Action-related auditory attenuation: event-related potential studies

János Horváth

Cognitive Neuroscience I. Research Group
Institute of Cognitive Neuroscience and Psychology
Research Centre for Natural Sciences
Hungarian Academy of Sciences

Akadémiai doktori értekezés

Budapest, 2014

Acknowledgements

I am indebted to my mentors, Prof. István Czigler and Prof. István Winkler for their trust and continuous support over the years. They provided a background which gave me many opportunities to grow, to get to know the scientific community, and to think freely without losing the rigorous scientific stance. It has been a privilege to work with them over the years, and I am grateful for their patient guidance, the meticulous methodological training and the passionate discussions.

I owe my warmest thanks to my third scientific role model, Prof. Erich Schröger for welcoming me in his research group, and for hosting me, and my family in Leipzig numerous times in the past fourteen years. I am grateful for his friendship, open-mindedness, and generosity to embrace my - often conflicting - approaches and views on topics pioneered by him.

I am indebted to my colleagues, Burkhard Maess, Pamela Baess, Annamária Tóth, and Annamária Burgyán for their contributions to parts of the work presented here. I am also grateful to Zsuzsanna D`Albini, Judit Roschéné Farkas, Yvonne Wolff and Zsuzsanna Kovács for assistance in data collection, and to Viktor Major for technical assistance. Although their hard work is reflected in the present thesis, all errors are solely mine.

I am grateful for the financial support for my research over the years from the Hungarian Scientific Research Fund (Országos Tudományos Kutatási Alapprogramok), the Hungarian National Research and Development Programs (Nemzeti Kutatási és Fejlesztési Program), the European Community’s Seventh Framework Programme (MEIF-CT-2006-023924, PERG04-GA-2008-239393), the Deutscher Akademischer Austausch Dienst and the Hungarian Scholarship Board (Magyar Ösztöndíj Bizottság, MÖB-P/853) and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

Research would have been impossible without the support and love of my wife, Krisztina Bali, our children, Lili and Emília, and my parents. Thank you for all your patience, inspiration, and support that allowed me to dedicate time to this effort.

Contents

PART I.

An introduction to action-related auditory attenuation

Hearing provides a constant stream of information about the events of the environment. Filtering relevant pieces of information out from this flood is made possible by constantly updated, predictive sensory models, which capture regular aspects of the sound environment. Sounds, however, are not only generated by external sources, but also by ourselves, and by our interactions with the environment. We move around, interact with objects, talk, and perform various actions which result in predictable and unpredictable sound events. Some of these sounds may be useful: we can rely on the clicking of the keyboard to monitor that we have pressed a key with sufficient force, and a beeping sound may assure us that a touch screen registered our interaction with it. In other situations, the sounds generated by our own actions may make listening more difficult by masking external sounds.

The studies presented in this thesis investigated how voluntary actions and interactions with the environment influenced auditory processing in humans.

The last five years have seen a burst of effort in the research on action-related changes in auditory processing. Most studies in this field analyze event-related brain potentials (ERPs) recorded in various experimental psychological paradigms. The goal of the present thesis is to provide an overview of this endeavor, present one line of research in great detail, and suggest a synthesis – a tentative framework integrating the seemingly diverging results. Although one wished that research produced more answers than questions, the research presented here led to the questioning of the current scientific consensus on the cause of action-related auditory attenuation, and resulted in a host of new questions and a number of novel hypotheses.

Goal-directed behavior is impossible without knowing the environment and the consequences of our potential actions. Recent theoretical approaches to human auditory perception emphasize that the auditory system relies on a continuously updated dynamic model of the auditory environment, which allows the prediction forthcoming auditory events. These predictions form the bases of the information filtering capability of the auditory system by calling capacity-limited attention- and cognitive control processes only for stimuli which are incompatible with these predictions (Bendixen, SanMiguel, Schröger, 2012), thereby allowing continuous, undisrupted goal-directed behavior. Based on our goals and on our abstract knowledge about the auditory environment, we can also influence auditory processing in a top-down manner: establishing a selective attention set allows us to process task-relevant sounds more efficiently, while suppressing the processing of task-irrelevant sounds (Hillyard, Hink, Schwent, & Picton, 1973; Okamoto, Stracke, Wolters, Schmael, Pantev, 2007).

