required to those stimuli. This was due to the need to collect behavioral data from highly attended and less attended stimuli in order to be able to assess attention effects. For example, in the traditional Posner spatial orienting task (cf. Posner, 1980; see also Posner, 2014), attention can be directed by an arrow cue that indicates the side (left or right) on which an upcoming target is most likely to appear. Responses to validly (correctly) cued stimuli are compared with responses to invalidly cued stimuli. As the constraint of collecting behavioral responses to both attended and unattended targets does not apply to psychophysiological studies, filtering paradigms can be used, in which attention is exclusively allocated to the to-be-selected feature, and stimuli not comprising this feature are completely task-irrelevant and can be ignored. Therefore, attention can be manipulated solely via instruction or the intention of the participant to select the one feature and to ignore the other. Again, this has been performed in sustained attention situations (dichotic listening paradigms) or in trial-by-trial cueing studies (Posner-type cueing). Cognitive psychophysiology not only investigated the brain activity of to-be-attended and to-be-ignored sounds, but studies also exist that investigate the effects of attention before the to-be-attended sound is presented. This is usually performed by analyzing brain activity in sustained attention situations in the silent interval between sounds or in transient attention situations in the interval between the cue and the sound. In the following we will present some main findings of this research.

Attention effects on the processing of the sound

One of the most important findings in this field was published by Hillyard *et al.* (1973). In a dichotic listening paradigm, they showed that the amplitude of the auditory N1 is enhanced for sounds that are presented at the task-relevant, to-be-attended location (Fig. 3A). They presented short tones at a fast pace of ca. two tones per second randomly to either the left or the right ear. In one half of the experimental blocks, listeners were instructed to attend to the tones presented in the left ear, and in the other half to attend to the right-ear tones, and to discriminate the higher pitch tones from the regular pitch (standard) tones at that ear. By comparing the ERPs elicited by the standard tones presented at the attended ear and the ERPs elicited by the unattended tones, they found that the ERP for attended tones was enhanced in the N1 range. As the tones were physically identical and as the standard tones did not require a behavioral response, the difference in the ERP amplitude could be explained as an attentional effect. Hillyard *et al.* (1973) interpreted this attention effect as an enhancement of the auditory N1 component, which could be explained by an amplification of the afferent activity caused by the attended tones (attentional gain theory).

Näätänen *et al.* (1978) yielded similar results and interpreted this attention effect in a different way. They suggested that, in addition to the exogenously driven processes reflected in the N1, an additional, temporally overlapping attention-specific endogenous so-called processing negativity (PN) takes place. This PN indicates an attentional stimulus selection that is based on the comparison of the current sensory input against the memory representation template (the attentional trace) of the stimulus to be attended. As long as the comparison process yields a match between the sound features characterizing the to-be-attended (task-relevant) sound and the actual sound, the sound is processed to a larger extent than when no match occurs. When the template and the current sound no longer match, the PN is terminated and the ERPs elicited by relevant and irrelevant sounds start to differ, which is expressed as a negative difference. Thus, in a selective listening situation, each sound elicits a PN as long as the sound cannot be classified as irrelevant; the more similar that the relevant and irrelevant sounds are, the longer it takes before the PN to the irrelevant sound can be terminated and the negative difference emerges (Hansen & Hillyard, 1980; Alho *et al.*, 1987a,b). This conflicting interpretation of auditory selective attention on the auditory ERP provided by the attentional gain theory (Hillyard) and the attentional trace theory (Näätänen) triggered a long-lasting debate and plenty of interesting experiments. Meanwhile there is some agreement in the sense that there seem to exist two selective-attention mechanisms: one causes an enhancement of the N1 and the other generates the PN (Hansen & Hillyard, 1980; Näätänen *et al.*, 2002). In fact, the PN *per se* has been shown to consist of two (Näätänen, 1982) or even three (Giard *et al.*, 1988) subcomponents that can be distinguished in time and neural generator structure. The earlier PN occurring somewhere between 50 and 200 ms seems to reflect the matching process occurring at a secondary auditory cortex level, whereas the later PN occurring beyond 300 ms could possibly indicate the frontal lobe executive control of the attentional trace (cf. Näätänen, 1990).

In auditory trial-by-trial cueing tasks, in which attention was oriented in a probabilistic manner (a cue indicated the feature of the next target with a validity of 80%, but invalidly cued targets were also behaviorally relevant) or via instruction (participants should only respond to the cued target), enhanced negativities for validly cued sounds on the 100–200 ms interval have been obtained (Schröger, 1993, 1994; Schröger & Eimer, 1993, 1996). When cueing for the pitch of a sound, attention effects were frontally distributed; cueing for spatial location evoked additional parietally distributed attention effects.

In addition to the processing benefits observed for attended sounds, either in the form of gain enhancements or PNs, behavioral research has yielded evidence for a third mechanism consisting of the selective inhibition of processing of to-be-ignored sounds (Treisman, 1964). ERP evidence for this has been provided in spatial trial-by-trial cueing studies including 'neutral' or 'control' trials (Schröger & Eimer, 1997; Bidet-Caulet *et al.*, 2010). Comparing the ERPs elicited by sounds in attend vs. neutral trials yields the 'benefits' of attentional orienting, and comparing the ERPs in ignored sound vs. neutral trials yields the 'costs' of attentional orienting. In fact, both costs and benefits of similar size were obtained in these studies. However, costs and benefits can hardly be interpreted in terms of physiological terms of suppression or excitation at a cellular level. One reason for this is that it is difficult to calibrate attention in the neutral trials to zero level. Instead, the terms inhibition/suppression and facilitation/excitation are merely useful in the context of mental operations described at a cognitive level of information processing.

It should be mentioned that not only ERPs but also other electroencephalography measures are sensitive to auditory selective attention. For example, the auditory evoked gamma band (40 Hz) activity and the auditory steady-state potential were found to be enhanced for attended sounds (Tiitinen *et al.*, 1993; Debener *et al.*, 2003; Ross *et al.*, 2005; Saupe *et al.*, 2009), although not all studies found such an effect (Linden *et al.*, 1987). Even (microsaccadic) eye movements, which are sensitive to auditory selective attention (Valsecchi & Turatto, 2009; Yuval-Greenberg & Deouell, 2011), indicate that the brain can differentiate target from non-target sounds quite fast, already at 80–100 ms after sound onset (Widmann *et al.*, 2014).

Attention effects before the presentation of the sound

Some studies have investigated the brain activity in either sustained attention or trial-by-trial cueing situations before a sound occurred. In electroencephalography studies, in these situations, slow sustained waveforms are what is mostly seen, e.g. in the cue–target interval (Harter *et al.*, 1989). In a visual study, Harter *et al.* (1989) used arrow cues that indicated the location at which targets had to be detected. The comparison of right-cue vs. left-cue ERPs revealed a negative deflection over the hemisphere contralateral to the attention-directing cue between 200 and 400 ms relative to cue onset, termed the early directing-attention negativity. A subsequent response between 500 and 700 ms after cue onset consisted of a positivity over the hemisphere contralateral to the attention-directing cue (the late directing-attention positivity).

An auditory instance of such sustained waveforms is also found in spatial cueing studies. An anterior directing-attention negativity (Fig. 3B) is induced contralateral to the cued side in the 300–500 ms postcue intervals, which is followed by a contralateral late directing-attention positivity occurring at around 600–900 ms after the cue (Seiss *et al.*, 2007). Such effects in the cue–target interval have been interpreted as attentional control processes engaged in the preparatory modulation of the excitability of the respective sensory areas (cf., e.g. Foxe *et al.*, 2005; Talsma *et al.*, 2008). This type of preparation can be related to the first stage of a four-stage model of (visual) selective attention (Eimer, 2014), which may in principle also apply for the auditory modality.

The locus of attention effects

Many imaging studies have revealed effects of auditory attention in the auditory cortex, but other subcortical and cortical brain structures have been identified that are also modulated by attention (for reviews see, e.g. Giard *et al.*, 2000; Fritz *et al.*, 2007). In a recent meta-analysis including 115 fMRI studies, Alho *et al.* (2014) identified effects of actively vs. passively listening to pitch in the middle STG. Selectively attending to low or high tones delivered in a dichotic listening situation revealed effects centered in the STG or superior temporal sulcus. When attending to space, effects were centered in the posterior STG or planum temporale. In the right-hemisphere auditory cortex, the median locus of spatial attention-related modulations was in the superior temporal sulcus, significantly inferior to the median locus for passive spatial processing. Attention effects to speech were localized in the mid-STG/superior temporal sulcus region, which was also activated by speech and voice sounds when the sounds were not attended.

It seems that the attentional modulation of stimulus-evoked responses happens in those areas that are sensitive to the respective sounds even in the absence of attention. This is compatible with both Hillyard's attentional gain theory and Näätänen's attentional trace theory. However, as auditory attention effects have not only been localized in auditory cortical areas [effects have been found from the auditory periphery (as demonstrated by an attentional modulation of the evoked otoacoustic emissions, see Giard *et al.*, 1994) to cortical frontoparietal networks (Bidet-Caulet & Bertrand, 2005) and frontal-cortex-to-auditory-cortex networks (Fritz *et al.*, 2010)], it seems that attention to sounds can be dynamically allocated wherever the respective representation is of relevance. Giard *et al.* (2000) suggest that attention acts as a bandpass filter that tunes responses to task-relevant feature values. The higher the attentional demands, the sharper, more efficient and earlier is the filter tuning.

The anatomical pattern of activation underlying the sustained attention effects occurring before the onset of the sound has been investigated in imaging studies. These studies showed activity in brain regions associated with the control of attention, such as frontal and parietal networks, but also revealed that the relevant sensory cortices (i.e. the auditory cortex in this case) were activated (Jäncke *et al.*, 1999; Wu *et al.*, 2007). This activation occurred before the presentation of the sound, i.e. the mere expectation of an auditory input was found to activate the auditory cortex. Moreover, this activation (at least partly) represented the features of the to-be-attended input, as the activation was larger contralateral to the predicted side of presentation (Voisin *et al.*, 2006).

Mechanisms of auditory attention

In the previous sections, we discussed evidence for two types of attention mechanisms: (i) those causing a modulation of stimulus-evoked responses (which can consist of the enhanced amplitude of sensory responses to attended sounds and of the suppressed amplitude to to-be-ignored sounds, particularly the N1 and the early PN subcomponent, named negative difference early) reflecting an attentional gain of the bottom-up flow of information, and (ii) those generating activity in the absence of stimulation, which possibly reflect top-down attentional templates. The attentional trace that serves as a template for the to-be-selected sound features (e.g. Näätänen *et al.*, 1978, 2002; Näätänen, 1990) can be regarded as a top-down attentional template, which is possibly reflected in the later, frontally distributed part of the PN (negative difference late). Later we will reinterpret these two attention mechanisms in terms of (i) enhanced or reduced gain of the prediction error and (ii) top-down prediction activity.

As described above, the predictive coding theory explains perception as inference about causes of sensory input, which emerges in a

process of establishing and refining internal generative models. The process of perceptual inference is concerned with the content of the sensory input, as well as the confidence in the sensory data and therefore it needs to take into account the precision of the sensory data (Brown & Friston, 2013; Hohwy, 2013). Prediction error signals are responsible for the passage of sensory information from lower to higher levels of the inferential hierarchy in order to update the conditional expectations of the internal generative model. To minimize surprise, the perceptual inference mechanism needs to optimize the precision of the prediction error signal. Precision is thought to be optimized through the synaptic gain modulation of prediction error units on each level of the hierarchy. In this framework, attention can thus be defined as the optimization of precision in the service of hierarchical prediction error minimization (Friston, 2009; Feldman & Friston, 2010; Hohwy, 2013).

