International Journal of Psychophysiology 83 (2012) 120–131

Contents lists available at SciVerse ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Review

Early electrophysiological indicators for predictive processing in audition: A review☆,☆☆

Alexandra Bendixen *, Iria SanMiguel, Erich Schröger

Institute for Psychology, University of Leipzig, Germany

ARTICLE INFO

Article history:
Received 1 June 2011
Received in revised form 28 July 2011
Accepted 8 August 2011
Available online 23 August 2011

Keywords:
Auditory information processing
Predictive modeling
Match detection
Mismatch detection
Event-related potential (ERP)
Oscillatory activity

ABSTRACT

The auditory system essentially deals with sequential type of input and thus requires processing that is particularly suited to extract stimulus relations within a sequence. Evidence from a variety of paradigms converges to show that the auditory system automatically uses stimulus predictability for facilitating its sequential processing. This type of predictive processing does not require attentional processing of the sounds or cognitive control of the predictions, nor does it involve the preparation of motor responses to the auditory stimuli. We will present a taxonomy of paradigms and resulting electrophysiological indicators for such automatic predictive processing in terms of event-related potential components and oscillatory activity. These indicators will include signals of fulfilled predictions (match signals such as N1 attenuation, repetition positivity, and early evoked gamma band response enhancement) as well as signals of violated predictions (mismatch signals such as the mismatch negativity and stimulus omission responses). We will show how recent approaches have revealed particularly early indicators of predictive processing down to the level of the auditory middle-latency responses. We will discuss the strength of the various indicators in terms of a truly predictive account of auditory processing (as opposed to, e.g., a retrospective verification of predictions). Finally, we will discuss the benefits of a predictive system within and beyond auditory processing. In conclusion, we argue in favor of the overwhelming evidence for predictions in audition, flexibly instantiated on different levels and timescales, and we aim to provide guidance along a variety of research paradigms illustrating the existence of these predictions.

1. Introduction

The concept of predictive processing has a long tradition in psychology, albeit under different terms. For instance, cognitive psychology has long been investigating the concept of mental models (Craik, 1943). Mental models can be conceived as internal representations of reality and play an important role in making inferences and decisions (Johnson-Laird, 1983). By means of mental models, we simulate our reality. This allows us to anticipate, for instance, the consequences of a given action. It also helps us to prepare for future events, such as the rain that is to be expected, or the bad mood of our spouse. We can then take appropriate action to prevent undesirable effects (e.g., getting wet by the rain) or even to change the future in such a way that the bad mood of our spouse will not occur in the first place. In other words, the ability of generating predictions enables goal-directed behavior that goes much beyond fix stimulus–response associations. This is the basis of an enormous flexibility in interacting with our physical and social environment.

The concept of mental models is often associated with higher cognitive functions such as problem solving or language production. Yet the same notion is widespread in research on perception as a seemingly lower cognitive function. Examples are the theory of unconscious inferences of Hermann von Helmholtz (1867) or Irvin Rock (1983), according to which perception rests on logical inferences based on internal representations. However, these perceptual inferences are seen as retrospective rather than prospective in nature. More recently, the proactive or predictive aspect of mental models is increasingly discussed in perception as well (Friston, 2005; Baldeweg, 2006; Bar, 2007; Winkler et al., 2009; see also Gregory, 1980). Perception often needs to make extrapolations to facts

☆ Contribution to the Special Issue entitled "Predictive information processing in the brain: Principles, neural mechanisms and models" edited by J. Todd, E. Schröger, and I. Winkler.

☆☆ This paper is based on a German textbook chapter (Schröger E., SanMiguel, I., & Bendixen A., in press, Prädiktive Modellierung in der auditiven Wahrnehmung. In. E. Schröger & S. Koelsch (Eds.). Kognitive und Affektive Neurowissenschaften. Enzyklopädie der Psychologie (Serie II: Kognition, Band 9). Göttingen: Hogrefe).

* Corresponding author at: Institute for Psychology, University of Leipzig, Seeburgstraße 14-20, D-04103 Leipzig, Germany. Tel.: +49 341 9735907; fax: +49 341 9735969.
E-mail addresses: alexandra.bendixen@uni-leipzig.de (A. Bendixen), iria.sanmiguel@uni-leipzig.de (I. SanMiguel), schroger@uni-leipzig.de (E. Schröger).

0167-8760/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpsycho.2011.08.003

that have no immediate sensory basis. For instance in vision, objects are represented as complete even when not all of their elements are projected to the retina due to occlusion.

Recently, predictive modeling in perception has been described as the result of an interaction between processing of the current sensory information and the information that was known before (cf. Cheung and Bar, 2012-this issue). This process can be conceived as an evidence-based prediction or expectation of what might be present in the world. By an iterative comparison of information, the amount of anticipated events becomes increasingly smaller, until finally a positive identification is possible. Thus, two important determinants are the context into which the incoming stimulus is embedded and the contents of long-term memory (Bar, 2007).

Extrapolation towards future events is particularly helpful for signals that are highly variable in time (i.e., that need to be analyzed on-line) and in which relevant information is contained in the relations between successive events. In audition, the information that needs to be extracted is present exclusively with high variation over time, and missing information typically is the one that lies in the future (while in vision, missing information often is the one occluded by other objects). This is due to the nature of sound propagation and the construction of the hearing apparatus. The assumption of predictive modeling thus seems particularly plausible in the auditory modality.

Here, we will describe central experimental paradigms as well as important indicators and findings that reveal such predictive processing in audition. Throughout the contribution, we will use the terms “prediction”, “expectation”, “anticipation”, and “extrapolation” in a largely synonymous manner. Before we turn to describe specific paradigms, let us first advance an operational definition of what shall be understood by predictive modeling in audition in the present text. The type of predictive processing we refer to does not require attentional processing of the sounds or cognitive control of the predictions, nor does it involve the preparation of motor responses to the auditory stimuli. It is not the outcome of a mental model in the sense described in the opening of this contribution. Although it can, in most of the cases, be performed also at will by employing attentional resources and exerting active cognitive control, neither of these is a necessary ingredient. Instead, the type of predictive modeling of the auditory input we regard in this paper is an automatic process, and as such it cannot be switched off. Importantly, predicting entails more than just the extraction of a rule: This information has then to be used to generate an actual, more or less specific, prediction; a separate and independent entity which will occur ahead of or synchronously with the predicted stimulus. Thus, the process we will refer to as predictive modeling cannot be a retrospective or point by point comparison of the present stimulation to a stored template of some kind, even though such templates may indeed often play a role in the process of generating predictions. Finally, in order to provide solid evidence of prediction, we will ask the observed effects to be more than a difference between change and repetition, as no actual prediction is strictly necessary for such a difference to exist. For this purpose, we may propose a rule of thumb: A true measure of prediction violation should prove to be present to unexpected repetitions, while it should not be elicited by expected changes.

2. Empirical evidence for the existence of auditory predictions

Various paradigms have been developed in order to probe the existence of auditory predictions. These paradigms differ on a number of dimensions. For instance, some of them involve attentional allocation to the sounds while others do not, some paradigms are using purely auditory stimuli while others include multimodal links, and so on. All paradigms have in common that they induce a certain form of auditory predictability and then probe whether the processing of upcoming stimuli is modulated by their conformity with the prediction — if it is, one can infer that a prediction has been generated, or at least the predictability must have been detected to allow for a post-hoc check of congruence. Comparisons are typically made between prediction-conforming and prediction-violating stimuli, or between stimuli with different degrees of certainty of the prediction. Fig. 1 provides an overview of the various sources of inducing predictability (left part of Fig. 1) and the psychophysiological indicators of predictive processing (right part of Fig. 1). Predictability can be induced in various ways: by means of the history of auditory stimulation which is temporarily stored in auditory sensory memory (e.g., the next tick of a clock is predictable after a few seconds' exposure), by entries in auditory long-term memory (e.g., the continuation of many tunes is well-known), by cues from other modalities (e.g., watching a window shut by the wind causes the expectation of a loud noise), or by means of self-generating an auditory stimulus (e.g., clapping one's hands causes the expectation of a well-known sound). The psychophysiological indicators may directly reflect the predictive activity of the brain under certain circumstances (e.g., the P50 component of the event-related potential can be elicited without an actual stimulus input). In other situations, the psychophysiological indicators may indirectly reflect the outcome of a comparison process between the prediction and the representation of the actually experienced stimulus. If the comparison results in a congruence, *match* signals are elicited (e.g., the *repetition positivity*, RP). If the comparison results in an incongruence, *mismatch* signals are elicited (e.g., the *mismatch negativity*, MMN). Some of these psychophysiological indicators and the paradigms in which they can be obtained will be presented in the following sections on the basis of exemplary studies. Of course there is a considerably larger number of studies investigating the processing of sounds that are conforming or non-conforming with a prediction. Most of them can, however, be integrated into the scheme presented here.

2.1. Match paradigms: Processing of regularity-conforming stimuli

One approach towards predictive modeling is to investigate the effects of the increasing predictability of a stimulus on the attentive and preattentive processing of that stimulus. In the so-called *roving-standard* paradigm, Haenschel et al. (2005) presented a tone of a specific frequency 2, 6, or 36 times and then replaced it by a tone of a different frequency, which again was presented 2, 6, or 36 times, and so on. The repeatedly presented tone is considered a *standard* stimulus, while the first tone with a frequency change is considered a *deviant* stimulus. If the system learns to “expect” the occurrence of repetition in this paradigm, the event-related potentials (ERPs) elicited by standard and deviant stimuli should differ — which indeed they do, as we shall consider in Section 2.2. Here we focus on the processing of the standard stimuli per se. The confidence in expecting a standard should increase with the number of repetitions. Haenschel and colleagues investigated whether this is the case by comparing the ERPs elicited by standard stimuli in positions 2, 6, and 36. They found a positivity over frontocentral electrodes that was increasing with the number of repetitions — the so-called repetition positivity (RP) (for illustration, see Fig. 2). The RP was observed in a latency range from 50 to 250 ms after tone onset, and was present both when participants were attending the tones and when their attention was diverted from the tones by means of a muted video. Haenschel et al. (2005) interpret the RP as an ERP correlate of fast stimulus-specific adaptation underlying the formation of a sensory memory representation. They suggest that the neural processing of a tone is increasingly suppressed with the number of repetitions (see also Baldeweg, 2006). From the data obtained in the roving-standard paradigm, it remains unclear whether the underlying mechanism is of predictive or retrospective nature. One may argue that the RP is a post-hoc phenomenon caused by a fast comparison of an incoming stimulus with the sensory