Beside these well-known ways in which auditory processing can be tuned in accord with our goals, an idea that recently gained popularity is that predictive modeling in audition also relies on *internal forward modeling* of the auditory consequences of voluntary actions. In brief, it is assumed that actions and their immediate sensory consequences are represented by internal forward models (Miall & Wolpert, 1996). When engaging in an action, internal forward models allow the adjustment of sensory processing to accommodate the predictably occurring sensory events due to the action itself (reafference, Crapse & Sommer, 2008b). Similarly to the dynamic models of the auditory environment, this type of modeling would also strongly support the information-filtering capacity of the auditory system, and the maintenance of undisrupted goal-directed behavior. There is strong evidence that forward modeling plays an important role in sensorimotor integration (Wolpert, Ghahramani, & Jordan, 1995), and predictions provided by internal forward models are also used by various cognitive subsystems beyond those directly involved in the control of the given effector (Davidson & Wolpert, 2005).

Because forward models capture causal action-effect mappings, it seems plausible that forward modeling may play a role whenever voluntary actions produce consistent patterns of reafference. In the auditory modality a number of studies reported that self-generated speech sounds elicited attenuated auditory N1 event-related potentials (ERPs, Ford & Mathalon, 2004; Flinker, Chang, Kirsch, Barbaro, Crone, & Knight, 2010), in comparison to that elicited by the same sounds when they were only listened to. Because N1 reflects auditory event-detection and sound feature processing (Näätänen & Winkler, 1999), it was generally assumed that these ERP attenuations reflected the central cancellation of auditory reafference, that is, the workings of an internal forward model. Interestingly, N1 (and P2) attenuation was found for button-press-induced sounds as well (Schäfer & Marcus, 1973; Martikainen, Kaneko, & Hari, 2005). These findings are generally explained in the forward modeling framework, and form the bases of a number of recent contributions to research on speech production (Hickok, 2012), or understanding sensory deficits in schizophrenia (e.g. Ford, Gray, Faustman, Roach, & Mathalon, 2007), for example.

The five studies presented in this thesis investigated the fundamental assumption that action-related auditory ERP attenuation reflected functions related to action-sound contingency representations. The results showing that actions randomly coinciding with sounds presented in an independent sequence result in attenuated auditory ERPs, challenge the current contingency-representation-based explanations of action-related auditory ERP attenuation, and lead to a series of hypotheses replacing or extending the forward-modeling account. The studies provided evidence compatible with the notion that auditory ERP attenuation reflects central processes, and suggested that well-known auditory selective attention effects did not substantially contribute to the action-related ERP attenuations. The exploration of these hypotheses lead to a detailed characterization of action-related auditory ERP attenuation phenomena, including the action-related attenuation of the T-complex, and the dissociation of the P2 from the N1 ERP waveform.

The results are integrated in a tentative interpretational framework suggesting that ERP attenuation effects obtained in the coincidence paradigm reflect internal forward modeling functions, whereas experiments utilizing contingent stimulation protocols reflect additional preparatory attention effects, or preactivation of the sensory effects related to the cognitive representation of actions in the given task setting.

The late auditory event-related brain potentials

Auditory stimulation is reflected in the electroencephalogram (EEG, Davis, 1939). By averaging EEG epochs time-locked to sound events, a number of ERP waveforms can be observed following the onset of a change in auditory stimulation, some of which originate from the cortex (see e.g. Geisler, Frishkopf, & Rosenblith,1958; Rapin, Schimmel, Tourk, Krasnegor, & Pollak, 1966, for a sumary see Näätänen and Picton, 1987). The present thesis focuses on the N1 and P2 waveforms of the so-called *late* part of the auditory ERP (Davis & Zerlin, 1966). The N1 is most prominently observable as a fronto-centrally negative waveform peaking at around 80-150 ms following change onset. P2 is a centrally peaking positive waveform following the N1 at around 150-200 ms from the change onset. The following summary focuses mainly on N1, because our knowledge on P2 is rather limited.