In line with the idea that attention can be understood as the optimization of the precision of prediction errors, Feldman & Friston (2010) employed computational modeling to simulate electrophysiological and psychophysical findings from the Posner spatial cueing task (Posner, 1980). The simulated electroencephalography data of Feldman & Friston (2010) show a stimulus-related enhancement for validly cued targets in the P1 and N1 range, which mimics the effects obtained in actual experiments. These reflect an increased precision (i.e. gain) of prediction error signals for the validly cued targets, and a low precision for invalidly cued targets. Moreover, the increased P3 response for invalidly cued targets was interpreted as reflecting a prediction error that updates and revises the probabilistic context encoded by the inferred hidden states. This explanation is consistent with the sensory gain explanation described above, the difference being that, in the predictive coding models, transient electrophysiological responses are considered to be signals generated by prediction error units (Feldman & Friston, 2010; Brown & Friston, 2013), whereas in the classic attention models the gain operates on information channels (Broadbent, 1958).

Based on results from animal and human research, Scheich et al. (2007) proposed that the nature of the task (attention) modifies the stimulus representations in the auditory cortex. This notion is partly justified by the fact that two-thirds of the input to the auditory cortex are descending from other cortical areas. Similarly, Fritz et al. (2007) suggested that attention serves as the key trigger that initiates a cascade of events resulting in changes of the spectrotemporal receptive fields of the primary auditory cortex in ferrets. Thus, the effects of attention are triggered by top-down activations of the respective auditory areas. These proposals are compatible with the predictive coding theory that the gain and the precision contribute to the attention effects. Also, Ahveninen et al. (2011) argued that auditory attention cannot be solely due to a gaining mechanism, when competitor sounds activate overlapping receptive fields, but that an additional sharp pre-tuning of the to-be-attended pitch is required (cf. Okamoto et al., 2007). Moreover, for attention effects occurring at the somewhat later time range of about 150–300 ms, these authors argue in favor of an endogenous attentional trace (as suggested by Näätänen, 1982) (see above). More specifically, they explain that the respective attentional effect occurs at the receiving end of top-down communications from the prefrontal to the auditory cortex. The recurrent feedback loop may be concerned with the identification of the target as a target and may send the features characterizing the target in working memory to auditory areas. Thus, we have sensory gain and sharpening due to frequency or spatial tuning that can exert early effects, and the re-entrance of frontal cortex representations into the auditory cortex, which takes place somewhat later.

Studies on the relation between auditory attention and prediction

The suppression effect for self-generated sounds: prediction or attention?

Studies comparing self-generated and externally-generated sound responses have provided converging evidence for reduced responses to self-generated sounds taking place in the early sensory processing stages. These could be interpreted as reflecting predictive processing in the brain. However, the paradigms used suffer from several flaws when it comes to unequivocally attributing suppression effects to predictive processing (for a review, see Hughes et al., 2012). Perhaps the most striking confound is the lack of control for concomitant attention effects in most of these studies. Suppression effects may be due to differences in the amount of attention devoted to sound processing during passive listening compared with during self-generation. For example, attentional resources may be directed to the button-press task, so that less attention is devoted to the sounds during self-generation, resulting in a decrease of the N1 amplitude. However, the few studies that have directly addressed this confound do not support this interpretation. In one of those studies, Timm et al. (2013) manipulated the allocation of attention during the performance of a self-generation paradigm. In three different attention allocation conditions, three different aspects of the experimental stimulation were made task-relevant, thereby directing attention to sounds (counting sounds, both externally-generated and self-generated), visual stimuli (counting extended fixation crosses), or one's own motor acts (counting button presses). N1 suppression was present in all three attention allocation conditions. Importantly, the N1 amplitude was also enhanced by attention; however, attention and suppression effects were independent of each other. This was further confirmed by the analysis of different N1 subcomponents: N1 enhancement for attended sounds was present for all three N1 subcomponents (N1a, N1b and N1c), whereas N1 suppression for self-initiated sounds was confined to N1b and N1c. Similarly, Saupe et al. (2013) included a novel active listening condition in their self-generation paradigm. They observed an enhancement by attention and a suppression by prediction (self-generation) on the N1 component and its lateral subcomponent Tb. The scalp distribution of the Tb-suppression effect over the temporal cortices was clearly dissociable from the frontocentral scalp distribution of the attention effect. Taken together, these studies suggested that the effects of sensory predictions (generated by the self-initiation of sounds) and voluntary selective attention on the auditory N1 are two distinct processes that behave in a rather additive fashion and have dissociable electrophysiological correlates.

However, interactions between the prediction and attention effects have been reported in a study by Jones et al. (2013). In this study, action–effect prediction and attention were orthogonally manipulated. Importantly, the self-generation effects were strictly controlled to include only specific self-generation effects, and to discard any unspecific self-generation effects. Thus, specific action–effect associations, linking the execution hand to sound identity, were built up in the acquisition phase of the experiment, i.e. left-hand or right-hand button presses generated high-frequency or low-frequency sounds. These associations were either matched or violated in a subsequent test phase of the experiment. Electrophysiological responses to congruent (prediction match) vs. incongruent (violation of prediction) tones were compared. Attention was manipulated in each block in a filtering manner by asking participants to respond to catch tones if they appeared in one ear (attended ear), and to ignore all tones

presented to the other ear (unattended ear). Both independent and interdependent ERP effects of prediction and attention were reported. Typical attention effects were found on the auditory N1. However, action–effect prediction influenced only the later stages of processing. Compared with congruent tones, incongruent tones elicited a P3 response over the anterior electrodes, reflecting an orienting response to prediction-violating stimuli (Friedman *et al.*, 2001). Finally, incongruent tones also elicited a late (ca. 200 ms) response over the posterior electrodes. Crucially, this effect interacted with attention, as it was only present in the unattended condition, and not in the attended condition. The authors suggested two possible interpretations for this interactive effect. First, they proposed that the component could reflect an evaluative process regarding action–effect associations, which would take place under attentive conditions regardless of whether there is a match or mismatch, but would only be called into action when there is a mismatch under non-attentive conditions. Alternatively, the interaction could reflect a ceiling effect in the gain of the brain signals, in the sense that, under attention conditions, the amplitude could not be further increased for incongruent sounds. One critical difference between this study and the two studies described above is that here, the interactive effects were found exclusively at a late stage of processing on error evaluation signals, when predictions were violated, whereas the two previous studies focused on prediction-match signals, investigating possible influences of attention on the N1-suppression effects when comparing predictable (self-generated) vs. unpredictable stimulation.

From within the predictive coding framework, Brown *et al.* (2013) have proposed an explanation for sensory attenuation based on active inference, which would indeed relate self-generation effects to an attentional mechanism. They propose that sensory attenuation is a necessary consequence of reducing the precision of sensory evidence during movement, to allow the expression of proprioceptive predictions that incite movement (see Adams *et al.*, 2013; for a detailed explanation of this account of motor control). In their view, during movement, the precision of sensory evidence is attenuated, resulting in a decrease in gain. This is in fact equivalent to attending away from the consequences of self-made acts.

Thus, it remains to be resolved whether self-generation suppression effects reflect only prediction, only attention or an additive or interactive effect of these two factors. Following from our analysis on specific and unspecific contributions to these effects, we could speculate that perhaps specific effects are related to the degree of match between the content of motor–sensory predictions and the received input, whereas unspecific effects might be rather related to modulations in gain during active movements.

The interaction between attention and prediction

The evidence reviewed in this article so far shows that attention and prediction mechanisms leave distinct signatures on electrophysiological responses (cf. Fig. 3 for typical findings of auditory attention and prediction in human ERPs). Whereas attention leads to an increase of neural responses to attended stimuli, prediction seems to suppress responses to predicted stimuli. However, when we inspect responses in the absence of stimulation, they seem to produce similar effects, activating sensory templates. The interactions observed in some studies between attention and prediction point to their interrelatedness. The predictive coding framework offers a unifying view on prediction and attention (for reviews, see Hohwy, 2013 and Clark, 2013), which takes into account their synergistic effect.

In some classic attention paradigms, such as Posner's spatial cueing task, the task relevance of the stimuli is manipulated by probabilistic cues indicating the likelihood of their occurrence at a particular spatial location. Attentional cues therefore provide combined information about the probability of stimulus occurrence and its task relevance (Summerfield & Egner, 2009, 2014). Attention is thus always oriented to the feature (e.g. spatial location) predicted by the cue. Information about probability and relevance is usually related, but in some cases might constitute potentially orthogonal sources of information flow. Behavioral chronometric studies have provided some evidence that selective attention can be modulated by the probabilistic information (Klein, 1994; see also Klein & Hansen, 1990; Zhao *et al.*, 2013). For instance, Klein (1994) has shown that endogenous spatial orienting of attention interacted with the probability of stimulus–response likelihood; thus the benefit of valid spatial symbolic cues was present for frequent and therefore expected stimulus–response bindings, but was diminished for infrequent and therefore unexpected stimulus–response bindings. Moreover, an electrophysiological study by Lasaponara *et al.* (2011) also demonstrated that top-down attentional orienting is modulated by bottom-up probabilistic information. In the predictive condition, the authors observed orienting costs reflected in slower responses to less likely invalidly cued trials than to more frequent neutral trials, together with a reduced visual P1 component for invalid trials. However, both behavioral and ERP effects were diminished in the non-predictive condition, in which valid, neutral and invalid trials were distributed with equal probabilities (see also Doricchi *et al.*, 2010 for fMRI evidence on such an interactive effect). According to Macaluso & Doricchi (2013), the evidence that top-down endogenous attentional control is contingent upon bottom-up probabilistic information challenges the traditional dichotomy between top-down attentional control associated with the activation of the dorsal frontoparietal network and bottom-up, stimulus-driven attentional control subserved by the ventral frontoparietal network. Furthermore, Awh *et al.* (2012) also pointed out that the dissociation between top-down and bottom-up attentional control is insufficient, as the notion of top-down attention conflates orthogonal interacting factors, one related to current goals and one related to the history of previously selected stimuli. As we further argue, predictive coding might serve as an integrative framework to account for these interactions.

The evidence reviewed in Section Cognitive psychophysiology of auditory prediction suggests that evoked responses for predicted or more probable input are suppressed, which can be interpreted in line with the assumption of predictive coding that the prediction error for predicted and thus explained-away input is reduced. This explanation is seemingly contradictory to some findings of increased responses for expected stimuli (see, e.g. Tervaniemi *et al.*, 1994; Doherty *et al.*, 2005; Chaumon *et al.*, 2008; for a review, see also Lange, 2013) that might be attributed to attentional modulation. However, these findings can be reconciled if we consider prediction and attention as two forms of perceptual inference that act in a synergistic manner. Perceptual inference is, from the predictive coding perspective, understood as an iterative process of prediction error minimization; therefore the internal models of the sensory input (and thus predictions generated by it) are only more (leading to smaller prediction error) or less (leading to larger prediction error) accurate approximations of the sensory input. Considering that the external environment and our brains are inherently noisy, prediction errors are very unlikely to be zero (see e.g. Hohwy, 2013; Chapter 8). As Hohwy (2013, Chapter 2) argues: "Heuristically, a model that fits data perfectly is itself highly implausible, because it violates our prior beliefs that noise is always greater than zero." That is why the additional aim of perceptual inference is to correctly estimate the level of noise, or variability of the signal (the inverse of which is

equal to its precision). The internal generative model is hence involved in two types of inference performed simultaneously: the inference on the content of the prediction error (sensory signal) and the inference about the expected precision of the prediction error, or its variability. According to Hohwy (2012, 2013), these can be metaphorically understood as the first-order statistical inference (similar to estimating the mean) and the second-order statistical inference (similar to estimating the variance). The latter type of inference is related to attention, as attention is understood as a precision-weighting mechanism that regulates the gain of prediction error. The two types of inference are inherently related, prediction error signals being modulated by their expected precision. What follows from the scheme is that the gain of the prediction error can be amplified for predicted (more likely) input to a larger extent as the predicted input is expected to be more precise. A study by Kok et al. (2012b) provides fMRI evidence for this hypothesis. The authors orthogonally manipulated probabilistic information about the side on which a visual grating would appear (block-wise cues), and the task relevance of the gratings (trial-by-trial cues). Their findings revealed an interaction between predictability and task relevance in the Blood Oxygenation Level Dependent signal measured in the primary visual cortex (V1). The Blood Oxygenation Level Dependent response was suppressed for predicted gratings as compared with unpredicted gratings when they were task-irrelevant, consistent with the hypothesis that prediction error is silenced for explained-away input. However, the pattern of responses was reversed when gratings were task relevant: the Blood Oxygenation Level Dependent response was enhanced for predicted and task-relevant gratings. The enhancement by attention was also observed in higher level visual areas (V2, V3), whereas the suppression by attention was observed only in V1. To conclude, the enhancement of responses by the attentional precision-weighting mechanism seems to be dependent upon predictability, as predictable gratings receive greater gain (see Hohwy, 2013; but see also Kok et al., 2012a for an fMRI study reporting no interaction between task relevance and prediction).