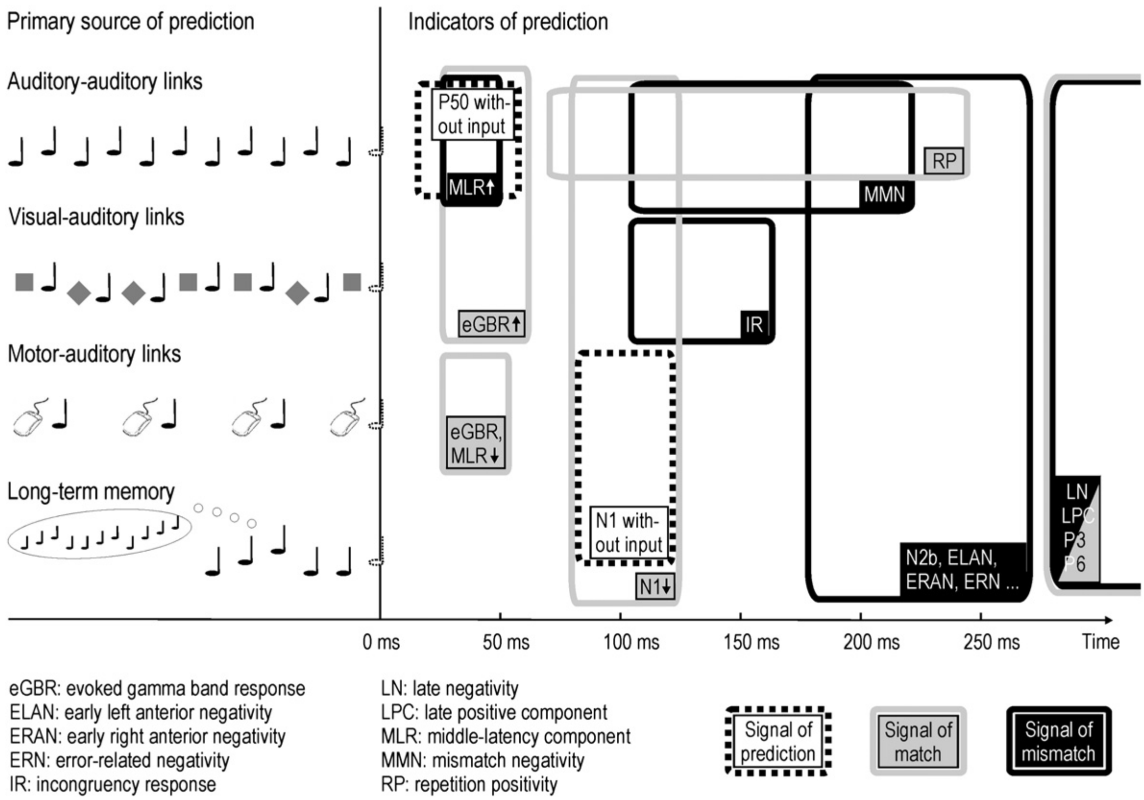

Fig. 1. Left: Classification of sources for auditory predictions. A prediction can be based on regularities extracted from an auditory sequence (auditory–auditory links), on systematic contingencies between visual and auditory stimuli (visual–auditory links), on relations between a motor act and the resulting sensory effects (motor–auditory links), and on knowledge acquired through experience (long-term memory). Although some predictions are generated on the fly, based on sensory or motor representations, of course memory processes are engaged in these cases as well. Right: Electrophysiological effects that are regarded as evidence in favor of predictive processing — either as direct correlates of anticipation (brain activity without an incoming stimulus) or indirectly as consequences of match or mismatch results of a comparison process between predicted and actual auditory event. The effects are classified, on the one hand, by the latency of their elicitation (from early effects after ca. 25 ms until relatively late effects from 300 to 500 ms after sound onset). On the other hand, effects are classified by the source of prediction: Some effects (e.g., N1 attenuation) have been demonstrated for all sources of auditory predictions, others (e.g., repetition positivity, RP) have so far been demonstrated only for certain sources. Adapted from Schröger E., SanMiguel, I., & Bendixen A., 2012, Prädiktive Modellierung in der auditiven Wahrnehmung. In. E. Schröger & S. Koelsch (Eds.). Kognitive und Affektive Neurowissenschaften. Enzyklopädie der Psychologie (Serie II: Kognition, Band 9). Göttingen: Hogrefe; with permission of Hogrefe Publisher.

regularity extracted from the previous stimuli, rather than by an extrapolation of the regularity towards the next incoming stimulus.

Fig. 2 illustrates that the presentation of a tone will lead to a characteristic auditory ERP including the N1 and P2 components (middle row of Fig. 2). This situation is denoted as "without the existence of a predictive model", which means that the current tone with its specific feature values (frequency, intensity, duration etc.) has no specific history of having been presented several times before or being otherwise expectable from the stimulus context. In contrast, when a tone or tone feature value is repeatedly presented, an auditory model is generated that predicts the re-occurrence of this feature value. This situation is depicted in the top row of Fig. 2 ("with predictive model"). If an incoming stimulus corresponds with the predicted stimulus, the ERP elicited by this stimulus will be positive (repetition positivity, RP) relative to the ERP elicited by the same stimulus without the existence of a predictive model. The ERPs are shown in the column "Observable brain response". The column "Indicators of predictive processing" shows the RP as the difference wave (black line) obtained by subtracting the ERP elicited by a stimulus without the existence of a predictive model (dotted line) from the ERP elicited by the same stimulus with the existence of a predictive model (gray line).

The processing of stimuli conforming to a regularity established by the immediate auditory history has also been investigated by Schadow et al. (2009). However, they used an abstract regularity (instead of the repetition of a specific frequency), and they measured oscillatory activity in the gamma band as an additional variable. In their study, the regularity was defined by sequences of six tones that were either rising or falling in frequency throughout. The regularity should be established more firmly in later positions within the sequence than in earlier positions. If the RP is based on prediction and not on simple stimulus repetition effects (e.g., increased refractoriness of frequency-specific neurons), it should be observed in the design of Schadow et al. (2009) as well. Although no explicit analysis was carried out by the authors, a comparison of the ERPs elicited by regularity-conforming tones in positions 3 and 5 of the sequence shows clear signs of RP (see Fig. 2 in Schadow et al., 2009).

RP has been reported with even more complex regularities as well (Bendixen et al., 2008). It should be mentioned, however, that some other studies have failed to observe RP (e.g., Bendixen et al., 2007; Ylinen and Huotilainen, 2007) although their stimulus protocol was quite similar to the original one employed by Haenschel et al. (2005). On the other hand, Baldeweg et al. (2004) as well as Costa-Faidella et al. (2011) obtained findings in line with Haenschel et al. (2005). Thus the conditions necessary and sufficient to elicit RP are still a matter of investigation.

The study of Schadow et al. (2009) was not designed towards analyzing RP but focused on evoked gamma band activity. This is an early (ca. 40–100 ms after stimulus onset) response of the brain that is maximal in a range of about 40 Hz (therefore termed 40-Hz response or gamma band response; for a review, see Kaiser and Lutzenberger,

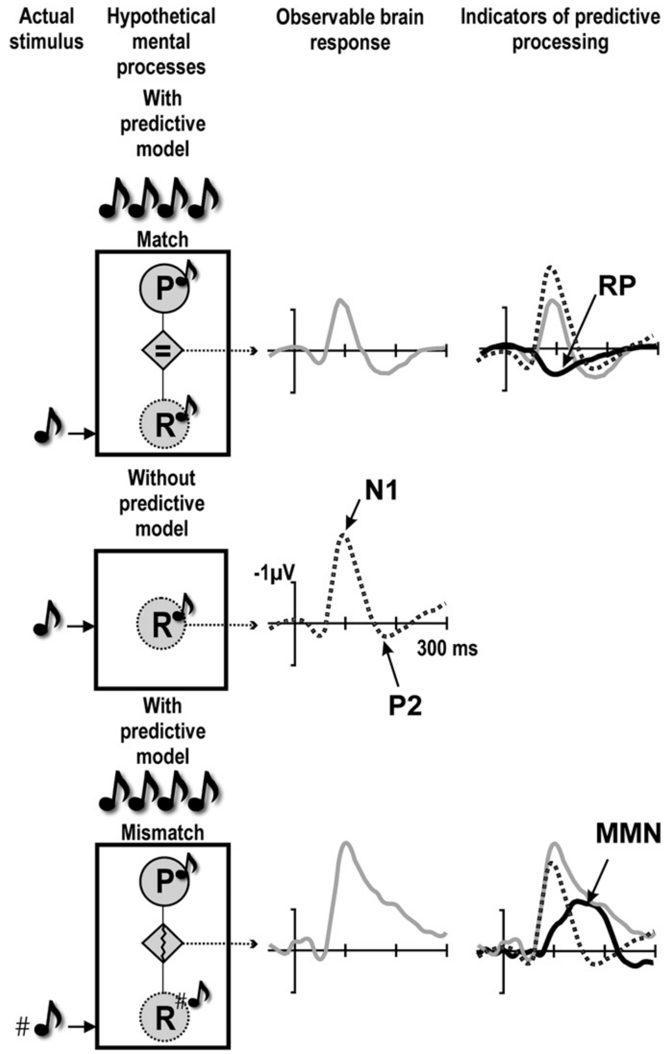

Fig. 2. Processing of an auditory stimulus depending on the existence of a predictive model and its match with the predictive model. The experimentally manipulated "current stimulus" is shown on the left. The two columns on the right show the experimentally measurable responses of the system as raw responses in the various conditions ("Observable brain response") and as visualization of the prediction effect by forming difference waves between conditions ("Indicators of predictive processing"; strong black lines). In this case, the brain responses are event-related potentials (ERPs) or differences between ERPs. The x axis of the voltage-over-time plot shows the time range from 50 ms preceding stimulus onset to 300 ms after stimulus onset. The y axis shows voltage in µV. The indicators of predictive processing, *repetition positivity* (RP) and *mismatch negativity* (MMN), are marked with arrows. The "hypothetical mental processes" between current stimulus and brain response are putative processing stages in a system operating with predictive models. "P" denotes the prediction generated by the model regarding a specific feature of the stimulus (e.g., its frequency). "R" denotes the representation of this feature extracted from the sensory input. If prediction and representation are identical, the system detects a *match*, which typically shows as a positivity in the measurable brain responses (upper row). If prediction and representation are not in agreement, the system detects a prediction violation (*mismatch*), which typically shows as a negativity (lower row). For the stimulus in the middle row, no relevant predictive model exists. The stimulus is thus unpredictable regarding the relevant feature "tone frequency". Therefore, its actual frequency cannot be compared against a regularity or prediction. Strictly speaking, a comparison condition without any predictive model should always be implemented in order to distinguish between indicators of match and mismatch. However, this is not always experimentally feasible. Adapted from Schröger E., SanMiguel, I., & Bendixen A., 2012, Prädiktive Modellierung in der auditiven Wahrnehmung. In. E. Schröger & S. Koelsch (Eds.). Kognitive und Affektive Neurowissenschaften. Enzyklopädie der Psychologie (Serie II: Kognition, Band 9). Göttingen: Hogrefe; with permission of Hogrefe Publisher.