The auditory N1 has been the subject of numerous studies. In the following the aspects of this research relevant to the present thesis are summarized, following the influential review of the literature by Näätänen and Picton (1987), and Näätänen and Winkler (1999).

The N1 waveform comprises at least three subcomponents:

1) When sounds follow long (> 5-6 s) silent periods, the elicited N1 waveform is dominated by a centrally maximal (negative) subcomponent, the non-specific N1 (Hari, Kaila, Katila, Tuomisto, & Varpula, 1982, see also Näätanen, 1988), which is also elicited by non-auditory stimuli (Davis, Davis, Loomis, Harvey, Hobart, 1939; Davis & Zerlin, 1966).

2) When sounds are presented with short (< 4 s) inter-stimulus intervals, the N1 waveform shows a more frontal or fronto-central distribution. When the EEG is recorded with a nose reference, the fronto-centrally peaking negative waveform often shows a polarity inversion at mastoid sites (Vaughan & Ritter, 1970), suggesting that one of the subcomponents originates from a supratemporal generator structure. Indeed, event-related magnetic field (ERF) counterparts of N1 (the N1m) recorded with MEG closely correspond to fields generated by tangentially oriented dipoles located in the supratemporal auditory cortex (see Eberling, Bak, Kofoed, Lebech, & Særmark, 1980; Hari, Aittoniemi, Järvinen, Katila, & Varpula, 1980).

3) The T-complex (Wolpaw & Penry, 1975) comprising a positivity at around 90-100 ms (Ta), and a negative peak at around 140-150 ms (Tb), which are most readily observable at the temporal (T3 and T4, according to the 10-20 electrode-placement system, see Jasper, 1958) electrodes. The T-complex is thought to be generated in

secondary auditory cortices (e.g. Scherg and von Cramon, 1985, 1986; Ponton, Eggermont, Khosla, Kwong, Don, 2002).

The parameters of N1 reflect the characteristics of the eliciting sound-event. For example, sounds of higher intensity elicit N1s with higher amplitudes and shorter latencies (Bak, Lebech, & Saermark, 1985; Lütkenhöner & Klein, 2007), similarly to larger changes in tone frequency (Dimitrijevic, Michalewski, Zeng, Pratt, & Starr, 2008). N1 is also sensitive to stimulus presentation rate: both non-specific and supratemporal N1 subcomponents increase with increasing inter-sound-intervals (Davis, Mast, Yoshie, & Zerlin, 1966). It is important to note that beside a between-condition comparison of N1s elicited by sound sequences presented at different rates, N1 amplitudes also change *within* a sound sequence: The first sound within a train of sounds elicits a higher N1-P2 amplitude than the second, with consecutive sounds eliciting lower (or equal) amplitude ERPs (Ritter, Vaughan, Costa, 1968; Fruhstorfer, Soveri, & Järvilehto, 1970; Fruhstorfer, 1971), which is attributed to refractoriness (Näätänen & Picton, 1987): a lower responsiveness of neural structures generating the N1. Refractoriness of the N1-generators is also stimulus specific: pure tones closer in frequency to an immediately preceding tone elicit lower amplitude N1 than those further away (e.g. Butler, 1972). Based on these results, auditory N1 is generally interpreted as a reflection of auditory event-detection and sound feature processing (Näätänen & Winkler, 1999).