Empirical research has just begun to address the interaction of attention and prediction understood as the interplay of predictability factors and precision-related factors. Only a few studies have directly addressed this intriguing issue by orthogonally manipulating these two factors. Hsu et al. (2014) investigated the potential interaction of attention and prediction on auditory electrophysiological responses. In a stream-segregation task, two stimulus streams, one of high-frequency tones and one of low-frequency tones, were interleaved. One of the stimulus streams was predictable and consisted of pairs of tones, the second tones of the pairs always being two natural keys higher than the first tones. The second, unpredictable stream consisted of randomly arranged tones without repetitions. Attention was manipulated block-wise by instructing participants to attend to one of the streams and to detect catch tones of attenuated loudness that would appear in the stream. Electrophysiological responses to attended predictable/unpredictable and unattended predictable/unpredictable tones were compared. Analysis of the auditory N1 showed larger responses for attended and predictable stimuli, as compared with all other conditions, which did not differ from each other. Hence, the enhancement by attention was dependent on the predictability of the stimuli, and the responses to attended and unattended stimuli differed only when they were predictable. However, predicted and unpredicted stimuli differed only when they were attended to. The second finding seems consistent with an fMRI study by Jiang et al. (2013), which applied multivariate neural pattern classifiers to discriminate between expected (predicted) and unexpected (unpredicted) visual stimuli (faces or houses) when they

were attended (task-relevant) or unattended. They showed that voxel activation patterns for predicted and unpredicted stimuli differed from each other more substantially when stimuli were attended to compared with when they were ignored. Taking both findings together, they seem to support a conclusion that attention increases the disparity between neural responses to predicted and unpredicted input.

The experimental evidence on interactive effects between attention and prediction that has been collected so far might seem to show a somewhat divergent pattern of results. In addition to the obvious differences in the methodology applied to study neural responses (e.g. Blood Oxygenation Level Dependent signal or ERPs) and the modality studied (auditory, visual or multimodal stimuli), it is important to note that each study applied different operational definitions of attention and prediction. The first divergence refers to different manipulations of prediction. In some studies (Kok et al., 2012b; Jiang et al., 2013; see also Jones et al., 2013) prediction effects reflect the comparison between predicted (more likely) and unpredicted (unlikely) input, whereas in another (Hsu et al., 2014), a fully predictable condition is contrasted with a non-predictable, random condition. Interestingly, in the latter case the attention effect was shown to be fully contingent upon predictability. This is consistent with the idea that the mechanism modulating the gain of prediction error (i.e. attention) operates through inference about expected precision and increases gain for predicted input. Intriguingly, in the latter case, no suppression by prediction in the unattended condition was observed. When we compare unpredicted input vs. predicted input, the difference might be interpreted as a suppression of responses for predicted input or as an enhancement of responses for the unpredicted input. As suggested by Larsson & Smith (2012), increased responses for unpredicted stimuli could reflect an increased gain of prediction error for the unpredicted input, leading to an orienting response towards unexpected events in the environment. Hypothetically, this effect could be present at the later processing stages or higher up in the cortical hierarchy (cf. Jones et al., 2013). This idea seems consistent with the interpretation that P3 responses, traditionally thought to reflect involuntary attention orienting towards unexpected events (Friedman et al., 2001), are within the predictive coding framework seen as prediction errors revising the internal generative model.

The second divergence regarding the operational definitions of attention and prediction in the studies reviewed above is related to the manipulation of attention. Firstly, in some studies (cf. Jiang et al., 2013; Jones et al., 2013; Hsu et al., 2014), attention was manipulated in the 'filtering' fashion, as participants were directed by block-wise cue to pay attention to only one type of stimuli (e.g. stimulus stream of certain frequency, stimuli of certain feature) while ignoring all other stimuli. In the study by Kok et al. (2012b), however, attention (understood as task relevance) was manipulated in the trial-by-trial fashion by symbolic spatial cues (i.e. arrows). Electrophysiological studies that have considered both types of attentional manipulations have shown that attention-related enhancements of early negativities are usually more pronounced when attention is manipulated in the filtering fashion (sustained attention) rather than in a trial-by-trial fashion (transient attention) (see, e.g. Eimer, 1993; Schröger, 1993; for a discussion, see Lange, 2013). As discussed by Hsu et al. (2014, p. 6), in the case of the filtering paradigms the unattended input might be suppressed to a larger extent at earlier stages of processing. However, in the case of trial-by-trial cueing, unattended stimuli are likely to be processed to a larger extent. It can be assumed that, in these two types of paradigms, the internal generative model of the experimental situation differs. In

the sustained, filtering paradigm, the precision of the prediction error for unattended input is relatively lower than it is in the case of transient, trial-by-trial cueing, in which even unattended stimuli are expected to have a higher level of precision. This might lead to a relatively more pronounced gain for the attended input in the former than in the latter. Importantly, irrespective of whether attention is manipulated in the sustained or transient fashion, the attention filter might be applied at different levels of the processing hierarchy depending on the behavioral goals (Serences & Kastner, 2014). For instance, in the study by Hsu *et al.* (2014), the participants' task was to detect the tones of attenuated loudness, and therefore the pitch of the sound, which could be either predictable or unpredictable, was irrelevant for the correct performance of the task and could be ignored. However, in two other studies, the predictability was manipulated within features (Jiang *et al.*, 2013) or spatial locations (Kok *et al.*, 2012b) that were to a larger extent relevant for the correct task performance. Thus, the locus of selection and degree of task relevance might determine where and when in the processing hierarchy the gain modulation of prediction errors occurs.

According to Wacongne *et al.* (2011, see Section Prediction effects uncovered in the absence of stimulation: omission paradigms in the current paper), auditory perception can be understood as an internal model of the sensory input that is a result of a hierarchical cascade of prediction errors being compared with top-down predictions at different levels of the processing hierarchy. Prediction errors at different levels might be differentially modulated by top-down attentional engagement. This hypothesis was investigated by Chennu *et al.* (2013). In their study, they manipulated the degree of bottom-up predictability by presenting tones that deviated in frequency (monoaural deviants) or laterality (interaural deviants) at either the local sequence or the global sequence level (cf. study by Wacongne *et al.*, 2011 described above). Attention, defined as "increased precision in perceptual inference" (Chennu *et al.*, 2013, p;. 11194), was engaged by making the auditory stimulation task-relevant, or diverted away from auditory stimuli by a demanding visual task. The authors also distinguished another source of top-down influence, namely the top-down expectation as biasing attention "in favor of contextually anticipated stimuli" (Chennu *et al.*, 2013, p. 11194). Top-down expectation was manipulated in the two attentional engagement conditions by different task instructions: attend to individual sounds (count the number of local deviants) or attend to sequences (count the number of global deviants). At the early level of the auditory hierarchy, MMN proved to be sensitive to the degree of bottom-up predictability, as indicated by larger responses to the less predictable interaural than monoaural deviants. No interaction between attention and bottom-up predictability was found, although responses to deviants were in general attenuated when attention was directed away. Interestingly, MMN for both types of deviants was reduced as a function of the top-down expectation. The interaction between predictability and attention was found at a higher level of the hierarchy. The P300 to interaural deviants was larger than to monoaural deviants with attention, whereas the P300 responses to both types of deviants were diminished when attention was directed away. Hence, prediction error responses at the higher level of the predictive hierarchy that are likely to reflect updating of the internal generative model were contingent upon attentional engagement. The latency of P300 was in turn modulated by top-down expectations; the P3 latency was shorter in the attend-to-tones condition compared with the attend-to-sequences condition.

Chennu *et al.* (2013) were able to demonstrate differential influences of 'bottom-up predictability', reflected in the effects of 'local' deviance, and two types of top-down influence: attention and top-

down expectation. The theoretical model of Lange (2013) of the effects of temporal orienting of attention on the auditory N1 also draws a distinction between these three terms (although the suggested underlying mechanisms are slightly different). Temporal orienting is defined as the selection of information according to stimulus occurrence at a particular, task-relevant point in time, resulting in prioritized processing of that stimulus (see, e.g. Nobre & Coull, 2010) (in other contexts temporal orienting has been termed the 'temporal' or 'when' aspect of prediction as opposed to the 'structural' or 'what' aspect, Schwartze *et al.*, 2012). Lange (2013) suggests that selection based on the temporal dimension is based on the opposite effects of attention and prediction. She defines attention as a multiplicative function of task relevance and *a-priori* expectation, whereas prediction is a multiplicative function of conditional probability (the probability of stimulus occurrence after a particular time has elapsed) and *a-priori* expectation. According to this model, attention and prediction contain a common term, namely the *a-priori* expectation considering *a-priori* probability (i.e. probabilistic information delivered by the contingencies between cue and task-relevant stimuli) and rhythmic regularity (triggering expectations based on periodic stimulus presentation), thus interactions between attention and prediction should occur. In fact, her model is compatible with many of the empirical phenomena in this field.

In sum, in line with Friston's predictive coding theory and Lange's model of temporal orienting, attention and prediction may exert both independent and interactive effects on the processing of sounds. However, the effects of prediction, as well as their interaction with attentional mechanisms might be observed at different processing stages. This suggests that interactions might take place at different stages along the cortical hierarchy of predictive models (cf., Chennu *et al.*, 2013). However, in order to be able to propose a concise framework based on the experimental evidence, precise definitions of the manipulated variables should be considered. In the next section, we propose a framework to organize the existing research and guide future attempts for investigating prediction and attention.

Linking the cognitive psychophysiology lesson and predictive coding theory in order to organize research on prediction and attention in audition

### Conclusions from the cognitive psychophysiology lesson

We would like to bring the threads presented throughout this article together (Fig. 4). The evidence reviewed has clarified three main points: (1) prediction and attention are different mechanisms; (2) prediction and attention have often been confused in previous research; and (3) prediction and attention are interdependent.

Prediction and attention are different mechanisms

Attention and prediction are separate and differentiated mechanisms, which operate in a distinct fashion. Sections Cognitive psychophysiology of auditory prediction and Cognitive psychophysiology of auditory attention show evidence for different effects of the two mechanisms in terms of whether brain activity is enhanced or suppressed, the locus of the respective attention and prediction effect, and the time-course of the effect.