2005). Indeed, the amplitude of the gamma band response elicited by regularity-conforming tones was clearly enhanced as early as 50 ms after tone onset. Such enhancement was not found for regularity-violating tones. The authors interpret this finding as an indicator of an early match process between the incoming stimulus and the memory trace for the expected tone (cf. Figs. 1 and 3). Although this comes closer to a retrospective explanation, the underlying mechanism could also be conceived in a predictive manner. As in the case of RP, the empirical results do not allow for a dissociation of these two possibilities, although the early onset argues for the predictive account.

In contrast to RP, the effect on the gamma band response reported by Schadow et al. (2009) was not modulated by the position of the stimulus within the sequence. This might suggest that the RP and the evoked gamma band response reflect different aspects of processing regularity-conforming stimuli. An alternative explanation is that the modulation by position is obtained only for simple repetition regularities, and is then due to refractoriness contributions. Evidence and arguments along these lines have been put forward by Bendixen and Schröger (2008), Horváth et al. (2008), as well as Sculthorpe and Campbell (in press).

In the study of Schadow et al. (2009), the expectation of a particular tone was induced by the current auditory sequence. Widmann et al. (2007) were able to show that auditory expectations, as measured by evoked gamma band responses, can be generated even based on visual context. In this study, participants were shown a visual pattern on a screen consisting of four to six rectangles. Each of these rectangles appeared on one of two vertical positions, indicating low or high frequency of a tone in a melody that was played to participants after short exposure to the visual "score" (see Fig. 3). The melody contained as many tones as there were rectangles, and participants were asked to compare tone-by-tone whether the vertical position of the rectangle was consistent with the pitch of the corresponding tone. Responses were given at the end of each four- to six-tone sequence. In half of the sequences, one of the tones was incongruent with the visual symbol. Evoked gamma band responses were analyzed separately for tones that were congruent and tones that were incongruent with the visual symbol. Differences were observed as early as 17–70 ms after tone onset: A strong gamma band response was evoked for congruent but not for incongruent tones. Fig. 3 shows the observed brain responses in a time–frequency plot. The gamma band enhancement for congruent tones is marked by a black rectangle. In case of incongruent tones, no gamma response enhancement is obtained in the same time range (again marked by a rectangle in Fig. 3). Such early gamma band responses are seen as indicative of a match process between incoming sensory information and an activated memory representation (Herrmann et al., 2004). There is also a gamma band response enhancement in a later time range (dotted rectangle), which could reflect processing when the sound is perceived as matching the prediction. The results of Widmann et al. (2007) thus suggest that a match was detected in case of congruency, but not in case of incongruency. Given the very early onset of the response (after 17 ms), these results can best be explained by the generation of preparatory activity based on the visual symbolic information. The left-hemispheric preponderance of this early gamma-band effect suggests a stronger involvement of left- compared with right-hemispheric auditory areas. In a more generalized description, a prediction has been generated about the to-be-expected sound that is due to the learned association between the elevation of the visual rectangle and the pitch of the sound.

Fig. 3 shows not only the gamma band response, but also the ERPs that were elicited in the study of Widmann et al. (2007) in case of audio–visual congruence and incongruence. The ERP elicited by an incongruent stimulus differs from that elicited by a congruent stimulus in several latency ranges. In the present context, the negativity observed around 100–140 ms is relevant, as it probably indicates the detection of mismatch between the visually-based auditory prediction and the actual auditory stimulus. This response was termed the *incongruency response* (IR) (cf. Widmann et al., 2004). For completeness, we shall mention that the IR effect could also be explained by assuming the elicitation of an additional positivity in case of a match, rather than a negativity in case of a mismatch. For

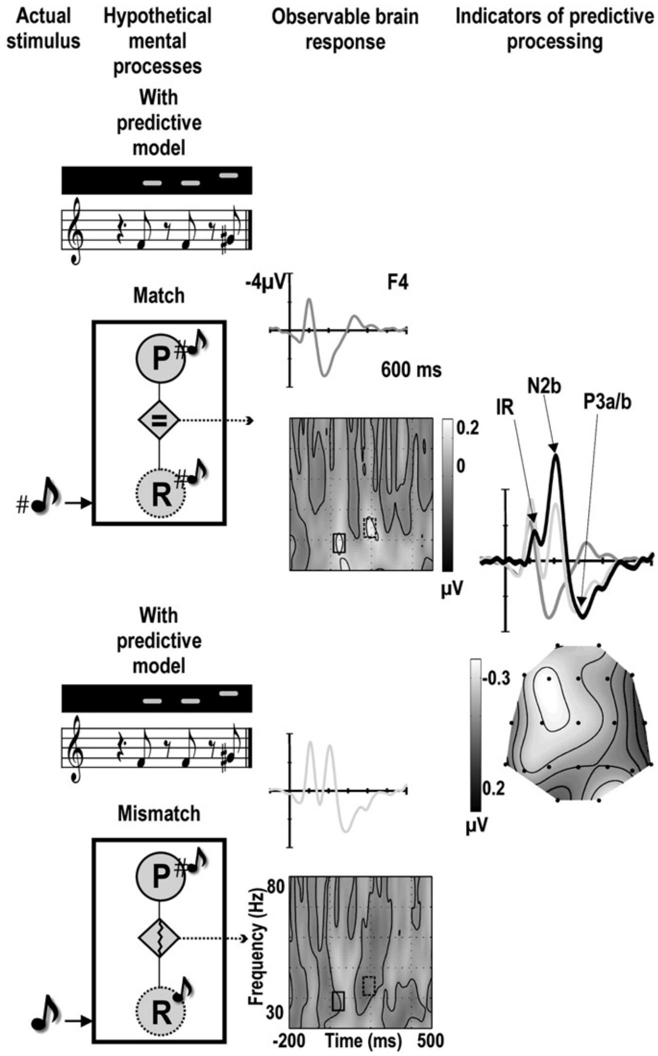

Fig. 3. Processing of an auditory stimulus depending on its congruence with a predictive model based on visual information (Widmann et al., 2007; Widmann et al., 2004). Indicators and processes as in Fig. 2. In contrast to Fig. 2, there is no neutral condition without a predictive model here. The ERPs show an *incongruency response* (IR) when auditory and visual information do not match (lower row). The IR bears resemblance to the MMN as a mismatch indicator based on purely auditory information (cf. Fig. 2). In addition, evoked gamma band activity was investigated in this experiment, showing an enhancement in case of a match between auditory and visual information (upper row). Such enhancement was also observed with purely auditory stimuli (Schadow et al., 2009), but so far only with attentive processing of the tones. In case of mismatch, the gamma band response is not enhanced (lower row; the regions in which enhancement was observed for congruent auditory and visual inputs are marked by rectangles). Differences in topographical distribution of the gamma band amplitudes are shown in the right column. Please note the differences in temporal resolution for the voltage–time plots (ERPs) and the frequency–time plots (gamma band responses). Adapted from Schröger E., SanMiguel, I., & Bendixen A., 2012, Prädiktive Modellierung in der auditiven Wahrnehmung. In. E. Schröger & S. Koelsch (Eds.). Kognitive und Affektive Neurowissenschaften. Enzyklopädie der Psychologie (Serie II: Kognition, Band 9). Göttingen: Hogrefe; with permission of Hogrefe Publisher.

distinguishing whether the IR reflects match- or mismatch-related processing, a neutral comparison condition is needed in which the ERP is measured without an underlying predictive model. Yet it is unlikely that the IR belongs to the indicators of matching processes as it bears much resemblance to the MMN as a classical indicator of mismatch, which we will discuss in the following Section 2.2.

2.2. Mismatch paradigms: Processing of regularity-violating stimuli

Experiments designed to investigate match processes, as described in the previous section, typically do not only present regularity-conforming stimuli but also regularity-violating stimuli. In the simplest case, one tone (the standard) is repeatedly presented and occasionally replaced by a different tone (the deviant). The difference usually refers to a change in one of the feature values (e.g., frequency, intensity, location, onset time, duration, and timbre.). Deviants typically elicit the *mismatch negativity* (MMN) component of the ERP, whether participants are attending the auditory stimuli or not (for detailed discussion of attention effects on the MMN, see Sussman, 2007). MMN has been reported in close to 2000 publications (for comprehensive reviews, see Kujala et al., 2007; Näätänen et al., 2007; Winkler, 2007). It consists in a frontocentrally negative and (with nose reference) posterolaterally positive enhancement in the ERP elicited by regularity-violating relative to regularity-conforming stimuli. MMN is observed from around 100–250 ms after onset of the deviation and is generated at least partially in auditory areas of supratemporal cortex (Giard et al., 1990; Molholm et al., 2005; Rosburg et al., 2005; Sabri et al., 2006; Shalgi and Deouell, 2007).

Fig. 2 illustrates the MMN and the underlying mechanism of elicitation. The MMN consists in a negativity of the ERP in case of mismatch both relative to the ERP in case of match and relative to the ERP elicited by the same stimulus when no predictive model concerning the critical stimulus feature exists (Schröger, 2007). The comparison with the situation without a relevant predictive model allows for the inference that the MMN is indeed a mismatch indicator (as opposed to a change in processing for the regularity-conforming stimulus).