Whereas N1 has been the subject of numerous studies, which yielded a body of data allowing the establishment of a systematic view of the waveform and its subcomponents, not much is known about P2 (for a review, see Crowley & Colrain, 2004). Although early studies often used N1-P2 peak-to-peak measurements, a few studies show that the N1 and P2 waveform (or at least some of their subcomponents) reflect different processes (e.g. because of their differing sensitivity to temporal separation from preceding stimuli: Roth, Krainz, Ford, Tinklenberg, Rothbart, & Kopell, 1976; or differential impact of lesions on them: Knight, Hillyard, Woods, & Neville, 1980).

Only few studies attempted speculations regarding the functional significance of P2. Currently, the most likely possibility seems to be that P2 reflects a process supporting auditory perceptual learning. Exposure to, and interaction with sounds in a perceptual task context leads to enhanced P2 ERPs even when the sounds are presented in an inattentive situation, and this enhancement is retained for months (Seppänen, Hämäläinen, Pesonen, & Tervaniemi, 2012; Tremblay, Ross, Inoue, McClannahan, & Collet, 2014).

The attentional enhancement of N1[1]

The utility of the N1 waveform for cognitive psychological research comes - in part - from the fact that N1 does not only reflect the physical parameters of the stimulation, but it is also sensitive to the cognitive state of the participants, that is, it is also an endogenous ERP. In the following, the two cognitive effects on the N1 (and P2) central to the present thesis are summarized: the N1-enhancement related to selective attention, and the action-related N1-attenuation.

Following several early experiments (for a summary, see Näätänen & Picton, 1987) the effect of selective attention on the auditory N1 was demonstrated by Hillyard, Hink, Schwent, & Picton (1973). They recorded ERPs to relatively rapid (250-1250 ms inter-onset intervals) sound sequences presented in the left and right ears, and instructed participants to detect rare target stimuli in the sequence presented in one ear. It was found that sounds (and not only target sounds) presented in the attended ear elicited higher-amplitude N1 waveforms than those presented in the unattended one. Similar N1 (and N1m) enhancements has been found for attention sets induced by various types of task demands (e.g. for attended ears: Rif, Hari, Hämäläinen, & Sams, 1991; Woldorff & Hillyard, 1991; for attended frequencies: Kauramäki, Jääskeläinen, & Sams, 2007; Kauramäki, Jääskeläinen, Hänninen, Auranen, Nummenmaa, Lampinen, & Sams, 2012; Okamoto, Stracke, Wolters, Schmael, & Pantev, 2007; attended frequency or ear: Ozaki, Jin, Suzuki, Baba, Matsunaga, & Hashimoto, 2004; a given attended moment in time: e.g., Lange, Röder & Rösler, 2003; for a recent summary see Lange, 2013).

The correspondence between N1 amplitude and attention (i.e. the sound event being attended or not) allows the utilization of N1 amplitude measurements to *monitor the focus of input attention* in relation to a sound event: higher N1 amplitudes signal that attention was focused on the sound event eliciting the N1, whereas lower N1 amplitudes signal that attention was not focused on the event. The removal of attentional N1 enhancement by distracting stimuli can also be used to monitor the unfolding of involuntary attention changes: task-relevant sound events following rare, unpredictably occurring distracting events elicit lower-amplitude N1 (and P2, Horváth & Winkler, 2010; Horváth, 2014a, 2014b). In other words, N1 waveform amplitude contrasts can be used to assess whether the input attention set was optimal for the processing of the given stimuli or not.

The nature of the attentional enhancement of the N1 waveform has been the topic of a long debate, which lead to a consensus that the enhancement of the N1 waveform is due to the superposition of different ERPs or ERP-effects: Beside a “genuine” N1-modulation, an ERP of different origin (Alho, Paavilainen, Reinikainen, Sams, & Näätänen, 1986; Knight, Hillyard, Woods, & Neville, 1981) overlapping the N1 - termed *processing negativity* (PN, Näätänen, Gaillard & Mäntysalo, 1978) or

---

[1] Based, in part, on the introduction in Horváth, J. (2014) Probing the sensory effects of involuntary attention change by ERPs to auditory transients: Probing the sensory impact of distraction. *Psychophysiology, 51(5)*, 489–497. doi:10.1111/psyp.12187

negative difference (Nd, Hansen & Hillyard, 1980) - also contributes to the amplitude increase. Whereas the "genuine" auditory N1 modulation is assumed to reflect the attentional amplification of the task-relevant aspects of auditory input processing on the sensory level, PN (or Nd) is thought to reflect additional, voluntary, task-relevant processing beyond the registration of the auditory event (Näätänen & Michie, 1979), possibly related to matching the event to a voluntarily maintained stimulus template (attentional trace, Näätänen, 1982, 1990).