In line with the predictive coding (Friston, 2009, 2010), we suggest that prediction is occupied with making inferences about the causes of sensory input and their expected precision, whereas attention is involved in optimizing the precision of sensory input and

FIG. 4. Attention in the predictive coding framework. The predictive (generative) model makes inferences on the content of and the confidence in the sensory evidence [i.e. the prediction error (PE) from a previous level]. The difference between the input and the content of the prediction is expressed in the PE, which updates the model. Sensory ERPs are understood as the transient expression of the PE. Once the input is explained by the prediction, the perceptual problem is solved and we perceive our interpretation of the world. The relevant content of the inference (attentional template or attentional trace) is what we attend to in a particular experimental situation. Confidence in the attended content is higher than the unattended content. Attention improves the precision of the PE (i.e. it reduces the estimated variability) for the attended content via the predictive model. This results in a modulation of the gain of the PE. Thus, attention will usually increase the amplitude of the PE, which will in turn make the predictive model more accurate. As we usually have several coexisting predictive models, which compete for dominance, attention exerts an indirect bias towards the model that contains the relevant content. The impact of attention on the precision of the PE constitutes an independent factor in the dynamic system of feedback/feedforward recurrent loops, which may interact with other factors of the predictive model (e.g. with the predictability of the input). Thus, the (partly) contradictory effects of attention and prediction reported from ERP research can be explained by the interaction between content prediction and precision estimation.
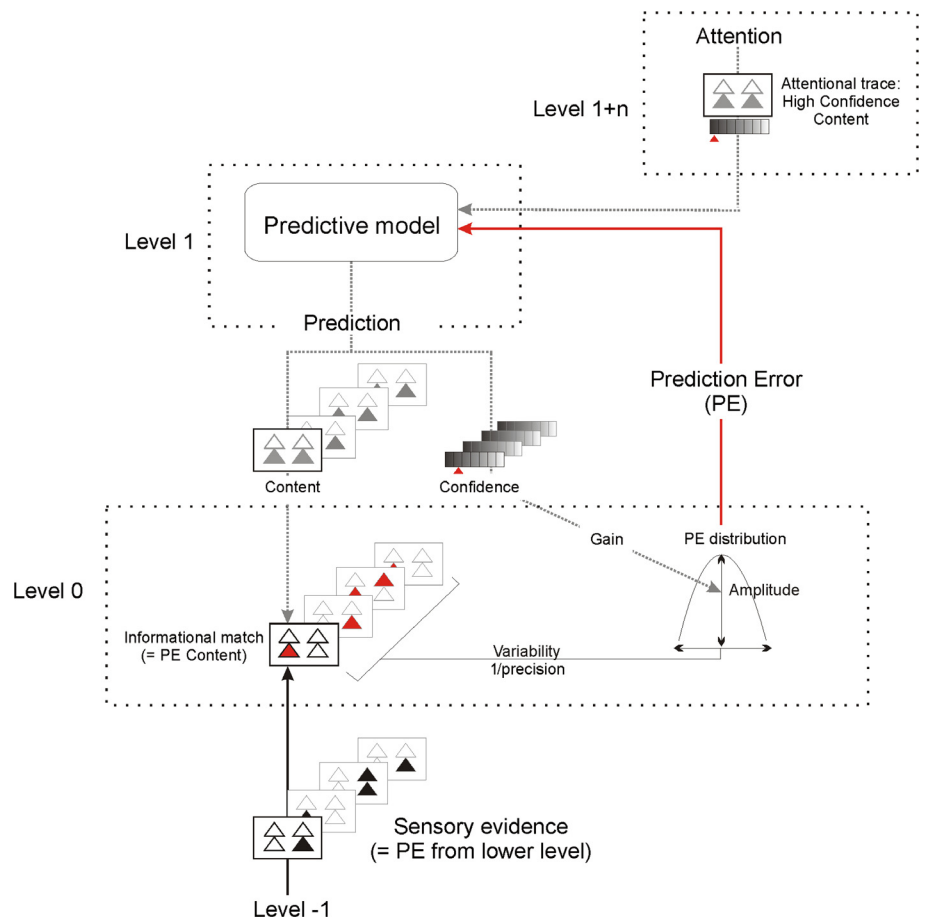
regulating the gain of the feedforward signals, i.e. prediction errors (but see, e.g. Bowman *et al.*, 2013 for the discussion of the limitations of this definition). Staying faithful to this, in the following, the term attention will be confined to refer to gain modulations and the term prediction to refer to the inferences. As we will see, this will diverge considerably from what has been considered as attention or prediction in the previous literature as summarized in Section Prediction and attention have often been confused in previous research.

Prediction and attention have often been confused in previous research

A great deal of confusion regarding the differentiation of these two mechanisms has come from varying definitions of the terms. The operational definitions of prediction and attention used in previous research have been reasonable, but also somewhat arbitrary. In many of the current attention and prediction studies, prediction is induced by cues containing probabilistic information about the forthcoming sound (e.g. its spatial location or pitch), whereas attention is induced by task instructions defining one type of stimuli (i.e. stimuli occurring on the left side) as task-relevant (i.e. they may require a behavioral response) and stimuli not comprising the task-relevant feature value as task-irrelevant (i.e. they do not require a behavioral response). In previous behavioral research on attention, where it was required that participants also responded to unattended target stimuli, attention was manipulated by the probabilistic information. In other words, what is regarded as prediction in some current studies has been regarded as attention in the older literature. A proper differenti-

ation of attention and prediction would make it easier to manipulate them in an orthogonal fashion. In the present context, it seems advantageous to define prediction as being related to probability and attention as being related to instructional (and possibly even motivational, emotional) relevance (e.g. Summerfield & Egner, 2009; Kok *et al.*, 2012b).

Another usage of concepts in which there has been poor agreement is the usage of the endogenous/exogenous dichotomy and top-down/bottom-up dichotomy, especially when used as a synonym to the voluntary/involuntary attention dichotomy. Detailed analyses of the various meanings that different researchers have assigned to these concepts and their overlaps and disagreements have been provided elsewhere (Engel *et al.*, 2001; Summerfield & Egner, 2009). Here, top-down and bottom-up will be interpreted in a strictly anatomical sense: top-down refers to the passage of information from higher levels of the hierarchy to lower levels of the hierarchy (what in the physiological studies is referred to as feedback), and bottom-up is feedforward. Top-down signals carry predictions and bottom-up signals carry sensory evidence, i.e. prediction error (see also Rauss & Pourtois, 2013, for a similar conceptualization). With endogenous, we will refer to that brain activity that is not directly elicited by external input, which we will refer to as exogenous. The term voluntary will be reserved to those (conscious) operations that are induced by the task or goal of the participant.

Prediction and attention are interdependent

Section Studies on the relation between auditory attention and prediction has provided evidence that prediction and attention interact

in various ways. Three main reasons for their interdependence can be derived from our definitions of attention and prediction.

iBecause sensory feedforward signals are an expression of prediction error, they are dependent on the underlying predictions generated by the internal model. Attention therefore operates on and modulates signals, which in fact reflect our predictions about the world.

iiBecause perceptual inference is not only occupied with making inferences about the content of the prediction errors, but also about their expected precision, and the expected precision is directly related to the gain of the feedforward signals, i.e. to attentional modulation.

iiiBecause increasing the gain means sampling the environment (or checking the internal model against the input) more efficiently, by which more efficient updating of the model will take place, and this ultimately results in better predictions.

Framework to organize research on prediction and attention in audition

In the following we attempt to parse these effects and their possible interactions in a systematic way, and try to link each of the effects and/or mechanisms to psychological constructs from the attention and prediction literature. We would like to emphasize that this is a suggestion of how to organize the concepts and findings, which we hope may generate new experimental hypotheses. However, we do not claim that it is an exhaustive list, or that all of what is listed is an axiomatic truth. This remains to be tested.

How is the content and amplitude of feedforward responses determined? There are two aspects affecting the feedforward prediction error signals, one of which determines the content, and the other the magnitude.

The content depends on how well the input signal matches the prediction. The computation of the differences cancels out the activity of units where input and prediction match (explains away), and transmits the activity of those units where input and prediction mismatch (cf. Figs 1 and 2). In this way, the new information (not yet contained in the predictive model) encoded by the responding neural units is identified and the model is informed about it. This is consistent with the interpretations that N1-suppression effects for predicted stimuli are due to a match between prediction and input. It is also what has been investigated with oddball paradigms, which typically compare brain activity for match vs. mismatch situations. With these studies it can also be monitored how the model and its predictions become better and better, and the informational match *per se* becomes more precise. Because with a better model the predictions will be more correct, a better model should in general lead to less units responding, which is also consistent with the interpretation that sharpening of responses can occur (Fritz *et al.*, 2007; Scheich *et al.*, 2007; Waszak *et al.*, 2012). The responses elicited as a result of the informational matching process can be modulated by the factors affecting gain modulations, which we discuss next.

The magnitude of the prediction error signals is given by the gain. The gain can be modulated by the intention of the listener and by the predictability of the input, as well as by the interaction between the two.

The gain can be adjusted voluntarily, by virtue of intention, which has often been called top-down or endogenous attention. How do we decide what we want to attend to? From our perspective, this content (what I want to attend to) is simply an inference at a very high level of the hierarchy. Just as we infer the causes of sensory signals, at a higher level we also infer what is relevant given such a model of the situation as a whole. This inference takes many different sources of information into account, i.e. several processing modules of the brain may exert an influence, e.g. motives and emotions. We will call the content of this inference on what is relevant the attentional template. We propose that this is equivalent to the attentional trace (Näätänen *et al.*, 1978; Näätänen, 1982), and can also be related to the proposal of Ahveninen *et al.* (2011). Regardless of how the template of what I want to attend to is decided, it is used to increase the gain of that representation. In other words, gain can be increased because at a high level of the hierarchy we have an attentional template that encodes what we want to attend to. This results in greater precision for the respective representations (predictions), resulting in increased gain of their prediction errors. When we are trying hard to listen to something, we fine tune our predictions regarding that one feature. This explains the enhanced stimulus-evoked responses for attended input, and is akin to the gain theories of attention effects (e.g. Hillyard *et al.*, 1973).

However, gain can also be adjusted depending on the predictability of the input. Because this depends on the statistical characteristics of the input and not on the intentions of the subject, it would fall under the category of what has previously been termed by many 'bottom-up', exogenous, or involuntary attention. It is the characteristics of the input that is driving the gain modulations (i.e. driving attention). There are two aspects to this. The first refers to the difference between predictable and unpredictable contexts. In more predictable contexts, the system will be able to form a more accurate model of the situation, and thereby the inferred precision of the prediction errors will be higher. In principle, more precise prediction errors receive greater gain (cf. Hohwy, 2013). The second aspect relates to the fact that gain can be adjusted depending on the amount of prediction error the system is receiving. Because the system's goal is to minimize prediction error (Friston, 2010), large prediction error signals force reconfigurations of the internal model, and possibly call for the involvement of voluntary attention mechanisms. Thus, our predictive coding perspective reveals links between the classical concepts of voluntary and involuntary attention that are not self-evident. Salient sounds initially result in a large prediction error (when no adequate top-down predictions are present, the prediction error corresponds to the input), and this may require model reconfiguration at higher levels, which may activate 'voluntary' top-down influence, which, in turn, may additionally amplify the impact of the salient sounds due to greater gain. To what degree this happens is related to how predictable the context is. In more predictable contexts, as we have just explained, we have more confidence in our predictions, and therefore, if there is a violation of the rule, the prediction error elicited by the mismatch will be larger than if we are in an unpredictable context. Intuitively, this relates to a situation in which the current model should be seriously reconsidered, or, as there are almost always several concurrent models available (Mill *et al.*, 2013), turn to another model. The former scenario corresponds to what has been termed involuntary orienting of attention, indicated by the P3a, a higher level prediction error response (e.g. Escera *et al.*, 2000). The latter scenario has been investigated in ambiguous hearing conditions, e.g. in the auditory streaming literature (Bendixen *et al.*, 2013).