MMN is often interpreted as the result of a comparison process detecting the mismatch between a memory trace of the currently valid auditory regularities and the representation of the actual stimulus (Schröger, 2007; Winkler, 2007; see, however, May and Tiitinen, 2010, for a partly different account). Since all the investigated regularities are of predictive nature, it is suggestive to interpret the MMN as a result of prediction violation. However, predictability alone does not justify the inference that a prediction was indeed made, and that the MMN is based on this prediction. It could be that the comparison process operates post-hoc: The incoming stimulus might be compared with the regularity once it has occurred, and not with a prediction made from the regularity. Thus once more, it remains unclear whether the mechanism is truly predictive or "just" retrospective. The retrospective interpretation seems possible when considering a recent study by Schwartze et al. (2011). They recorded MMN to frequency deviations in passive and active oddball conditions with isochronous (600 ms inter-stimulus interval, ISI) or random (200–1000 ms ISI) timing. There were no effects of random vs. isochronous timing on MMN in the passive condition. As one would expect a predictive system to perform better when there is no uncertainty about the timing of the stimuli, these results may be interpreted as arguing against a predictive account of MMN. However, as the MMN system mainly relies on temporal information of less than 400 ms (e.g. Grimm et al., 2006), one may not necessarily have expected an advantage of the isochronous ISI condition.

In any case, the elicitation of the MMN component directly indicates that a regularity violation was detected, and thus indirectly implies that the regularity was extracted. The range of extractable regularities is immense; the above example of the repetition of one stimulus and occasional deviations in a single feature is at the very simple end of a continuum of increasingly complex auditory stimulus configurations comprising more and more abstract regularities. Abstract regularities can be defined within single features or through combinations of features within single stimuli or across successive stimuli. For example, MMN is elicited when the frequency relation between successive tones in a pair (with the absolute frequencies varying across a huge range) is changed (e.g., Paavilainen et al., 1999; Schröger et al., 2007). The most complex regularity reported so far (at least in settings in which the auditory stimuli are not in the focus of attention) is a contingency rule connecting different features of temporally separate events (of the form, “the duration of the current stimulus predicts the frequency of the next stimulus”). Violations of this regularity lead to MMN elicitation (Bendixen et al., 2008; Paavilainen et al., 2007). For comprehensive reviews on the different kinds of abstract regularities whose violation has been shown to elicit MMN, see Näätänen et al. (2001, 2010).

In addition to the MMN, several other ERP components are elicited by violations of expectation on different levels of abstraction. The *incongruency response* (IR) first reported by Widmann et al. (2004) was already introduced above. The *early right anterior negativity* (ERAN) responds to violations of musical expectations (Koelsch et al., 2000; Rohrmeier and Koelsch, 2012-this issue). In the language domain, the *early left anterior negativity* (ELAN) reflects violations of syntactic rules (e.g., Hahne and Friederici, 1999). It occurs 150–200 ms after a syntax violation and shows an anterior distribution lateralized to the left hemisphere. Other linguistic violations are typically followed by electrophysiological indicators in later time ranges, such as the N400 that occurs ca. 250–400 ms after a semantically unexpected stimulus (Kutas and Hillyard, 1980), and the P600 (Osterhout, and Holcomb, 1992) or SPC (syntactic positive shift; Hagoort et al., 1993) which are elicited ca. 600 ms after phrase structure violations in spoken sentences (e.g., Friederici, 2002). Another important ERP component associated with the violation of expectation is the *error-related negativity* (Ne; Falkenstein et al., 1991) or ERN. This component is elicited when an own response is detected to be an error, thus violating the expected consequences of own action (cf. Hoffmann and Falkenstein, this issue). In the present framework, all these components share with the MMN the “problem” that their elicitation does not prove the existence of an underlying prediction. Nevertheless, a number of findings strongly suggest a predictive account, and most clearly so for MMN-related processing (as this has received most attention in the “prediction research community”).

One recent finding that is suggestive of predictive modeling in audition was presented by Grimm et al. (2011). They applied a classical oddball paradigm with repeatedly presented standard tones and occasional frequency deviants. In contrast to the multitude of previous studies employing the auditory oddball protocol, Grimm and colleagues measured not only late (*long-latency*) ERPs but also *middle-latency* potentials. These reflect a characteristic sequence of ERP deflections elicited by auditory stimuli (P0, Na, Pa, Nb) with a latency of 10 to 50 ms (e.g., Yvert et al., 2001). Grimm et al. (2011) found that regularity-violating tones did not only elicit MMN (as reported in many studies before), but also an enhanced Nb component. Thus the processing of deviant tones was already different from that of standard tones in a latency range of 37–47 ms. This early mismatch effect is hardly explainable as a post-hoc phenomenon. The assumption of predictive modeling underlying auditory processing in such situations thus gets more and more plausible. One may object that the effect in the middle-latency range was driven by increased refractoriness of the frequency-specific neurons responding to the standard frequency, and thus can be explained without the need of predictive modeling (for similar arguments on the MMN component, see Jääskeläinen et al., 2004; May and Tiitinen, 2010; Walker et al., 2001). Yet Grimm et al. (2011) also applied a control condition (see Näätänen and Alho, 1997; Schröger, 2007) that (in some sense) excludes the explanation based on differential refractoriness of frequency-specific neurons. Since the Nb enhancement could also be shown relative to this control condition, the evidence in favor of predictive processing seems convincing. Extending the findings of Grimm et al. (2011), Slabu et al. (2010) have shown that effects of prediction violation can be observed even earlier, already on the Pa component of the middle-latency potentials around 30 ms after stimulus onset. A review focusing on such early auditory deviance detection is provided by Grimm and Escera (in press).

Another piece of evidence in favor of the predictive character of the change detection system indexed by the MMN was provided by Grimm and Schröger (2007). In this study, an unmodulated sinusoidal tone was used as a standard stimulus. Short frequency modulations within this tone were used as deviant stimuli. In a 90–10 condition, these frequency modulations occurred after 200 ms with a probability of 10%. In a 70–30 condition, the modulations occurred again after 200 ms but now with a probability of 30%. In a 70–10–10–10 condition, the modulations occurred after 100, 200 or 300 ms with a probability of 10% each. Thus the probability of a deviation at any point within the tone was 30% (as in the 70–30 condition), while the probability of a modulation after 200 ms was 10% (as in the 90–10 condition). The results of this study revealed that the MMN in the 70–10–10–10 condition was as large as the MMN in the 90–10 condition, both being of higher amplitude than the MMN in the 70–30 condition. This result is best explained by assuming that the incoming stimulus is compared with the predicted stimulus point by point along the time-axis (*zip metaphor*). Therefore, the system estimates the probability of a frequency modulation not by the global probability of any modulation occurring during the stimulus, but by the local probability of a modulation occurring exactly at this point in time. The predictions thus derived are of a spectrotemporal nature, that is, representations of the spectral properties of the stimulus along the time axis. In other words, the auditory system seems to be able to make predictions stretched out over time.

2.3. Omission paradigms: Processing of missing stimuli as a special case of prediction violation

As argued in the previous section, the detection of regularity violations is highly compatible with the notion of predictive modeling. We will now consider an extreme case of regularity violation in which the expected stimulus is not occurring at all (rather than occurring with a change in one of its expected properties). In this situation, the unexpected event is again detected by the auditory system, which is reflected in the *omission MMN* — an ERP component specifically signaling the absence of an expected stimulus (Raij et al., 1997; Yabe et al., 1997; Yabe et al., 1998). This is interesting in and of itself as it demonstrates that the auditory system does not need an external trigger for detecting a deviation from its expectations. What can be even more conclusive, though, is to investigate the brain activity during the missing stimulus for signs of “the expectation”. The advantage of this approach lies in the absence of new auditory input. Overlapping activity caused by the processing of concurrent signals can thus be excluded, and the predictive activity of the system might be more unequivocally identified. Certainly, care must be taken not to measure late indicators of processing the previous tone and mistake this for predictive activity. This issue can be solved by comparing conditions in which the preceding tones elicit similar processing steps yet induce different amounts of predictability of the forthcoming tone (that is to be omitted). Applying this logic, Janata (2001) showed that a missing tone that had been highly expected elicits brain responses that are very similar to the initial sensory processing of a tone that is actually presented. Janata measured this by means of the auditory N1 component. However, he found such “N1 imitation” particularly when participants turned their expectations into active imagery — that is, when they were internally singing the sequence of tones to themselves.

A similar finding was reported by Kraemer et al. (2005). These authors showed that the auditory cortex exhibits higher activity during a missing segment of 2–5 s within a known piece of music than during a missing segment of equal duration within an unknown piece. Interestingly, the activity pattern was reversed during the piece itself: Unknown music elicited higher activity in auditory cortex than known music. Consequently, the difference in activity during the omitted segment is not due to prolonged processing of what was previously heard, but can be explained by prediction or mental continuation of the melody.

In the studies of Janata (2001) as well as Kraemer et al. (2005), it is difficult to rule out that the auditory signals during the omitted segments were caused by active imagery (mental continuation). More recent results of Bendixen et al. (2009) suggest that a similar form of predictive activity can be observed when predictable tones are processed outside the current focus of attention. The authors presented short sequences of tones in which every other tone was exactly predictable because it was a repetition of the immediately preceding tone. The other half of the tones were unpredictable with respect to their frequency, which was randomly chosen from a wide range (see Fig. 4). The authors then occasionally omitted one of the unpredictable tones or one of the predictable (repeated) tones from the sequence. Results reveal that the ERP elicited by the omission of a predictable tone is similar to the ERP elicited by the actual tone up until the P50 component (Fig. 4). This similarity in processing is not observed when comparing the omission of an unpredictable tone with the actual tone presentation. This finding suggests a stronger predictive activity at the expected onset of a predictable tone than at the expected onset of a tone that is somewhat unpredictable because one of its features (in this case, frequency) cannot yet be specified. The auditory system seems to generate predictions automatically as soon as it has enough information about the stimulus — in this case, as soon as it can specify the frequency of the tone in advance.