Action-related N1 and P2 attenuation

The focus of the present thesis is the action-related auditory ERP attenuation, which was first reported by Schafer & Marcus in 1973. They recorded ERPs to click sounds initiated by the participants' own key-presses. The ERPs occurring later than 100 ms (starting with N1) elicited by self-induced clicks were smaller than those elicited by the playback of the previously self-produced sound sequence. These results were explained in terms of uncertainty: It was suggested that because participants initiated the sounds themselves, uncertainty regarding the stimulation could be reduced, which, conversely, allowed the preservation of processing resources. Following-up on, and extending the experimental paradigm used by Schafer & Marcus (1973), McCarthy & Donchin (1976) replicated the initial results regarding the attenuation of the N1 waveform (but did not find significant attenuation for P2). The N1 effect was found in a number of further studies administering stimulation arrangements in which sounds were induced by button-presses (Martikainen, Kaneko, & Hari, 2005; Ford, Gray, Faustman, Roach, & Mathalon, 2007; Baess, Jacobsen, & Schröger, 2008; Aliu, Houde, & Nagarajan, 2009; Baess, Horváth, Jacobsen, & Schröger, 2011; Knolle, Schröger, Baess & Kotz, 2012; Sowman, Kuusik, Johnson, 2012; Knolle, Schröger, & Kotz, 2013a, 2013b; Ford, Palzes, Roach, & Mathalon, in press).

Interestingly, a number of studies reported N1-attenuations to probe tones presented during vocalizations (Ford, Mathalon, Kalba, Whitfield, Faustman, Roth, 2001), or that elicited by self-produced speech sounds (Curio, Neuloh, Numminen, Jousmäki, & Hari, 2000; Houde, Nagarajan, Sekihara & Merzenich, 2002; Ford & Mathalon, 2004; Heinks-Maldonado, Mathalon, Gray & Ford, 2005; Heinks-Maldonado, Nagarajan & Houde, 2006; Ford, et al., 2007; Ventura, Nagarajan, Houde, 2009; Flinker, Chang, Kirsch, Barbaro, Crone, & Knight, 2010; Niziolek, Nagarajan, & Houde, 2013; Sitek, Mathalon, Roach, Houde, Niziolek, & Ford, 2013).

Although most of these studies did not explicitly investigate whether attenuation occurred in the P2 time interval, P2-suppression effects could be seen on the ERP figures of such studies as well (see e.g. Ford & Mathalon, 2004; Baess et al., 2011). A number of recent studies (including the studies presented in detail in this thesis) consistently demonstrated that action-related P2 attenuation generally co-occurs with N1 attenuation.

As described before, the scalp-recorded, fronto-centrally or centrally negative N1 ERP sums multiple subcomponents. In contrast, the magnetic counterpart of N1 (N1m) mainly reflects the activity of the supratemporal generator (N1m, Näätänen, 1988) due to the relative "insensitivity" of magnetoencephalography (MEG) to non-tangentially oriented, and deeply located dipolar sources (Hämäläinen, Hari, Ilmoniemi, Knuutila, Lounasmaa, 1993). Therefore, measuring ERF is useful in assessing the contribution of the supratemporal (auditory) component to N1-effects. Because many of the studies reporting action-related N1 attenuation used MEG methodology, it is generally assumed that N1(m)-attenuations reflect, at least in part, the attenuation of the supratemporal generator, that is, auditory processing.