Importantly, voluntary gain adjustment and gain adjustment due to the predictability of the input can, and almost always do, interact. First, voluntary attention makes use of current models and selectively enhances those parts of the model that are relevant, and also uses the model to dampen those things that are irrelevant (and that we already know of). This latter point is probably one aspect of the effects observed for self-generated stimulation (e.g. Brown *et al.*,

2013), and also relates to the 'attention costs' for selectively ignored stimuli when compared with neutral conditions (Schröger & Eimer, 1997; Bidet-Caulet *et al.*, 2010). This interaction is also described in the study by Kok *et al.* (2012b). The selective enhancement and dampening of gain depending on relevance can be performed much more efficiently when we have a good model. If we do not have a good model, it will be very difficult to decide which parts of the stimulation are relevant and irrelevant. Voluntary attention operates on the predictions that we have; thus, if our predictions are not accurate, attention will operate less effectively. Second, predictability has an influence on what we decide is relevant. As we said previously, the 'what' is relevant can be an inference drawn on a higher level, which is affected by the input it receives from the lower levels (and from different processing modules). One example of such a situation is when the voluntary attention mechanism is captured by large prediction error (see above). Speculatively, if the prediction error prompts an updating of the model that is efficient in quickly explaining away the prediction error again, then we may remain unaware of this. In this case we would not consider that the event has captured our voluntary attention. This case is illustrated in typical passive oddball paradigms in which irregular sounds only elicit MMN, but no higher error response of the brain such as P3a. However, if the issue cannot be resolved at the lower level of the hierarchy where MMN is generated, and it forces a reconfiguration of the model at the top of the hierarchy, then we would inevitably become aware of it, and thus we consider that this has captured the voluntary attention mechanism. This scenario is what has often been interpreted to happen when a more salient irregular event elicits a P3a response (Escera *et al.*, 2000).

Traditionally, salient events or large physical transients in stimulation are regarded as one cause of eliciting involuntary attention, in addition to events violating a model/prediction, which also can result in the involuntary orienting of attention (e.g. Eimer *et al.*, 1996). Indeed, a larger input signal will activate more units resulting in larger prediction error, transferring more content to the next level in the hierarchy that has to be explained. However, according to our proposal, this will only happen in the absence of a prediction for that large input. In this sense, the prediction error elicited by salient sounds is not fundamentally different from the prediction error elicited by less salient but still prediction mismatching sounds. In any case the prediction error resulting from a salient sound (with no complementing prediction) will call for attention as a considerable model configuration is required, which may engage voluntary attention. This, in turn, will increase the gain and result in even higher prediction error. Nevertheless, according to Feldman & Friston (2010) and Hohwy (2013), a strong signal can also immediately increase the gain of the prediction error associated with the signal.

To recapitulate what has been described so far, feedforward prediction error signals are determined by the degree of (mis)match between prediction and input, which determines their content, and by several variables that can influence the gain, which determines their magnitude. Thus, until here, we have explained factors determining the stimulus-evoked responses that are elicited (N1, MMN and P3) and how their amplitude is determined. As we have seen, these factors also relate to different psychological constructs from the attention and prediction literature, i.e. voluntary and involuntary attention and the like. We have also seen that traditional dichotomies, such as top-down vs. bottom-up, might be insufficient attempts to dichotomize empirical phenomena (Anderson, 2011) that do not capture the mutual interdependence and constant interaction of these aspects (Rauss & Pourtois, 2013).

However, throughout Sections Cognitive psychophysiology of auditory prediction and Cognitive psychophysiology of auditory attention, we have also described neurophysiological signals that are observed in the absence of stimulation (i.e. omission responses and sustained attention signals). These are purely endogenous signals and as such we believe that they are a reflection of prediction, i.e. of top-down signals. These signals have been observed less frequently, and in consequence our knowledge about them is less precise. First, a clear differentiation must be made between omission responses and the slow waveforms found in sustained attention situations. Despite being purely endogenous signals, omission responses are still prediction error signals. This issue has already been discussed in Section Cognitive psychophysiology of auditory prediction, and we will not further dwell on it here. Regarding the sustained attention signals, we would like to propose that they are at least in part a direct reflection of the attentional template. This is line with previous theories (Näätänen *et al.*, 2002) and in agreement with the predictive coding framework. Hence, such signals should reflect the content of the representations (predictions) that we choose to attend to.

To close our discussion of the framework, we should emphasize once more the hierarchical nature of the system. In each level of the hierarchy resides a somewhat different internal model, and hence the content of the predictions formulated also differs. When we assess match and mismatch effects, we should take into consideration at which level of the hierarchy the model that the input is matching or mismatching resides. Moreover, gain can also be adjusted at different levels of the hierarchy. This relates to the ideas outlined by Giard *et al.* (2000) of attention acting as a band-pass filter, whose efficiency and locus of effect can be adjusted depending on the attentional demands. Finally, one should also consider interactions that can take place across hierarchical levels (as illustrated by Wacongne *et al.*, 2011 and Chennu *et al.*, 2013, described above).

Definitions and manipulations to which researchers should increase their gain

When one systematically analyses the literature on attention and prediction research in the light of our framework, it becomes immediately clear that there are certain operational definitions and design manipulations that are particularly relevant, and to which one should pay special attention.

First, of course, is the issue of how attention and prediction are defined. In this respect, it is particularly important to assess to what degree the operational definitions are orthogonal. As we have seen, interactions between the two will occur, and thus it is important to assess whether the attention and prediction manipulations are affecting the same or different content. For example, one should always consider the possibility that, even though the prediction information is in principle not relevant for the task (e.g. it informs about the location of the stimulus, but responses are required regarding the identity), it is often possible to translate it into a relevance cue. In this example, if the stimulus is predicted on the right, and responses are required to high but not low tones, following the rules outlined above, it should be possible to selectively increase the gain for high tones presented on the right side.

Second, one should pay attention to the type of comparisons made to assess the different effects, in particular to whether prediction effects are assessed comparing predictable and unpredictable contexts, or predicted (prediction-matching) and unpredicted (prediction-mismatching) stimulation. The first comparison rather taps into

the confidence (precision) of the predictions and the second rather taps into the content of the predictions.

Third and finally, one should consider at which level of the hierarchy the attentional filter is applied, and the predictions being induced and/or violated.

Closing remarks

We have reported a cognitive psychophysiology approach to study the processes underlying attention and prediction. We have briefly discussed some, but by no means all relevant, studies in these fields, and introduced the core principles of the prevailing theoretical accounts of the results obtained (without going into too much detail). It turned out that attention often increases various parameters of brain activity, whereas prediction often results in the attenuation of brain activity, although there are studies yielding opposite effects. We have argued that the predictive coding theory can serve as an integrative theory (Gigerenzer, 2010) successfully relating prediction and attention. As another prospect of this common framework, the permeability between different disciplines (such as biology and psychology) investigating attention and prediction is expected to be enhanced.

Acknowledgements

Abbreviations

ERP, event-related potential; fMRI, functional magnetic resonance imaging; MEG, magnetoencephalography; MMN, mismatch negativity; PN, processing negativity; STG, superior temporal gyrus.

References

Adams, R.A., Shipp, S. & Friston, K.J. (2013) Predictions not commands: active inference in the motor system. *Brain Struct. Funct.*, **218**, 611–643.

Ahveninen, J., Hämäläinen, M., Jääskeläinen, I.P., Ahlfors, S.P., Huang, S., Lin, F.H., Raij, T., Sams, M., Vasios, C.E. & Belliveau, J.W. (2011) Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. *Proc. Natl. Acad. Sci. USA*, **108**, 4182–4187.

Alho, K., Töttölä, K., Reinikainen, K., Sams, M. & Näätänen, R. (1987a) Brain mechanism of selective listening reflected by event-related potentials. *Electroen. Clin. Neuro.*, **68**, 458–470.

Alho, K., Donauer, N., Paavilainen, P., Reinikainen, K., Sams, M. & Näätänen, R. (1987b) Stimulus selection during auditory spatial attention as expressed by event-related potentials. *Biol. Psychol.*, **24**, 153–162.

Alho, K., Rinne, T., Herron, T.J. & Woods, D.L. (2014) Stimulus-dependent activations and attention-related modulations in the auditory cortex: a meta-analysis of fMRI studies. *Hearing Res.*, **307**, 29–41.

Aliu, S.O., Houde, J.F. & Nagarajan, S.S. (2009) Motor-induced suppression of the auditory cortex. *J. Cognitive Neurosci.*, **21**, 791–802.

Anderson, B. (2011) There is no such thing as attention. *Front. Psychol.*, **2**, 246.

Arnal, L.H. & Giraud, A.L. (2012) Cortical oscillations and sensory predictions. *Trends Cogn. Sci.*, **16**, 390–398.

Arnal, L.H., Wyart, V. & Giraud, A.L. (2011) Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat. Neurosci.*, **14**, 797–801.

Awh, E., Belopolsky, A.V. & Theeuwes, J. (2012) Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends Cogn. Sci.*, **16**, 437–443.

Baess, P., Widmann, A., Roye, A., Schröger, E. & Jacobsen, T. (2009) Attenuated human auditory middle latency response and evoked 40-Hz response to self-initiated sounds. *Eur. J. Neurosci.*, **29**, 1514–1521.

Baess, P., Horváth, J., Jacobsen, T. & Schröger, E. (2011) Selective suppression of self-initiated sounds in an auditory stream: an ERP study. *Psychophysiology*, **48**, 1276–1283.

Baldeweg, T. (2007) ERP repetition effects and mismatch negativity generation. *J. Psychophysiol.*, **21**, 204–213.

Bäss, P., Jacobsen, T. & Schröger, E. (2008) Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. *Int. J. Psychophysiol.*, **70**, 137–143.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P. & Friston, K.J. (2012) Canonical microcircuits for predictive coding. *Neuron*, **76**, 695–711.

Behroozmand, R., Liu, H. & Larson, C.R. (2011) Time-dependent neural processing of auditory feedback during voice pitch error detection. *J. Cognitive Neurosci.*, **23**, 1205–1217.

Bendixen, A., Schröger, E. & Winkler, I. (2009) I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. *J. Neurosci.*, **29**, 8447–8451.

Bendixen, A., SanMiguel, I. & Schröger, E. (2012) Early electrophysiological indicators for predictive processing in audition: a review. *Int. J. Psychophysiol.*, **83**, 120–131.

Bendixen, A., Bőhm, T.M., Szalárdy, O., Mill, R., Denham, S.L. & Winkler, I. (2013) Different roles of similarity and predictability in auditory stream segregation. *Learn. Percept.*, **5**, 37–54.

Bendixen, A., Scharinger, M., Strauß, A. & Obleser, J. (2014) Prediction in the service of comprehension: modulated early brain responses to omitted speech segments. *Cortex*, **53**, 9–26.

Bidet-Caulet, A. & Bertrand, O. (2005) Dynamics of a temporo-fronto-parietal network during sustained spatial or spectral auditory processing. *J. Cognitive Neurosci.*, **17**, 1691–1703.

Bidet-Caulet, A., Mikyska, C. & Knight, R.T. (2010) Load effects in auditory selective attention: evidence for distinct facilitation and inhibition mechanisms. *NeuroImage*, **50**, 277–284.

Blakemore, S.J., Rees, G. & Frith, C.D. (1998) How do we predict the consequences of our actions? A functional imaging study *Neuropsychologia*, **36**, 521–529.

Bourbon, W.T., Will, K.W., Gary, H.E. & Papanicolaou, A.C. (1987) Habituation of auditory event-related potentials: a comparison of self-initiated and automated stimulus trains. *Electroen. Clin. Neuro.*, **66**, 160–166.

Bowman, H., Filetti, M., Wyble, B. & Olivers, C. (2013) Attention is more than prediction precision. *Behav. Brain Sci.*, **36**, 206–208.

Broadbent, D.A. (1958) *Perception and Communication*. Pergamon Press, Oxford.