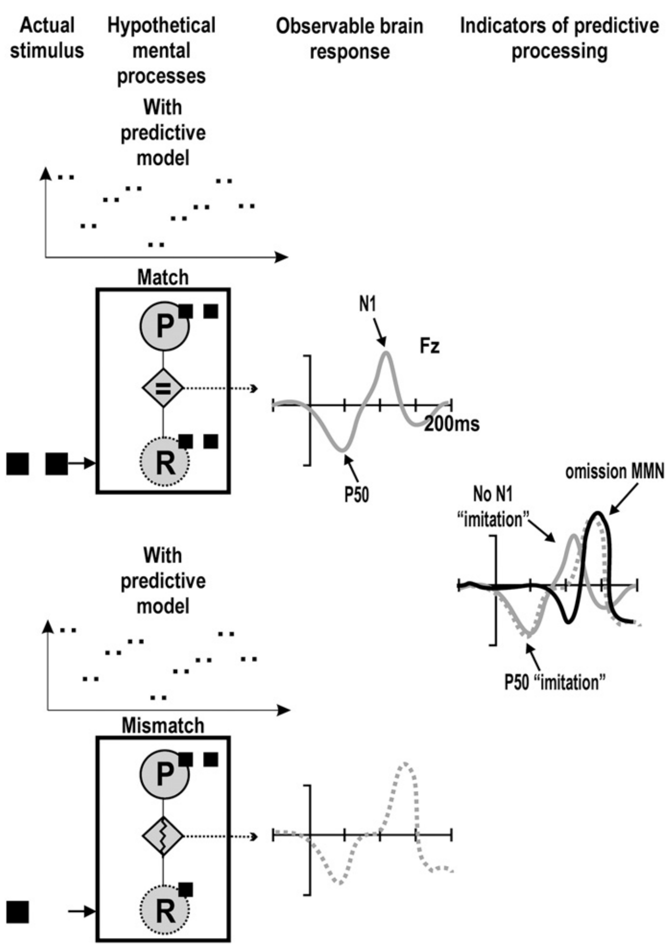

Fig. 4. Processing of a missing sound. Indicators and processes as in Figs. 2 and 3. The omission of an expected stimulus, just like other kinds of expectation violations, elicits the MMN component (here termed *omission MMN*). Another indicator of predictive modeling can be observed long before the MMN. Although no tone is presented, ERPs are elicited at the time of expected stimulus onset which resemble those elicited by the actual tone presentation. In experiments in which the auditory stimuli were processed outside the current focus of attention, the "imitation" of tone processing during omissions has been shown for the time range of the P50 ERP component (Bendixen et al., 2009). In experiments in which the tones were attentively processed, even the generation of an N1 component without physical input has been shown (Janata, 2001). Adapted from Schröger E., SanMiguel, I., & Bendixen A., 2012, Prädiktive Modellierung in der auditiven Wahrnehmung. In. E. Schröger & S. Koelsch (Eds.). Kognitive und Affektive Neurowissenschaften. Enzyklopädie der Psychologie (Serie II: Kognition, Band 9). Göttingen: Hogrefe; with permission of Hogrefe Publisher.

2.4. Self-generation paradigms: Processing of self-generated stimuli–Prediction in the context of action

It has long been known that many organisms take the consequences of their own action into account for adjustments of sensory processing. A well-known example is the phenomenon that we are hardly able to tickle ourselves, while the same kind of tactile stimulation will make us laugh when applied by another person. Another example is the fact that the visual world stays stable when we move our eyes. This is astonishing because the excitation dynamics of the retina during an eye movement should lead to the impression that the environment is moving. The impression of a moving environment is indeed caused when our eyes are moved passively (e.g., by slightly pushing on the eyeball). Such phenomena follow the so-called *re-afference principle*. They have been demonstrated in many species, from insects up to mammals. As a mechanism for explaining such phenomena, it is assumed that a copy of the efferent (motor) signal is sent to the sensory system (*efferency copy*: von Holst and Mittelstaedt, 1950) so that the sensory system is able to predict the signal changes caused by the motor act (*corollary discharge*: Sperry, 1950) and subtract them from the actual sensory signals.

On a very general level, sensory effects caused by own motor acts can be predicted, and the processing of self-generated sensory changes can be adapted appropriately. Such phenomena are well-known in the auditory modality as well. Several studies have reported that the evoked potentials or evoked magnetic fields (in particular, the N1 or N1m component ca. 100 ms after tone onset) are reduced in amplitude for a self-generated relative to an externally generated tone (Baess et al., 2008; Hazemann et al., 1975; Martikainen et al., 2005; McCarthy and Donchin, 1976; Schäfer and Marcus, 1973). This phenomenon is called N1 suppression. A typical result is displayed on Fig. 5. Recently it was shown that this suppression occurs not only on the N1 component but already on the Pa component between 27 and 33 ms after stimulus onset (Baess et al., 2009).

In the typical N1 suppression studies, self-generated and externally generated stimuli are presented in different experimental blocks. Therefore one might argue that the suppression effects do not result from predictions of the sensory consequences of own action, but merely reflect context effects. To rule out this alternative explanation, Baess et al. (2011) presented self-generated and externally generated tones in the same experimental block. Again, N1 suppression was obtained; indeed, it was even larger than in a comparison condition with the classical protocol. Thus what underlies the suppression seems to be the "knowledge" about the consequences of one's own motor acts.

One could interpret the suppression by considering that self-generated tones reveal no new information and thus do not need to be processed to a large extent. In contrast, externally generated tones are unknown and thus require more detailed processing. An alternative explanation posits that the knowledge about the effect that one will soon cause results in an anticipation and, in turn, in a kind of emulation of stimulus processing. When the stimulus is then indeed perceived, the neural circuits responsible for its processing are refractory, which weakens stimulus processing. The exact nature of the effect remains unknown up to now, but both interpretations can be regarded as predictive phenomena.

Suppression effects have also been shown in numerous vocalization studies. For instance, spoken language elicits smaller evoked

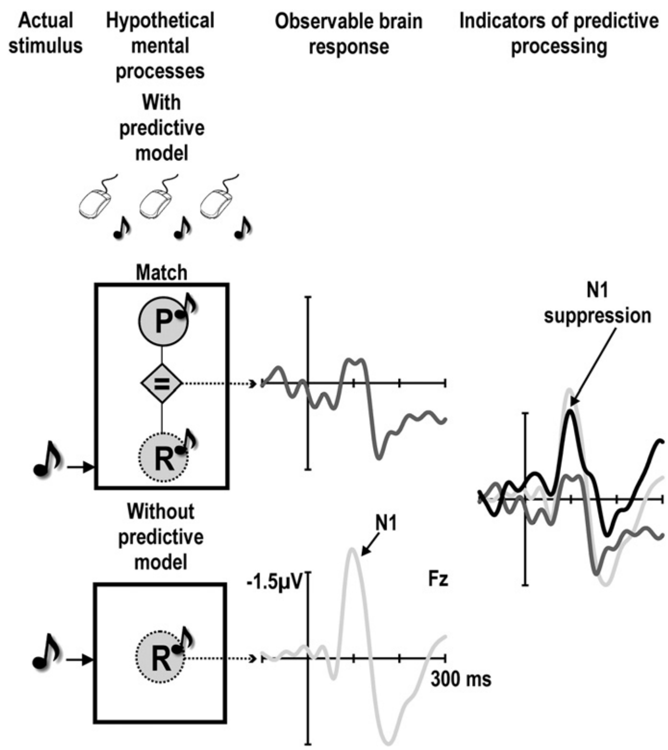

Fig. 5. Processing of a self-generated auditory stimulus, compared with processing of an externally generated stimulus. Indicators and processes as in Figs. 2 to 4. A self-generated stimulus elicits attenuated sensory processing, which can most easily be shown by reduction of the N1 component around 100 ms after stimulus onset. Similar suppression effects have meanwhile been reported for earlier components (30 ms after stimulus onset) (Baess et al., 2009, not depicted). Adapted from Schröger E., SanMiguel, I., & Bendixen A., 2012, Prädiktive Modellierung in der auditiven Wahrnehmung. In. E. Schröger & S. Koelsch (Eds.). Kognitive und Affektive Neurowissenschaften. Enzyklopädie der Psychologie (Serie II: Kognition, Band 9). Göttingen: Hogrefe; with permission of Hogrefe Publisher.

potentials in the speaker than in the listener (Curio et al., 2000; Gunji et al., 2001; Numminen and Curio, 1999; Numminen et al., 1999; Ventura et al., 2009). Animal studies have revealed that most neurons in auditory cortex show reduced responses during vocalization (Eliades and Wang, 2003; Müller-Preuss and Ploog, 1981). Similar findings were obtained in human intracranial recordings from superior and middle temporal gyrus in epileptic patients (Creutzfeldt et al., 1989). The suppression can be quite unspecific, and it can even precede the vocalization (Creutzfeldt et al., 1989; Eliades and Wang, 2003). Suppression effects are strongest for unchanged own vocalizations, and responses to self-generated speech sounds are in fact enhanced when the auditory feedback of own vocalizations is altered (Behroozmand et al., 2009, 2011; Fu et al., 2006), following the pattern of other mismatch responses described above. Therefore, the effect cannot be a simple, unspecific one of blocking any kind of information processing during vocalization (Fu et al., 2006; Hashimoto and Sakai, 2003; Heinks-Maldonado et al., 2005; Hirano et al., 1997; McGuire et al., 1996). It is thus reasonable to assume that specific predictions concerning self-generated sounds play a role in suppressing the neural responses elicited by an auditory stimulus (cf. Ford and Mathalon, 2012-this issue).

2.5. Caveats for testing the existence of predictions

From the taxonomy of paradigms and brain indices described this far it stands out that we dispose of numerous valuable tools to probe predictive processes in audition. However, it is also clear that each of these have possible pitfalls. A few of the abovementioned omission studies exemplify the need to draw a line between the automatic generation of predictions as a fundamental part of basic auditory perceptive processing, and actively controlled auditory imagery. In other cases, the difficulty rather resides in the distinction between effects based on predictive processes and effects caused by purely attentional differences. This is particularly critical in self-generation paradigms. Both potential confounds may be partially relieved by studies proving that predictive processes take place also outside of the focus of attention.

A returning issue is whether the effects may be explained on the basis of differential refractoriness of the underlying neural populations alone. Whether or not predictive modeling may come about as an emergent property of complex adaptation dynamics of the underlying neural populations' firing patterns; the effects that are presented as proof of predictive modeling must prove to be effects of prediction rather than of repetition by itself. We refer here again to our rule of thumb, contrasting unexpected repetitions with expected changes. Proper refractoriness controls should be used to rule out the possibility of mere repetition effects. Also, the role that both residual activity and general widespread modulations may play in the effects must be taken into consideration. In this respect, proof of specific predictions outweighs that of unspecific effects. Early cross-modal effects are also more convincingly not due to residual activity or contextual effects. Finally, prediction effects that require the extrapolation of abstract rules are less likely to be based on refractoriness differences alone.

Prediction should be prospective rather than retrospective. As discussed, this is a distinction that cannot be made with certainty in many cases. However, very early effects are more likely to be due to prospective mechanisms than to a post-hoc retrospective comparison to a stored template. Also, omission studies that would allow looking into the predictive activity per se rather than at the outcome of a prediction (or template) vs. actual stimulation comparison would provide arguments in favor of a prospective account. Finally, whenever match or mismatch effects are presented, one should be careful with interpreting the direction of the effects if an appropriate neutral baseline is not included.