It seems, however, likely that the non-specific component is attenuated as well. In their seminal study, Schafer & Marcus (1973) suggested that it was the non-specific N1 subcomponent that was attenuated, because self-induced visual stimulation also resulted in the attenuation of the N1 measured at the central leads. Most studies utilize relatively long (typically around 4 s) between-sound intervals, for which the contribution of the non-specific subcomponent to the N1 waveform is already substantial (Näätänen, 1988). Using 0.8, 1.6, or 3.2 s. inter-sound (and corresponding between-action intervals) SanMiguel, Todd & Schröger (2013) found that the magnitude of N1 attenuation decreased with decreasing inter-sound interval, and suggested that the N1-effect might be due to the attenuation of the non-specific component.

In some cases, ERPs may also allow conclusions regarding the involvement of the supratemporal generators, because N1 often exhibits a polarity reversal at the mastoids when the EEG is recorded with a nose-reference (Vaughan and Ritter, 1970). An N1-effect showing such a polarity reversal signals that the effect (at least in part) originates from the supratemporal generator. The lack of a polarity reversal, on the other hand, does not mean that the supratemporal component is not involved, since it may simply be overlapped by the non-specific N1 subcomponent or other ERPs.

To-date most studies on auditory ERP attenuation do not allow for a clear-cut separation of the two subcomponents, and N1-attenuation is interpreted as a general reflection of attenuated sensory (i.e. auditory) processing.

Interpreting action-related N1 and P2 attenuation in the framework of internal forward modeling

The most widely accepted interpretation of action-related N1-attenuation is that it reflects the *cancellation of auditory reafference caused by our own actions*.

Many studies suggest that causal relationships between self-produced movements and their sensory effects are represented by *internal forward models* (Miall & Wolpert, 1996). When engaging in an action, a copy of the outgoing motor commands (the *efference copy*, von Holst, & Mittelstaedt, 1950) is produced, which is translated by an *internal forward model* into a special sensory signal representing the associated consequences (a *corollary discharge*, Sperry, 1950; for a recent summary, see Crapse & Sommer, 2008a, 2008b). The corollary discharge is special, because it does not only allow one to compare the predicted sensory consequences of the action with the actual sensory input *after* the action took place (feedback), but it can also be used *in parallel with the action* to adjust sensory processing so that it can accommodate (some of the) predictably occurring sensory events due to the action itself. Although many studies formulate this as an explicit comparison process between the predicted and the actual re-afference, such a process would defeat the purpose of a forward model, that is, that one does not have to wait for feedback when the action produces the predictable consequence before continuing with the next action (see Miall & Wolpert, 1996; Hickok, 2012).

Forward modeling supports many functions of the neural system. There is strong evidence that forward modeling of the sensory consequences of self-produced movements plays an important role in sensorimotor integration (Wolpert, Ghahramani, & Jordan, 1995). Beside the sensory input, forward modeling adds a source of information which can be used to improve movement performance (Vaziri, Diedrichsen, Shadmehr, 2006; Shadmehr, Smith, & Krakauer, 2010). Forward models also make it possible to *cancel reafference*, that is, the stimulation inherently resulting from the action due to mechanics of the actor's own body. Whereas it seems plausible that proprioceptive and skin-stretch reafference for muscle movements (Proske & Gandevia, 2012) originating from the moving bodyparts is represented by forward models, predictions provided by internal forward models are also used by various cognitive subsystems beyond those directly involved in the control of the given effector (Davidson & Wolpert, 2005). For example, ticklishness on the palm (as well as the concurrent activation in somatosensory cortex) is reduced when the stimulation is self-produced (e.g. Blakemore, Wolpert, & Frith, 1998). Active, voluntary head-movements lead to the cancellation of reafference in the vestibular system (Cullen, 2004; Roy & Cullen, 2004; Cullen, Brooks, Sadeghi, 2009). Forward modeling supports the stabilization of the visual field despite eye-movements (Duhamel, Colby, Goldberg, 1992); and arm-movements influence the saccadic eye-movement system (Ariff, Donchin, Nanayakkara & Shadmehr, 2002; Thura, Hadj-Bouziane, Meunier, & Boussaoud, 2011), as well as motor imagery (Gentili, Cahouet Ballay & Papaxanthis, 2004).