Brown, H.R. & Friston, K.J. (2013) The functional anatomy of attention: a DCM study. *Front. Hum. Neurosci.*, **7**, 784.

Brown, H., Adams, R.A., Parees, I., Edwards, M. & Friston, K. (2013) Active inference, sensory attenuation and illusions. *Cogn. Process.*, **14**, 411–427.

Busse, L. & Woldorff, M.G. (2003) The ERP omitted stimulus response to "no-stim" events and its implications for fast-rate event-related fMRI designs. *NeuroImage*, **18**, 856–864.

von Carlowitz-Ghori, K.M., Hohlefeld, F.U., Bayraktaroglu, Z., Curio, G. & Nikulin, V.V. (2011) Effect of complete stimulus predictability on P3 and N2 components: an electroencephalographic study. *NeuroReport*, **22**, 459–463.

Carmel, P.W. & Starr, A. (1963) Acoustic and nonacoustic factors modifying middle-ear muscle activity in waking cats. *J. Neurophysiol.*, **26**, 598–616.

Chapman, C.E., Bushnell, M.C., Miron, D., Duncan, G.H. & Lund, J.P. (1987) Sensory perception during movement in man. *Exp. Brain Res.*, **68**, 516–524.

Chaumon, M., Drouet, V. & Tallon-Baudry, C. (2008) Unconscious associative memory affects visual processing before 100 ms. *J. Vision*, **8**, 10.1–10.10.

Chen, Z., Chen, X., Liu, P., Huang, D. & Liu, H. (2012) Effect of temporal predictability on the neural processing of self-triggered auditory stimulation during vocalization. *BMC Neurosci.*, **13**, 55.

Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibáñez, A., Owen, A.M. & Bekinschtein, T.A. (2013) Expectation and attention in hierarchical auditory prediction. *J. Neurosci.*, **33**, 11194–11205.

Christoffels, I.K., Formisano, E. & Schiller, N.O. (2007) Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Hum. Brain Mapp.*, **28**, 868–879.

Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science *Behav. Brain Sci.*, **36**, 181–204.

Cohen, L.G. & Starr, A. (1987) Localization, timing and specificity of gating of somatosensory evoked potentials during active movement in man. *Brain*, **110**(Pt 2), 451–467.

Colder, B. (2011) Emulation as an integrating principle for cognition. *Front. Hum. Neurosci.*, **5**, 54.

Cooper, R.J., Atkinson, R.J., Clark, R.A. & Michie, P.T. (2013) Event-related potentials reveal modelling of auditory repetition in the brain. *Int. J. Psychophysiol.*, **88**, 74–81.

Costa-Faidella, J., Baldeweg, T., Grimm, S. & Escera, C. (2011) Interactions between "what" and "when" in the auditory system: temporal predictability enhances repetition suppression. *J. Neurosci.*, **31**, 18590–18597.

Creutzfeldt, O., Ojemann, G. & Lettich, E. (1989) Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects own voice. *Exp. Brain Res.*, **77**, 476–489.

Crone, N.E., Hao, L., Hart, J., Boatman, D., Lesser, R.P., Irizarry, R. & Gordon, B. (2001) Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology*, **57**, 2045–2053.

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V. & Hari, R. (2000) Speaking modifies voice-evoked activity in the human auditory cortex. *Hum. Brain Mapp.*, **9**, 183–191.

Debener, S., Herrmann, C.S., Kranczioch, C., Gembris, D. & Engel, A.K. (2003) Top-down attentional processing enhances auditory evoked gamma band activity. *NeuroReport*, **14**, 683–686.

Desimone, R. (1996) Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. USA*, **93**, 13494–13499.

Doherty, J.R., Rao, A., Mesulam, M.M. & Nobre, A.C. (2005) Synergistic effect of combined temporal and spatial expectations on visual attention. *J. Neurosci.*, **25**, 8259–8266.

Doricchi, F., Macci, E., Silvetti, M. & Macaluso, E. (2010) Neural correlates of the spatial and expectancy components of endogenous and stimulus-driven orienting of attention in the Posner task. *Cereb. Cortex*, **20**, 1574–1585.

Eimer, M. (1993) Spatial cueing, sensory gating and selective response preparation: an ERP study on visuo-spatial orienting. *Electroen. Clin. Neuro.*, **88**, 408–420.

Eimer, M. (2014) The neural basis of attentional control in visual search. *Trends Cogn. Sci.*, **18**, 526–535.

Eimer, M., Nattkemper, D., Schröger, E. & Prinz, W. (1996) Involuntary attention. In Neumann, O. & Sanders, A.F. (Eds), *Handbook of Cognition: Perception and Action*. Academic Press, London, pp. 155–184.

Eliades, S.J. & Wang, X. (2003) Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J. Neurophysiol.*, **89**, 2194–2207.

Engel, A.K., Fries, P. & Singer, W. (2001) Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.*, **2**, 704–716.

Escera, C., Alho, K., Schröger, E. & Winkler, I. (2000) Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol. Neuro-Otol.*, **5**, 151–166.

Feldman, H. & Friston, K.J. (2010) Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.*, **4**, 215.

Flinker, A., Chang, E.F., Kirsch, H.E., Barbaro, N.M., Crone, N.E. & Knight, R.T. (2010) Single-trial speech suppression of auditory cortex activity in humans. *J. Neurosci.*, **30**, 16643–16650.

Foxe, J.J., Simpson, G.V., Ahlfors, S.P. & Saron, C.D. (2005) Biasing the brain's attentional set: I. Cue driven deployments of intersensory selective attention. *Exp. Brain Res.*, **166**, 370–392.

Friedman, D., Cycowicz, Y.M. & Gaeta, H. (2001) The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neurosci. Biobehav. R.*, **25**, 355–373.

Friston, K. (2005) A theory of cortical responses. *Philos. T. Roy. Soc. B.*, **360**, 815–836.

Friston, K. (2009) The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.*, **13**, 293–301.

Friston, K. (2010) The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.*, **11**, 127–138.

Friston, K. & Kiebel, S. (2009) Cortical circuits for perceptual inference. *Neural Networks*, **22**, 1093–1104.

Fritz, J.B., Elhilali, M., David, S.V. & Shamma, S.A. (2007) Auditory attention - focusing the searchlight on sound. *Curr. Opin. Neurobiol.*, **17**, 437–455.

Fritz, J.B., David, S.V., Radtke-Schuller, S., Yin, P. & Shamma, S.A. (2010) Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat. Neurosci.*, **13**, 1011–1019.

Fu, C.H., Vythelingum, G.N., Brammer, M.J., Williams, S.C., Amaro, E., Andrew, C.M., Yágüez, L., van Haren, N.E., Matsumoto, K. & McGuire, P.K. (2006) An fMRI study of verbal self-monitoring: neural correlates of auditory verbal feedback. *Cereb. Cortex*, **16**, 969–977.

Galambos, R. & Rupert, A. (1959) Action of the middle ear muscles in normal cats. *J. Acoust. Soc. Am.*, **31**, 349–355.

Gallagher, S. (2005) *How the Body Shapes the Mind*. Cambridge University Press, Cambridge.

Garrido, M.I., Kilner, J.M., Kiebel, S.J. & Friston, K.J. (2007) Evoked brain responses are generated by feedback loops. *Proc. Natl. Acad. Sci. USA*, **104**, 20961–20966.

Garrido, M.I., Kilner, J.M., Stephan, K.E. & Friston, K.J. (2009) The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.*, **120**, 453–463.

Giard, M.H., Perrin, F., Pernier, J. & Peronnet, F. (1988) Several attention-related wave forms in auditory areas: a topographic study. *Electroen. Clin. Neuro.*, **69**, 371–384.

Giard, M.H., Collet, L., Bouchet, P. & Pernier, J. (1994) Auditory selective attention in the human cochlea. *Brain Res.*, **633**, 353–356.

Giard, M.H., Fort, A., Mouchetant-Rostaing, Y. & Pernier, J. (2000) Neurophysiological mechanisms of auditory selective attention in humans. *Front. Biosci.*, **5**, D84–D94.

Gigerenzer, G. (2010) Personal Reflections on Theory and Psychology. *Theor. Psychol.*, **20**, 733–743.

Greenlee, J.D., Jackson, A.W., Chen, F., Larson, C.R., Oya, H., Kawasaki, H., Chen, H. & Howard, M.A. (2011) Human auditory cortical activation during self-vocalization. *PLoS One*, **6**, e14744.

Grill-Spector, K., Henson, R. & Martin, A. (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.*, **10**, 14–23.

Gruber, T. & Müller, M.M. (2005) Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. *Cereb. Cortex*, **15**, 109–116.

Gunji, A., Hoshiyama, M. & Kakigi, R. (2000) Identification of auditory evoked potentials of one's own voice. *Clin. Neurophysiol.*, **111**, 214–219.

Gunji, A., Hoshiyama, M. & Kakigi, R. (2001) Auditory response following vocalization: a magnetoencephalographic study. *Clin. Neurophysiol.*, **112**, 514–520.

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. *J. Neurosci.*, **25**, 10494–10501.

Hansen, J.C. & Hillyard, S.A. (1980) Endogenous brain potentials associated with selective auditory attention. *Electroen. Clin. Neuro.*, **49**, 277–290.

Harter, M.R., Miller, S.L., Price, N.J., Lalonde, M.E. & Keyes, A.L. (1989) Neural processes involved in directing attention. *J. Cognitive Neurosci.*, **1**, 223–237.

Hashimoto, Y. & Sakai, K.L. (2003) Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. *Hum. Brain Mapp.*, **20**, 22–28.

Hazemann, P., Audin, G. & Lille, F. (1975) Effect of voluntary self-paced movements upon auditory and somatosensory evoked potentials in man. *Electroen. Clin. Neuro.*, **39**, 247–254.

Heinks-Maldonado, T.H., Mathalon, D.H., Gray, M. & Ford, J.M. (2005) Fine-tuning of auditory cortex during speech production. *Psychophysiology*, **42**, 180–190.

Heinks-Maldonado, T.H., Nagarajan, S.S. & Houde, J.F. (2006) Magnetoencephalographic evidence for a precise forward model in speech production. *NeuroReport*, **17**, 1375–1379.

Von Helmholtz, H.V. (1867) *Handbuch der Physiologischen Optik*. L. Voss, Leipzig, Germany.

Hillyard, S.A., Hink, R.F., Schwent, V.L. & Picton, T.W. (1973) Electrical signs of selective attention in the human brain. *Science*, **182**, 177–180.

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., Ishizu, K., Yonekura, Y., Nagahama, Y. & Fukuyama, H. (1997) Cortical processing mechanism for vocalization with auditory verbal feedback. *NeuroReport*, **8**, 2379–2382.

Ho, H.T., Schröger, E. & Kotz, S.A. (2014) Selective attention modulates early human evoked potentials during emotional face-voice processing. *J. Cognitive Neurosci.*, doi: 10.1162/jocn_a_00734. [Epub ahead of print].

Hohwy, J. (2012) Attention and conscious perception in the hypothesis testing brain. *Front. Psychol.*, **3**, 96.

Hohwy, J. (2013) *The Predictive Mind*. Oxford University Press, Oxford.

von Holst, E. & Mittelstaedt, H. (1950) Das reafferenzprinzip. *Naturwissenschaften*, **37**, 464–476.

Hommel, B., Müsseler, J., Aschersleben, G. & Prinz, W. (2001) The Theory of Event Coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.*, **24**, 849–878.

Horváth, J. (2013a) Action-sound coincidence-related attenuation of auditory ERPs is not modulated by affordance compatibility. *Biol. Psychol.*, **93**, 81–87.

Horváth, J. (2013b) Attenuation of auditory ERPs to action-sound coincidences is not explained by voluntary allocation of attention. *Psychophysiology*, **50**, 266–273.