2.6. Considerations on the generation and neural implementation of predictive models

How can the predictions presumably underlying the *match*, *mismatch*, *omission* and *self-generation* effects come about? For self-generated tones, one may assume that the auditory consequences of own motor acts are “hard-wired” during development. For forming a predictive model later on, they need to be complemented by information as to which specific auditory stimulus is to be expected. For auditory predictions derived from visual stimuli, one could argue in a similar way via cross-modal associative mechanisms. Yet the case of predictions derived from the auditory sequence itself needs specification of a mechanism of regularity extraction. We suggest that the relations of successive stimuli are extracted and represented (phase 1). These relation representations are then compared (phase 2). On the basis of this comparison, a regularity can be extracted. If confidence in the regularity is large enough, a prediction for the next stimulus can be made (phase 3). The plausibility of a three-stage-model along these lines is supported by computational approaches (Kiebel et al., 2009).

Regularities seem to be extracted from auditory sequences amazingly fast. A single repetition of the relation representation is sufficient for the formation of a model and the derivation of predictions, whether the model is simple (e.g., “the next tone will have the same frequency as the current tone”) (Bendixen et al., 2007; Horváth et al., 2001) or a bit more abstract (e.g., “the next tone will be one semitone higher than the current tone”) (Bendixen and Schröger, 2008). With more complex predictive models, regularity extraction and the generation of predictions needs more time, or more repetitions of the relation representation (Bendixen et al., 2008). Once the predictive model is formed, it needs to be maintained and continuously checked for the validity of its predictions. If an incorrect prediction is detected, the model must be modified accordingly. Possibly the MMN component reflects this model update rather than violation detection per se (Winkler, 2007; Winkler and Czigler, 1998).

How could a prediction be neurally implemented? On the basis of research on oscillatory activity (Engel et al., 2001; Schroeder et al., 2008), it is reasonable to assume that oscillatory brain responses taking place in auditory areas are modulated by the respective auditory or visual input in a predictive manner. If the incoming stimulus hits upon pre-activated stimulus-specific brain responses, a form of resonance is caused which can be observed as an amplitude enhancement of oscillatory activity. This could explain the gamma-band enhancement findings of Schadow et al. (2009) and Widmann et al. (2007) described above. Moreover, in the auditory domain, the stimulation itself is often periodic. The mere assumption of an oscillatory process is sufficient for explaining that a response is generated in auditory cortex when a tone is missing (May and Tiitinen, 2001; May and Tiitinen, 2010). This could at first glance explain the activity in the P50 latency range elicited by sounds omissions in Bendixen et al. (2009). Yet the fact that the response was modulated by the predictability of the upcoming (to-be-omitted) sound is more difficult to explain in terms of a simple oscillatory process, and thus argues in favor of a predictive account.

Meanwhile, neuro-computational approaches exist that model the predictive behavior of the auditory system by interactions between neuronal populations on different hierarchical levels (Friston and Kiebel, 2009). According to Friston (2005), both *top-down* processes (conveying the predictions generated by higher cortical areas to sensory areas) and *bottom-up* processes (passing mismatch signals on to higher areas) are at play here. In our view, it is even possible that the interaction between prediction and prediction error takes place on a purely auditory level of information processing, without the involvement of higher (e.g., frontal) cortical areas. Of course, higher areas are involved in instantiating the prediction when the regularities become more complex, when intentional influences play a role, when long-term memory entries are relevant, or when other modalities are incorporated. There seems to be a close link between predictive processing and active sensing according to which motor behavior (and attention "behavior") has a strong impact on sensory processing (cf. Schroeder et al., 2010). Such behavior may entrain cortical rhythms of sensory areas and thus guide the processing of sensory information via amplification or inhibition. In other words, auditory prediction could be instantiated via the regulation of the (oscillatory) responsivity of auditory neural populations. The regulation may stem from auditory as well as from non-auditory areas. In case of the classical MMN or omission responses, it is likely to originate from auditory areas, whereas in case of the audio–visual gamma-band enhancement (e.g. Widmann et al., 2007) or N1 suppression for self-generated sounds (e.g. Baess et al., 2008), it is likely to originate from non-auditory processing modules such as visual or motor areas. Whether auditory prediction is based on auditory or non-auditory (visual, motor) information, other brain areas seem to assist in setting up the prediction. For example, one key structure seems to be the cerebellum, which receives input from motor and sensory areas and is likely to contribute to predictions. Indeed, according to neuro-imaging studies, the MMN seems to have generators in the cerebellum (Dittmann-Balçar et al., 2001). Moreover, Moberget et al. (2008) found that cerebellar patients reveal dysfunctional MMN for duration deviants. In addition, Knolle et al. (in revision) found reduced N1 suppression for self-generated sounds in cerebellar patients. A general framework for possible interactions between auditory areas and the cerebellum in generating and maintaining auditory predictions is given by Schwartze et al. (2012-this issue).

3. Implications of predictive modeling

In the previous section, we have proposed a framework for organizing various research paradigms along the lines of predictive processing. In our view, the experimental evidence convincingly suggests a predictive organization of the auditory system. It is difficult to imagine that this "predictive power" should have evolved as an emergent property of the system without an own functional role. In the following section, we will discuss the possible benefits of predictive modeling within and beyond the auditory system.

3.1. Implications for auditory information processing

A predictive system poses advantages for attentional mechanisms which can be sorted into two main groups. On the one hand, expected stimuli can be processed in more detail if they are currently task-relevant. For instance, for a tone whose onset time and pitch are predictable, one can judge faster whether a short gap was present or absent than for the same tone presented at an unexpected point in time or with an unexpected pitch (Lange, 2009; see also Rimmele et al., 2011, for similar evidence with temporal and spatial expectations). Thus in this case, expectations assist in allocating attentional resources at just the right moment in time, and thereby facilitate performance. On the other hand, stimuli that are currently irrelevant for the organism can be blocked from attentive processing — their information content does not need to be checked as their occurrence was predictable. Thus processing resources can be saved and invested into processes that would otherwise compete for resources (Sinkkonen, 1999). The fact that expected stimuli are sometimes processed with more and sometimes with fewer attentional resources – and that both of these cases are presented as evidence in favor of prediction–attention interactions – is occasionally puzzling, but readily explainable when taking the current goals of the organism into account.

Fading out certain predictable stimuli might represent more than just an attentional feature. In the case of self-generated stimuli, there is evidence that these are in fact perceived less intensely (Cardoso-Leite et al., 2010). This mechanism can be vital, as permanent sensory processing of self-generated stimulation could desensitize the receptors to an extent that relevant externally-generated stimulation would no longer be detected. For example, generating speech causes such a strong auditory feedback that if this was not filtrated, it would possibly render the individual temporarily deaf to other incoming sounds.

Whether predicted events are currently intensely processed or, on the contrary, blocked from attentive processing — the predictive mechanism remains active in the background. The permanent generation of predictions results in a very useful side effect: Each incoming stimulus can be checked for its conformance to the prediction. If it is found to be non-conforming, attention is immediately drawn towards the prediction violation in order to check its relevance (Escera et al., 1998; Schröger and Wolff, 1998). This allows identification of new stimuli in the environment and of stimuli that change their behavior in an unexpected way (e.g., that are suddenly approaching us). Likewise, the unexpected disappearance of a stimulus is immediately detected. The organism can then prepare an adequate response to the new situation. The predictive character of perception thus provides the organism with the opportunity of an optimal allocation of attentional resources towards its current goals and yet leaves it with the possibility to detect new information. Moreover, in the case of self-generated stimuli, detecting deviations from the predicted outcome is synonymous to detecting errors in one's own performance, and signals the need to adjust behavior in order to achieve the intended outcome. In this vein, checking for conformance with the prediction is proposed to lie at the basis for self-monitoring of speech (Fu et al., 2006).

Furthermore, a predictive system is particularly beneficial when objects in the environment are moving. Imagine constructing the auditory behavior of a fly buzzing around the head in a retrospective manner — movement perception would consist of bits and pieces. In contrast, a predictive system automatically leads to a coherent perceptual impression, and it helps in anticipating the movement's continuation (which, in turn, allows for the adaptation of own motor acts). For the visual system, anticipatory motor planning based on sensory predictions is well established (Diaz et al., 2009).

Regarding the information processing steps mentioned so far, predictions are equally helpful in all sensory modalities. In particular, one can always find visual analogs of the examples given here for the auditory system. Many predictive processes are actually investigated in more detail in the visual modality. Yet when directly comparing predictive processes and their impact on attention, effects are typically larger in the auditory than in the visual system (Bendixen et al., 2010b; Berti and Schröger, 2001; Boll and Berti, 2009). It could be that predictive mechanisms are implemented in a particularly efficient manner in the auditory system because of the nature of auditory stimuli. Possibly, handling the extremely fast occurrence of auditory signals is possible only by means of predictive mechanisms, since processing steps can be saved when predictable information does not have to be analyzed in detail. Yet predictions are not just helpful because of the highly variable input in audition — they also serve a very specific function in auditory scene analysis.

In addition to the high temporal variability, the auditory system is confronted with the problem that signals of concurrently active sources overlap (in contrast to the visual system, where competing stimuli typically occlude each other — unless they are transparent). The complex mixture of signals needs to be disentangled before relevant information can be retrieved; a process termed auditory scene analysis (Bregman, 1990). The decomposition could be massively facilitated if a part of the mixture could be subtracted immediately because it was predictable. This consideration (Jones, 1976; Jones et al., 1981) has received renewed interest lately (Denham and Winkler, 2006; Winkler et al., 2009). Experimental evidence for a role of predictability in the decomposition of auditory signals has been provided by Bendixen et al. (2010a) who showed that the representation of a signal as coming from a separate source can be stabilized if this signal shows predictable rather than unpredictable fluctuations in frequency and intensity. Consistent results have been obtained with temporal predictability by Andreou et al. (in press). Thus predictive modeling supports the auditory system in one of its central tasks: solving the auditory scene analysis problem. This may explain the exceptional position of predictions in audition.