Horváth, J. (2014) The role of mechanical impact in action-related auditory attenuation. *Cogn. Affect. Behav. Ne.*, **14**, 1392–1406.

Horváth, J. & Burgyán, A. (2013) No evidence for peripheral mechanism attenuating auditory ERPs to self-induced tones. *Psychophysiology*, **50**, 563–569.

Horváth, J., Maess, B., Baess, P. & Tóth, A. (2012) Action-sound coincidences suppress evoked responses of the human auditory cortex in EEG and MEG. *J. Cognitive Neurosci.*, **24**, 1919–1931.

Houde, J.F., Nagarajan, S.S., Sekihara, K. & Merzenich, M.M. (2002) Modulation of the auditory cortex during speech: an MEG study. *J. Cognitive Neurosci.*, **14**, 1125–1138.

Hsu, Y.F., Hämäläinen, J.A. & Waszak, F. (2014) Both attention and prediction are necessary for adaptive neuronal tuning in sensory processing. *Front. Hum. Neurosci.*, **8**, 152.

Hughes, H.C., Darcey, T.M., Barkan, H.I., Williamson, P.D., Roberts, D.W. & Aslin, C.H. (2001) Responses of human auditory association cortex to the omission of an expected acoustic event. *NeuroImage*, **13**, 1073–1089.

Hughes, G., Desantis, A. & Waszak, F. (2012) Mechanisms of intentional binding and sensory attenuation: the role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychol. Bull.*, **139**, 133–151.

Hughes, G., Desantis, A. & Waszak, F. (2013) Attenuation of auditory N1 results from identity-specific action-effect prediction. *Eur. J. Neurosci.*, **37**, 1152–1158.

Jackson, A.F. & Bolger, D.J. (2014) The neurophysiological bases of EEG and EEG measurement: A review for the rest of us. *Psychophysiology*, **51**, 1061–1071.

James, W. (1890) *The Principles of Psychology, vol 1*. Holt, New York.

Janata, P. (2001) Brain electrical activity evoked by mental formation of auditory expectations and images. *Brain Topogr.*, **13**, 169–193.

Jäncke, L., Mirzazade, S. & Shah, N.J. (1999) Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neurosci. Lett.*, **266**, 125–128.

Jessen, S. & Kotz, S.A. (2013) On the role of crossmodal prediction in audiovisual emotion perception. *Front. Hum. Neurosci.*, **7**, 369.

Jiang, J., Summerfield, C. & Egner, T. (2013) Attention sharpens the distinction between expected and unexpected percepts in the visual brain. *J. Neurosci.*, **33**, 18438–18447.

Jones, A., Hughes, G. & Waszak, F. (2013) The interaction between attention and motor prediction. An ERP study. *NeuroImage*, **83**, 533–541.

Klein, R.M. (1994) Perceptual-motor expectancies interact with covert visual orienting under conditions of endogenous but not exogenous control. *Can. J. Exp. Psychol.*, **48**, 167–181.

Klein, R. & Hansen, E. (1990) Chronometric analysis of apparent spotlight failure in endogenous visual orienting. *J. Exp. Psychol. Human.*, **16**, 790–801.

Knolle, F., Schröger, E., Baess, P. & Kotz, S.A. (2012) The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. *J. Cognitive Neurosci.*, **24**, 698–706.

Knolle, F., Schröger, E. & Kotz, S.A. (2013a) Cerebellar contribution to the prediction of self-initiated sounds. *Cortex*, **49**, 2449–2461.

Knolle, F., Schröger, E. & Kotz, S.A. (2013b) Prediction errors in self- and externally-generated deviants. *Biol. Psychol.*, **92**, 410–416.

Kok, P., Jehee, J.F. & de Lange, F.P. (2012a) Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, **75**, 265–270.

Kok, P., Rahnev, D., Jehee, J.F., Lau, H.C. & de Lange, F.P. (2012b) Attention reverses the effect of prediction in silencing sensory signals. *Cereb. Cortex*, **22**, 2197–2206.

Kokinous, J., Kotz, S.A., Tavano, A. & Schröger, E. (2014) The role of emotion in dynamic audiovisual integration of faces and voices. *Soc. Cogn. Affect Neur.*, doi: 10.1093/scan/nsu105. [Epub ahead of print].

Kudo, N., Nakagome, K., Kasai, K., Araki, T., Fukuda, M., Kato, N. & Iwanami, A. (2004) Effects of corollary discharge on event-related potentials during selective attention task in healthy men and women. *Neurosci. Res.*, **48**, 59–64.

Kühn, S. & Brass, M. (2010) Planning not to do something: does intending not to do something activate associated sensory consequences? *Cogn. Affect. Behav. Ne.*, **10**, 454–459.

Kühn, S., Seurinck, R., Fias, W. & Waszak, F. (2010) The internal anticipation of sensory action effects: when action induces FFA and PPA activity. *Front. Hum. Neurosci.*, **4**, 54.

Kujala, T., Tervaniemi, M. & Schröger, E. (2007) The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biol. Psychol.*, **74**, 1–19.

Lange, K. (2011) The reduced N1 to self-generated tones: an effect of temporal predictability? *Psychophysiology*, **48**, 1088–1095.

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. *Front. Hum. Neurosci.*, **7**, 263.

Larsson, J. & Smith, A.T. (2012) fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cereb. Cortex*, **22**, 567–576.

Lasaponara, S., Chica, A.B., Lecce, F., Lupianez, J. & Doricchi, F. (2011) ERP evidence for selective drop in attentional costs in uncertain environments: challenging a purely premotor account of covert orienting of attention. *Neuropsychologia*, **49**, 2648–2657.

Lee, T.S. & Mumford, D. (2003) Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am.*, **20**, 1434.

Linden, R.D., Picton, T.W., Hamel, G. & Campbell, K.B. (1987) Human auditory steady-state evoked potentials during selective attention. *Electroen. Clin. Neuro.*, **66**, 145–159.

Macaluso, E. & Doricchi, F. (2013) Attention and predictions: control of spatial attention beyond the endogenous-exogenous dichotomy. *Front. Hum. Neurosci.*, **7**, 685.

Makeig, S., Müller, M.M. & Rockstroh, B. (1996) Effects of voluntary movements on early auditory brain responses. *Exp. Brain Res.*, **110**, 487–492.

Martikainen, M.H., Kaneko, K. & Hari, R. (2005) Suppressed responses to self-triggered sounds in the human auditory cortex. *Cereb. Cortex*, **15**, 299–302.

May, P.J. & Tiitinen, H. (2010) Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, **47**, 66–122.

McCallum, W.C. (1980) Brain slow potential changes elicited by missing stimuli and by externally paced voluntary responses. *Biol. Psychol.*, **11**, 7–19.

McCarthy, G. & Donchin, E. (1976) The effects of temporal and event uncertainty in determining the waveforms of the auditory event related potential (ERP). *Psychophysiology*, **13**, 581–590.

McGuire, P.K., Silbersweig, D.A. & Frith, C.D. (1996) Functional neuroanatomy of verbal self-monitoring. *Brain*, **119**, 907–917.

Melcher, T., Weidema, M., Eenshuistra, R.M., Hommel, B. & Gruber, O. (2008) The neural substrate of the ideomotor principle: an event-related fMRI analysis. *NeuroImage*, **39**, 1274–1288.

Miall, R.C. & Wolpert, D.M. (1996) Forward models for physiological motor control. *Neural Networks*, **9**, 1265–1279.

Mill, R.W., Bőhm, T.M., Bendixen, A., Winkler, I. & Denham, S.L. (2013) Modelling the emergence and dynamics of perceptual organisation in auditory streaming. *PLoS Comput. Biol.*, **9**, e1002925.

Mock, J.R., Foundas, A.L. & Golob, E.J. (2011) Modulation of sensory and motor cortex activity during speech preparation. *Eur. J. Neurosci.*, **33**, 1001–1011.

Müller-Preuss, P. & Ploog, D. (1981) Inhibition of auditory cortical neurons during phonation. *Brain Res.*, **215**, 61–76.

Mustovic, H., Scheffler, K., Di Salle, F., Esposito, F., Neuhoff, J.G., Hennig, J. & Seifritz, E. (2003) Temporal integration of sequential auditory events: silent period in sound pattern activates human planum temporale. *NeuroImage*, **20**, 429–434.

Näätänen, R. (1982) Processing negativity: an evoked-potential reflection of selective attention. *Psychol. Bull.*, **92**, 605–640.

Näätänen, R. (1990) The role of attention in auditory by event-related potentials and other brain measures of cognitive function. *Behav. Brain Sci.*, **13**, 201–288.

Näätänen, R. & Picton, T. (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, **24**, 375–425.

Näätänen, R., Gaillard, A.W. & Mäntysalo, S. (1978) Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.*, **42**, 313–329.

Näätänen, R., Alho, K. & Schröger, E. (2002) Electrophysiology of attention. In Pashler, H. & Wixted, H. (Eds), *Stevens' Handbook of Experimental Psychology*: 3rd Edn. *vol 4: Methodology in Experimental Psychology*. John Wiley, New York, pp. 601–653.

Nittono, H. (2005) Missing-stimulus potentials associated with a disruption of human-computer interaction. *Psychologia*, **48**, 93–101.

Nittono, H. & Sakata, A. (2009) Effects of instructions on event-related brain potentials elicited by unresponsive computers. *Jpn. J. Phys. Psych. Psychophysiol.*, **27**, 215–223.

Niziolek, C.A., Nagarajan, S.S. & Houde, J.F. (2013) What does motor efference copy represent? Evidence from speech production *J. Neurosci.*, **33**, 16110–16116.

Nobre, K. & Coull, J.T. (2010) *Attention and Time.* Oxford University Press, Oxford.

Nordby, H., Hammerborg, D., Roth, W.T. & Hugdahl, K. (1994) ERPs for infrequent omissions and inclusions of stimulus elements. *Psychophysiology*, **31**, 544–552.

Numminen, J. & Curio, G. (1999) Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neurosci. Lett.*, **272**, 29–32.

Numminen, J., Salmelin, R. & Hari, R. (1999) Subject's own speech reduces reactivity of the human auditory cortex. *Neurosci. Lett.*, **265**, 119–122.

Okamoto, H., Stracke, H., Wolters, C.H., Schmael, F. & Pantev, C. (2007) Attention improves population-level frequency tuning in human auditory cortex. *J. Neurosci.*, **27**, 10383–10390.

Otsuka, A., Tamaki, Y. & Kuriki, S. (2008) Neuromagnetic responses in silence after musical chord sequences. *NeuroReport*, **19**, 1637–1641.

Ott, C.G. & Jäncke, L. (2013) Processing of self-initiated speech-sounds is different in musicians. *Front. Hum. Neurosci.*, **7**, 41.

Parmentier, F.B., Elsley, J.V., Andrés, P. & Barceló, F. (2011) Why are auditory novels distracting? Contrasting the roles of novelty, violation of expectation and stimulus change *Cognition*, **119**, 374–380.

Paus, T., Perry, D.W., Zatorre, R.J., Worsley, K.J. & Evans, A.C. (1996) Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur. J. Neurosci.*, **8**, 2236–2246.

Penney, T.B. (2004) Electrophysiological correlates of interval timing in the Stop-Reaction-Time task. *Brain Res. Cogn. Brain Res.*, **21**, 234–249.

Picton, T.W. & Hillyard, S.A. (1974) Human auditory evoked potentials. II. Effects of attention. *Electroen. Clin. Neuro.*, **36**, 191–199.

Polich, J. (2007) Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.*, **118**, 2128–2148.

Posner, M.I. (1980) Orienting of attention. *Q. J. Exp. Psychol.*, **32**, 3–25.

Posner, M.I. (2014) Orienting of attention: then and now. *Q. J. Exp. Psychol.*, doi: 10.1080/17470218.2014.937446. [Epub ahead of print].