3.2. Implications for other cognitive processes

The previous section highlights the implications that a predictive auditory system has for processing auditory signals in themselves. However, the impact of this modus operandi of the auditory system reaches far beyond sensory information processing. Particularly those predictions that are based on the self-generation of sensory effects are increasingly viewed as essential for the integrity of various cognitive functions. As we have described above, self-generated stimuli are processed differently from passively perceived stimuli, most likely due to their highly predictable nature. Thus, by means of sensory predictions, it is for example possible to decide whether an otherwise identical sensory event was self-generated or caused by another person. The implications of being able to make such distinction range from inhibiting an avoidance reflex towards a sudden loud self-generated sound to the understanding of agency, intentionality and causality of action. Prediction-based distinction between self and other might even contribute to developing a sense of self (Jeannerod, 2003). Not only is there evidence that self-recognition depends on predictive modeling, but there is also evidence for the converse conclusion: that dysfunction of predictive modeling might contribute to loss of self. A line of research proposes that a dysfunction in predictive mechanisms might underlie many symptoms of schizophrenia, which are often related to difficulties in distinguishing self from other (Feinberg, 1978; Frith et al., 2000). There is indeed evidence for an impairment of sensory attenuation to self-generated sounds in schizophrenia (e.g. Ford et al., 2007; Ford et al., 2001a; Ford et al., 2001b; see also Ford and Mathalon, 2012-this issue). Ford and Mathalon (2005) suggest that symptoms such as auditory hallucinations and “thought insertion” in schizophrenic patients could be due to a dysfunction in predictive mechanisms that would normally signal inner speech and thoughts as self-generated. These examples highlight the far-reaching repercussions of a fundamentally predictive auditory system into other cognitive domains.

4. Summary and conclusions

The above findings and considerations illustrate that predictions play a central role not only in higher cognitive functions such as problem solving, but also at the very core of perception. This is particularly true in auditory perception, but can be extended to other sensory modalities. Without predictive modeling, we would hardly be able to analyze the fast sensory signals adequately. We would have problems in auditory scene analysis (i.e., in disentangling the overlapping signals of simultaneously active sound sources) and in following the fast acoustic changes emitted by a single source. For instance, a human speaker can produce up to 15 phonemes per second (Levelt, 1999); still, we can follow the stream of speech without much effort. Actually, in view of the overwhelming evidence towards predictive processing, it seems that “following” a speaker may no longer be the right term. Fast information processing is enabled by the permanent generation of predictions on various levels of processing. Thus it appears more correct to say that the stream of speech is understood through an interaction of perceiving and constructing the underlying auditory signal.

Each predictive model is developed on the basis of specific experience in the environment and is continuously updated to account for new experiences. It thus rests both on the current context and on long-term memory. Up until now, research on auditory predictive modeling has focused on the “neutral” or “cognitive” aspects of the involved information processing stages. It is, however, highly plausible that motivational and emotional aspects play an important role in this process (cf. Barrett and Bar, 2009). Signals in the auditory environment often have an emotional valence (e.g., they might be comforting or threatening). The motivational and emotional aspects should not be perceived as counterparts to the cognitive aspects, but as distinguishable facets of an integrated system. After all, motivational and emotional content are “but” pieces of information used by the system for optimizing adaptive behavior. Nevertheless, a research deficit must be asserted in the relation between motivation/emotion and auditory predictive modeling.

Finally, we would like to point out once more that predictive modeling is by no means limited to within auditory perception, but has implications much beyond that, such as attentional allocation or the distinction between self and other.

Three core statements shall summarize and conclude this contribution. (1) We hear our acoustic environment through predictive models that we have acquired about it. (2) This predictive modeling in auditory perception has implications for a number of other cognitive functions. (3) Recent research has developed a variety of paradigms and indicators – summarized in the present taxonomy – that are suited to tap into auditory predictive processes of the human brain.

Acknowledgements

This work was supported by a DFG Reinhart-Koselleck grant on Auditory Prediction to ES (SCHR 375/20-1).

References

Andreou, L.-V., Kashino, M., Chait, M., in press. The role of temporal regularity in auditory segregation. Hear. Res.
Baess, 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.
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., 2006. Repetition effects to sounds: evidence for predictive coding in the auditory system. Trends Cogn. Sci. 10, 93–94.
Baldeweg, T., Klugman, A., Gruzelier, J., Hirsch, S.R., 2004. Mismatch negativity potentials and cognitive impairment in schizophrenia. Schizophr. Res. 69, 203–217.
Bar, M., 2007. The proactive brain: using analogies and associations to generate predictions. Trends Cogn. Sci. 11, 280–289.
Barrett, L.F., Bar, M., 2009. See it with feeling: affective predictions during object perception. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364, 1325–1334.
Behroozmand, R., Karvelis, L., Liu, H., Larson, C.R., 2009. Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. Clin. Neurophysiol. 120, 1303–1312.
Behroozmand, R., Liu, H., Larson, C.R., 2011. Time-dependent neural processing of auditory feedback during voice pitch error detection. J. Cogn. Neurosci. 23, 1205–1217.
Bendixen, A., Schröger, E., 2008. Memory trace formation for abstract auditory features and its consequences in different attentional contexts. Biol. Psychol. 78, 231–241.
Bendixen, A., Roeber, U., Schröger, E., 2007. Regularity extraction and application in dynamic auditory stimulus sequences. J. Cogn. Neurosci. 19, 1664–1677.
Bendixen, A., Prinz, W., Horváth, J., Trujillo-Barreto, N.J., Schröger, E., 2008. Rapid extraction of auditory feature contingencies. Neuroimage 41, 1111–1119.
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., Denham, S.L., Gyimesi, K., Winkler, I., 2010a. Regular patterns stabilize auditory streams. J. Acoust. Soc. Am. 128, 3658–3666.
Bendixen, A., Grimm, S., Deouell, L.Y., Wetzel, N., Mädebach, A., Schröger, E., 2010b. The time–course of auditory and visual distraction effects in a new crossmodal paradigm. Neuropsychologia 48, 2130–2139.
Berti, S., Schröger, E., 2001. A comparison of auditory and visual distraction effects: behavioral and event-related indices. Cogn. Brain Res. 10, 265–273.
Boll, S., Berti, S., 2009. Distraction of task-relevant information processing by irrelevant changes in auditory, visual, and bimodal stimulus features: a behavioral and event-related potential study. Psychophysiology 46, 645–654.
Bregman, A.S., 1990. Auditory scene analysis. The Perceptual Organization of Sound. MIT Press, Cambridge MA. 772 pp.
Cardoso-Leite, P., Mamassian, P., Schütz-Bosbach, S., Waszak, F., 2010. A new look at sensory attenuation. Action–effect anticipation affects sensitivity, not response bias. Psychol. Sci. 21, 1740–1745.
Cheung, O.S., Bar, M., 2012. Visual prediction and perceptual expertise. Int. J. Psychophysiol. 83, 156–163 (this issue).
Costa-Faidella, J., Grimm, S., Slabu, L., Díaz-Santaella, F., Escera, C., 2011. Multiple time scales of adaptation in the auditory system as revealed by human evoked potentials. Psychophysiology 48, 774–783.
Craik, K.J.W., 1943. The Nature of Explanation. University Press, Cambridge Eng. 134 pp.
Creutzfeldt, O., Ojemann, G., Lettich, E., 1989. Neuronal activity in the human lateral temporal lobe: 2. Responses to the subjects own voice. Exp. Brain Res. 77, 476–489.
Curio, G., Neuloh, G., Numminen, J., Jousmaki, V., Hari, R., 2000. Speaking modifies voice-evoked activity in the human auditory cortex. Hum. Brain Mapp. 9, 183–191.
Denham, S.L., Winkler, I., 2006. The role of predictive models in the formation of auditory streams. J. Physiol. Paris 100, 154–170.
Diaz, G.J., Phillips, F., Fajen, B.R., 2009. Intercepting moving targets: a little foresight helps a lot. Exp. Brain Res. 195, 345–360.
Dittmann-Balçar, A., Jüptner, M., Jentzen, W., Schall, U., 2001. Dorsolateral prefrontal cortex activation during automatic auditory duration-mismatch processing in humans: a positron emission tomography study. Neurosci. Lett. 308, 119–122.
Eliades, S.J., Wang, X.Q., 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., Winkler, I., Näätänen, R., 1998. Neural mechanisms of involuntary attention to acoustic novelty and change. J. Cogn. Neurosci. 10, 590–604.
Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L., 1991. Effects of crossmodal divided attention on late ERP components: II. Error processing in choice reaction tasks. Electroencephalogr. Clin. Neurophysiol. 78, 447–455.
Feinberg, I., 1978. Efference copy and corollary discharge — implications for thinking and its disorders. Schizophr. Bull. 4, 636–640.
Ford, J.M., Mathalon, D.H., 2005. Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? Int. J. Psychophysiol. 58, 179–189.
Ford, J.M., Mathalon, D.H., 2012. Anticipating the future: automatic prediction failures in schizophrenia. Int. J. Psychophysiol. 83, 120–131 (this issue).
Ford, J.M., Mathalon, D.H., Heinks, T., Kalba, S., Faustman, W.O., Roth, W.T., 2001a. Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. Am. J. Psychiatry 158, 2069–2071.
Ford, J.M., Mathalon, D.H., Kalba, S., Whitfield, S., Faustman, W.O., Roth, W.T., 2001b. Cortical responsiveness during talking and listening in schizophrenia: an event-related brain potential study. Biol. Psychiatry 50, 540–549.
Ford, J.M., Gray, M., Faustman, W.O., Roach, B.J., Mathalon, D.H., 2007. Dissecting corollary discharge dysfunction in schizophrenia. Psychophysiology 44, 522–529.
Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing. Trends Cogn. Sci. 6, 78–84.
Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 815–836.
Friston, K., Kiebel, S., 2009. Cortical circuits for perceptual inference. Neural Netw. 22, 1093–1104.
Frith, C.D., Blakemore, S.J., Wolpert, D.M., 2000. Abnormalities in the awareness and control of action. Philos. Trans. R. Soc. Lond. B Biol. Sci. 355, 1771–1788.
Fu, C.H.Y., Vythelingum, G.N., Brammer, M.J., Williams, S.C.R., Amaro, E., Andrew, C.M., et al., 2006. An fMRI study of verbal self-monitoring: neural correlates of auditory verbal feedback. Cereb. Cortex 16, 969–977.
Giard, M.-H., Perrin, F., Pernier, J., Bouchet, P., 1990. Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. Psychophysiology 27, 627–640.
Gregory, R.L., 1980. Perceptions as hypotheses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 290, 181–197.
Grimm, S., Escera, C., in press. Auditory deviance detection revisited: evidence for a hierarchical novelty system. *Int. J. Psychophysiol.*
Grimm, S., Schröger, E., 2007. The processing of frequency deviations within sounds: evidence for the predictive nature of the mismatch negativity (MMN) system. Restor. Neurol. Neurosci. 25, 241–249.
Grimm, S., Roeber, U., Trujillo-Barreto, N.J., Schröger, E., 2006. Mechanisms for detecting auditory temporal and spectral deviations operate over similar time windows but are divided differently between the two hemispheres. Neuroimage 32, 275–282.
Grimm, S., Escera, C., Slabu, L., Costa-Faidella, J., 2011. Electrophysiological evidence for the hierarchical organization of auditory change detection in the human brain. Psychophysiology 48, 377–384.
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.
Hagoort, P., Brown, C., Groothusen, J., 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. Lang. Cogn. Process. 8, 439–483.
Hahne, A., Friederici, A.D., 1999. Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. J. Cogn. Neurosci. 11, 194–205.
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. Electroencephalogr. Clin. Neurophysiol. 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.
Helmholtz, H.v., 1867. Handbuch der physiologischen Optik, Voss, Leipzig. 874 pp.
Herrmann, C.S., Munk, M.H.J., Engel, A.K., 2004. Cognitive functions of gamma-band activity: memory match and utilization. Trends Cogn. Sci. 8, 347–355.
Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., et al., 1997. Cortical processing mechanism for vocalization with auditory verbal feedback. Neuroreport 8, 2379–2382.
Hoffmann, S., Falkenstein, M., 2012. Predictive information processing in the brain: errors and response monitoring. Int. J. Psychophysiol. 83, 208–212 (this issue).
Horváth, J., Czigler, I., Sussman, E., Winkler, I., 2001. Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. Cogn. Brain Res. 12, 131–144.
Horváth, J., Czigler, I., Jacobsen, T., Maeß, B., Schröger, E., Winkler, I., 2008. MMN or no MMN: no magnitude of deviance effect on the MMN amplitude. Psychophysiology 45, 60–69.
Jääskeläinen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levänen, S., et al., 2004. Human posterior auditory cortex gates novel sounds to consciousness. Proc. Natl. Acad. Sci. U. S. A. 101, 6809–6814.
Janata, P., 2001. Brain electrical activity evoked by mental formation of auditory expectations and images. Brain Topogr. 13, 169–193.
Jeannerod, M., 2003. The mechanism of self-recognition in humans. Behav. Brain Res. 142, 1–15.
Johnson-Laird, P.N., 1983. Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness. Harvard University Press, Cambridge MA. 513 pp.
Jones, M.R., 1976. Time, our lost dimension: toward a new theory of perception, attention, and memory. Psychol. Rev. 83, 323–355.
Jones, M.R., Kidd, G., Wetzel, R., 1981. Evidence for rhythmic attention. J. Exp. Psychol. Hum. Percept. Perform. 7, 1059–1073.
Kaiser, J., Lutzenberger, W., 2005. Human gamma-band activity: a window to cognitive processing. Neuroreport 16, 207–211.

Kiebel, S.J., von Kriegstein, K., Daunizeau, J., Friston, K.J., 2009. Recognizing sequences of sequences. PLoS Comput. Biol. 5, e1000464.

Knolle, F., Schröger, E., Baess, P., Kotz, S.A., in revision. The cerebellum generates auditory predictions: ERP lesion data.

Koelsch, S., Gunter, T., Friederici, A.D., Schröger, E., 2000. Brain indices of music processing: "nonmusicians" are musical. J. Cogn. Neurosci. 12, 520–541.

Kraemer, D.J.M., Macrae, C.N., Green, A.E., Kelley, W.M., 2005. Musical imagery: sound of silence activates auditory cortex. Nature 434, 158.

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

Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205.

Lange, K., 2009. Brain correlates of early auditory processing are attenuated by expectations for time and pitch. Brain Cogn. 69, 127–137.

Levelt, W.J.M., 1999. Models of word production. Trends Cogn. Sci. 3, 223–232.

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., Tiitinen, H., 2001. Human cortical processing of auditory events over time. Neuroreport 12, 573–577.

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

McCarthy, G., Donchin, E., 1976. Effects of temporal and event uncertainty in determining waveforms of 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.

Moberget, T., Karns, C.M., Deouell, L.Y., Lindgren, M., Knight, R.T., Ivry, R.B., 2008. Detecting violations of sensory expectancies following cerebellar degeneration: a mismatch negativity study. Neuropsychologia 46, 2569–2579.

Molholm, S., Martinez, A., Ritter, W., Javitt, D.C., Foxe, J.J., 2005. The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. Cereb. Cortex 15, 545–551.

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

Näätänen, R., Alho, K., 1997. Higher-order processes in auditory-change detection. Trends Cogn. Sci. 1, 44–45.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., Winkler, I., 2001. 'Primitive intelligence' in the auditory cortex. Trends Neurosci. 24, 283–288.

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol. 118, 2544–2590.

Näätänen, R., Astikainen, P., Ruusuvirta, T., Huotilainen, M., 2010. Automatic auditory intelligence: an expression of the sensory-cognitive core of cognitive processes. Brain Res. Rev. 64, 123–136.

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.

Osterhout, L., Holcomb, P.J., 1992. Event-related brain potentials elicited by syntactic anomaly. J. Mem. Lang. 31, 785–806.

Paavilainen, P., Jaramillo, M., Näätänen, R., Winkler, I., 1999. Neuronal populations in the human brain extracting invariant relationships from acoustic variance. Neurosci. Lett. 265, 179–182.

Paavilainen, P., Arajärvi, P., Takegata, R., 2007. Preattentive detection of nonsalient contingencies between auditory features. Neuroreport 18, 159–163.

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.

Rimmele, J., Jolsvai, H., Sussman, E., 2011. Auditory target detection is affected by implicit temporal and spatial expectations. J. Cogn. Neurosci. 23, 1136–1147.

Rock, I., 1983. The Logic of Perception. MIT Press, Cambridge MA. 341 pp.

Rohrmeier, M.A., Koelsch, S., 2012. Predictive information processing in music cognition. Int. J. Psychophysiol. 83, 164–175 (this issue).

Rosburg, T., Trautner, P., Dietl, T., Korzyukov, O.A., Boutros, N.N., Schaller, C., et al., 2005. Subdural recordings of the mismatch negativity (MMN) in patients with focal epilepsy. Brain 128, 819–828.

Sabri, M., Liebenthal, E., Waldron, E.J., Medler, D.A., Binder, J.R., 2006. Attentional modulation in the detection of irrelevant deviance: a simultaneous ERP/fMRI study. J. Cogn. Neurosci. 18, 689–700.

Schadow, J., Lenz, D., Dettler, N., Fründ, I., Herrmann, C.S., 2009. Early gamma-band responses reflect anticipatory top-down modulation in the auditory cortex. Neuroimage 47, 651–658.

Schäfer, E.W.P., Marcus, M.M., 1973. Self-stimulation alters human sensory brain responses. Science 181, 175–177.

Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., Puce, A., 2008. Neuronal oscillations and visual amplification of speech. Trends Cogn. Sci. 12, 106–113.

Schroeder, C.E., Wilson, D.A., Radman, T., Scharfman, H., Lakatos, P., 2010. Dynamics of active sensing and perceptual selection. Curr. Opin. Neurobiol. 20, 172–176.

Schröger, E., 2007. Mismatch negativity: a microphone into auditory memory. J. Psychophysiol. 21, 138–146.

Schröger, E., Wolff, C., 1998. Behavioral and electrophysiological effects of task-irrelevant sound change: a new distraction paradigm. Cogn. Brain Res. 7, 71–87.

Schröger, E., Bendixen, A., Trujillo-Barreto, N.J., Roeber, U., 2007. Processing of abstract rule violations in audition. PLoS One 2, e1131.

Schwartze, M., Rothermich, K., Schmidt-Kassow, M., Kotz, S.A., 2011. Temporal regularity effects on pre-attentive and attentive processing of deviance. Biol. Psychol. 87, 146–151.

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 (this issue).

Sculthorpe, L.D., Campbell, K.B. (in press). Evidence that the mismatch negativity to pattern violations does not vary with deviant probability. Clin. Neurophysiol.

Shalgi, S., Deouell, L.Y., 2007. Direct evidence for differential roles of temporal and frontal components of auditory change detection. Neuropsychologia 45, 1878–1888.

Sinkkonen, J., 1999. Information and resource allocation. In: Baddeley, R., Hancock, P., Foldiak, P. (Eds.), Information Theory and the Brain. Cambridge University Press, Cambridge, pp. 241–254.

Slabu, L., Escera, C., Grimm, S., Costa-Faidella, J., 2010. Early change detection in humans as revealed by auditory brainstem and middle-latency evoked potentials. Eur. J. Neurosci. 32, 859–865.

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

Sussman, E.S., 2007. A new view on the MMN and attention debate: the role of context in processing auditory events. J. Psychophysiol. 21, 164–175.

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

von Holst, E., Mittelstaedt, H., 1950. Das Reafferenzprinzip (Wechselwirkungen zwischen Zentralnervensystem und Peripherie). Naturwissenschaften 37, 464–476.

Walker, L.J., Carpenter, M., Downs, C.R., Cranford, J.L., Stuart, A., Pravica, D., 2001. Possible neuronal refractory or recovery artifacts associated with recording the mismatch negativity response. J. Am. Acad. Audiol. 12, 348–356.

Widmann, A., Kujala, T., Tervaniemi, M., Kujala, A., Schröger, E., 2004. From symbols to sounds: visual symbolic information activates sound representations. Psychophysiology 41, 709–715.

Widmann, A., Gruber, T., Kujala, T., Tervaniemi, M., Schröger, E., 2007. Binding symbols and sounds: evidence from event-related oscillatory gamma-band activity. Cereb. Cortex 17, 2696–2702.

Winkler, I., 2007. Interpreting the mismatch negativity. J. Psychophysiol. 21, 147–163.

Winkler, I., Czigler, I., 1998. Mismatch negativity: deviance detection or the maintenance of the 'standard'. Neuroreport 9, 3809–3813.

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

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.

Ylinen, S., Huotilainen, M., 2007. Is there a direct neural correlate for memory-trace formation in audition? Neuroreport 18, 1281–1284.

Yvert, B., Crouzeix, A., Bertrand, O., Seither-Preisler, A., Pantev, C., 2001. Multiple supratemporal sources of magnetic and electric auditory evoked middle latency components in humans. Cereb. Cortex 11, 411–423.