Raij, T., McEvoy, L., Mäkelä, J.P. & Hari, R. (1997) Human auditory cortex is activated by omissions of auditory stimuli. *Brain Res.*, **745**, 134–143.

Rao, R.P. & Ballard, D.H. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.*, **2**, 79–87.

Rauss, K. & Pourtois, G. (2013) What is bottom-up and what is top-down in predictive coding? *Front. Psychol.*, **4**, 276.

Reznik, D., Henkin, Y., Schadel, N. & Mukamel, R. (2014) Lateralized enhancement of auditory cortex activity and increased sensitivity to self-generated sounds. *Nat. Commun.*, **5**, 4059.

Ross, J., Morrone, M.C., Goldberg, M.E. & Burr, D.C. (2001) Changes in visual perception at the time of saccades. *Trends Neurosci.*, **24**, 113–121.

Ross, B., Herdman, A.T. & Pantev, C. (2005) Stimulus induced desynchronization of human auditory 40-Hz steady-state responses. *J. Neurophysiol.*, **94**, 4082–4093.

Ruchkin, D.S., Sutton, S., Munson, R., Silver, K. & Macar, F. (1981) P300 and feedback provided by absence of the stimulus. *Psychophysiology*, **18**, 271–282.

Salomon, G. & Starr, A. (1963) Electromyography of middle ear muscles in man during motor activities. *Acta Neurol. Scand.*, **39**, 161–168.

SanMiguel, I., Saupe, K. & Schröger, E. (2013a) I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted "what" but not "when". *Front. Hum. Neurosci.*, **7**, 407.

SanMiguel, I., Todd, J. & Schröger, E. (2013b) Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. *Psychophysiology*, **50**, 334–343.

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N. & Schröger, E. (2013c) Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. *J. Neurosci.*, **33**, 8633–8639.

Saupe, K., Schröger, E., Andersen, S.K. & Müller, M.M. (2009) Neural mechanisms of intermodal sustained selective attention with concurrently presented auditory and visual stimuli. *Front. Hum. Neurosci.*, **3**, 58.

Saupe, K., Widmann, A., Trujillo-Barreto, N.J. & Schröger, E. (2013) Sensorial suppression of self-generated sounds and its dependence on attention. *Int. J. Psychophysiol.*, **90**, 300–310.

Schafer, E.W. & Marcus, M.M. (1973) Self-stimulation alters human sensory brain responses. *Science*, **181**, 175–177.

Scheich, H., Brechmann, A., Brosch, M., Budinger, E. & Ohl, F.W. (2007) The cognitive auditory cortex: task-specificity of stimulus representations. *Hearing Res.*, **229**, 213–224.

Schneider, D.M., Nelson, A. & Mooney, R. (2014) A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, **513**, 189–194.

Schröger, E. (1993) Event-related potentials to auditory stimuli following transient shifts of spatial attention in a Go/Nogo task. *Biol. Psychol.*, **36**, 183–207.

Schröger, E. (1994) Human brain potential signs of selection by location and frequency in an auditory transient attention situation. *Neurosci. Lett.*, **173**, 163–166.

Schröger, E. & Eimer, M. (1993) Effects of transient spatial attention on auditory event-related potentials. *NeuroReport*, **4**, 588–590.

Schröger, E. & Eimer, M. (1996) Effects of lateralized cues on the processing of lateralized auditory stimuli. *Biol. Psychol.*, **43**, 203–226.

Schröger, E. & Eimer, M. (1997) Endogenous covert spatial orienting in audition: "cost-benefit" analyses of reaction times and event related potentials. *Q. J. Exp. Psychol.*, **50**, 457–474.

Schröger, E., SanMiguel, I. & Bendixen, A. (2013) Prädiktive Modellierung in der auditiven Wahrnehmung. In Schröger, E. & Koelsch, S. (Eds), *Affektive Und Kognitive Neurowissenschaft (Enzyklopädie Der Psychologie Serie II: Kognition, Band 5)*. Hogrefe Verlag, Göttingen, pp. 11–45.

Schröger, E., Bendixen, A., Denham, S.L., Mill, R.W., Böhm, T.M. & Winkler, I. (2014) Predictive regularity representations in violation detection and auditory stream segregation: from conceptual to computational models. *Brain Topogr.*, **27**, 565–577.

Schwartze, M., Tavano, A., Schröger, E. & Kotz, S.A. (2012) Temporal aspects of prediction in audition: cortical and subcortical neural mechanisms. *Int. J. Psychophysiol.*, **83**, 200–207.

Seiss, E., Gherri, E., Eardley, A.F. & Eimer, M. (2007) Do ERP components triggered during attentional orienting represent supramodal attentional control? *Psychophysiology*, **44**, 987–990.

Serences, J.T. & Kastner, S. (2014) A multi-level account of selective attention. In Nobre, A.C. & Kastner, S. (Eds), *The Oxford Handbook of Attention*. Oxford University Press, Oxford, pp. 76–104.

Simson, R., Vaughan, H.G. & Ritter, W. (1976) The scalp topography of potentials associated with missing visual or auditory stimuli. *Electroen. Clin. Neuro.*, **40**, 33–42.

Sowman, P.F., Kuusik, A. & Johnson, B.W. (2012) Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. *Exp. Brain Res.*, **222**, 149–157.

Sperry, R.W. (1950) Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psych.*, **43**, 482–489.

Stapleton, J.M., O'Reilly, T. & Halgren, E. (1987) Endogenous potentials evoked in simple cognitive tasks: scalp topography. *Int. J. Neurosci.*, **36**, 75–87.

Summerfield, C. & Egner, T. (2009) Expectation (and attention) in visual cognition. *Trends Cogn. Sci.*, **13**, 403–409.

Summerfield, C. & Egner, T. (2014) Attention and decision-making. In Nobre, A.C. & Kastner, S. (Eds), *The Oxford Handbook of Attention*. Oxford University Press, Oxford, pp. 837–864.

Sutton, S., Tueting, P., Zubin, J. & John, E.R. (1967) Information delivery and the sensory evoked potential. *Science*, **155**, 1436–1439.

Synofzik, M., Vosgerau, G. & Newen, A. (2008) Beyond the comparator model: a multifactorial two-step account of agency. *Conscious. Cogn.*, **17**, 219–239.

Talsma, D., Kok, A., Slagter, H.A. & Cipriani, G. (2008) Attentional orienting across the sensory modalities. *Brain Cognition*, **66**, 1–10.

Tapia, M.C., Cohen, L.G. & Starr, A. (1987) Attenuation of auditory-evoked potentials during voluntary movement in man. *Audiology*, **26**, 369–373.

Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N. & Näätänen, R. (1994) Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biol. Psychol.*, **38**, 157–167.

Tian, X. & Poeppel, D. (2013) The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *J. Cognitive Neurosci.*, **25**, 1020–1036.

Tiitinen, H.T., Sinkkonen, J., Reinikainen, K., Alho, K., Lavikainen, J. & Näätänen, R. (1993) Selective attention enhances the auditory 40-Hz transient response in humans. *Nature*, **364**, 59–60.

Timm, J., SanMiguel, I., Saupe, K. & Schröger, E. (2013) The N1-suppression effect for self-initiated sounds is independent of attention. *BMC Neurosci.*, **14**, 2.

Timm, J., SanMiguel, I., Keil, J., Schröger, E. & Schönwiesner, M. (2014a) Motor intention determines sensory attenuation of brain responses to self-initiated sounds. *J. Cognitive Neurosci.*, **26**, 1481–1489.

Timm, J., Schönwiesner, M., SanMiguel, I. & Schröger, E. (2014b) Sensation of agency and perception of temporal order. *Conscious. Cogn.*, **23**, 42–52.

Todorovic, A., van Ede, F., Maris, E. & de Lange, F.P. (2011) Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *J. Neurosci.*, **31**, 9118–9123.

Towle, V.L., Yoon, H.A., Castelle, M., Edgar, J.C., Biassou, N.M., Frim, D.M., Spire, J.P. & Kohrman, M.H. (2008) ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. *Brain*, **131**, 2013–2027.

Treisman, A. (1964) Monitoring and storage of irrelevant messages in selective attention. *J. Verb. Learn. Verb. Be.*, **3**, 449–459.

Valsecchi, M. & Turatto, M. (2009) Microsaccadic responses in a bimodal oddball task. *Psychol. Res.*, **73**, 23–33.

Ventura, M.I., Nagarajan, S.S. & Houde, J.F. (2009) Speech target modulates speaking induced suppression in auditory cortex. *BMC Neurosci.*, **10**, 58.

Voisin, J., Bidet-Caulet, A., Bertrand, O. & Fonlupt, P. (2006) Listening in silence activates auditory areas: a functional magnetic resonance imaging study. *J. Neurosci.*, **26**, 273–278.

Vroomen, J. & Stekelenburg, J.J. (2010) Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *J. Cognitive Neurosci.*, **22**, 1583–1596.

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. *Proc. Natl. Acad. Sci. USA*, **108**, 20754–20759.

Wacongne, C., Changeux, J.P. & Dehaene, S. (2012) A neuronal model of predictive coding accounting for the mismatch negativity. *J. Neurosci.*, **32**, 3665–3678.

van Wassenhove, V., Grant, K.W. & Poeppel, D. (2005) Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. USA*, **102**, 1181–1186.

Waszak, F., Cardoso-Leite, P. & Hughes, G. (2012) Action effect anticipation: neurophysiological basis and functional consequences. *Neurosci. Biobehav. R.*, **36**, 943–959.

Weinberg, H., Walter, W.G., Cooper, R. & Aldridge, V.J. (1974) Emitted cerebral events. *Electroen. Clin. Neuro.*, **36**, 449–456.

Wetzel, N. & Schröger, E. (2014) On the development of auditory distraction: a review. *PsyCh J.*, **3**, 72–91.

Widmann, A., Engbert, R. & Schröger, E. (2014) Microsaccadic responses indicate fast categorization of sounds: a novel approach to study auditory cognition. *J. Neurosci.*, **34**, 11152–11158.

Williams, S.R. & Chapman, C.E. (2000) Time course and magnitude of movement-related gating of tactile detection in humans II. Effects of stimulus intensity. *J. Neurophysiol.*, **84**, 863–875.

Williams, S.R. & Chapman, C.E. (2002) Time course and magnitude of movement-related gating of tactile detection in humans III. Effect of motor tasks. *J. Neurophysiol.*, **88**, 1968–1979.

Williams, S.R., Shenasa, J. & Chapman, C.E. (1998) Time course and magnitude of movement-related gating of tactile detection in humans I. Importance of stimulus location. *J. Neurophysiol.*, **79**, 947–963.

Winkler, I., Denham, S.L. & Nelken, I. (2009) Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci.*, **13**, 532–540.

Wise, R.J., Greene, J., Büchel, C. & Scott, S.K. (1999) Brain regions involved in articulation. *Lancet*, **353**, 1057–1061.

Wu, C.T., Weissman, D.H., Roberts, K.C. & Woldorff, M.G. (2007) The neural circuitry underlying the executive control of auditory spatial attention. *Brain Res.*, **1134**, 187–198.

Yabe, H., Tervaniemi, M., Reinikainen, K. & Näätänen, R. (1997) Temporal window of integration revealed by MMN to sound omission. *NeuroReport*, **8**, 1971–1974.

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R.J. & Näätänen, R. (1998) Temporal window of integration of auditory information in the human brain. *Psychophysiology*, **35**, 615–619.

Yuval-Greenberg, S. & Deouell, L.Y. (2011) Scalp-recorded induced γ-band responses to auditory stimulation and its correlations with saccadic muscle-activity. *Brain Topogr.*, **24**, 30–39.

Zhao, J., Al-Aidroos, N. & Turk-Browne, N.B. (2013) Attention is spontaneously biased toward regularities. *Psychol. Sci.*, **24**, 667–677.