Because forward models capture causal (*contingent*) action-effect mappings connecting different cognitive subsystems, it seems plausible that forward modeling may also play a role in most functions where our own controlled actions produce consistent patterns of stimulation. Because N1(m) reflects the detection and feature-specific processing of auditory events, it is generally assumed that action-related N1 (and P2) -attenuation reflects the *cancellation of auditory reafference*. Because we have extensive experience with the control of, and sensory stimulation produced by our own speech production system, this notion is especially plausible in the case of speech-related N1-attenuation. The phenomenon is also often referred to as N1/P2 (or sensory) *suppression* probably to indicate that the effect on sensory processing is the goal/result - and not a side effect - of the processes involved. In the following, the more neutral *attenuation* will be used.

Studies showing N1-attenuation for non-speech-related actions are also explained in this framework. The core assumption of these studies is that capturing a contingent action-stimulus relationship occurs rapidly, at least within the order of minutes (Aliu et al., 2009), and the resulting forward model is then used to derive predictive sensory information, which is manifested in the attenuation of the N1 response. There is some evidence that this attenuation only happens for voluntary, but not for involuntary actions (e.g. those induced by transcranial magnetic stimulation, Timm, SanMiguel, Keil, Schröger, & Schönwiesner, 2014).

The promise of this line of research is that action-related auditory (ERP) attenuation may reflect functions related to agency attribution, that is, the attribution of sensory events to our own actions or to external sources. This notion is mostly suggested by studies showing that the magnitude of N1 attenuation is smaller in patients with schizophrenia (e.g. Ford et al., 2007, and Ford et al., in press; for a recent review see Ford & Mathalon, 2012). Although the argumentation is somewhat circular (in the sense that attenuation is regarded both as a result of the prediction *and* an indicator that the sensory events were self-produced), some studies show results compatible with this notion. Baess et al. (2011) found that when the same sound was brought about by one's own key-presses and also by external sources (in a mixed condition), N1 attenuation for the self-induced sounds increased in comparison to the condition in which the sounds were elicited only by the participant's actions. That is, auditory attenuation may reflect a mechanism enhancing the discriminability of self-induced and external stimulus events.

The neural underpinnings of action-related auditory (ERP) attenuation are not well-researched. Apart from a single, low-sample positron emission tomography-based study (Blakemore, Rees and Frith, 1998) administering conditions manipulating temporal sound predictability and self-induction showing that these resulted in different, interacting patterns of brain activations, not much is known about the neural structures contributing to the attenuation effects. Recent evidence suggests that contingent action-related auditory ERP attenuation may be subserved by cerebellar functions (Knolle et al., 2012; Knolle, Schröger & Kotz, 2013b): patients with cerebellar lesions did not exhibit N1 attenuation, which is compatible with the notion that internal forward models of the cerebellum (Wolpert, Miall, & Kawato, 1998) also provide auditory predictions that is action-contingent auditory processing adjustments (Knolle et al., 2012).

Measuring auditory processing activity in the presence of on-going action

Beyond the general problem of relating physiological or behavioral measurements to functional activity, measuring auditory processing activity in the presence of concurrent action-related activity is not trivial. In the simplest case, two measurements from separate conditions are compared: in one condition, the sound is presented in the absence of actions, in the other the sound is presented concurrently with, or at least in close temporal proximity to, an action. For some measurements and methods, this direct comparison may yield valid results, depending on the selectivity of the measure. For example, high spatial resolution (e.g. electrocorticography) may allow one to characterize auditory processing activity in the auditory cortex alone. Also, as mentioned previously, the relative "blindness" of MEG to certain sources of brain activity may allow a characterization of auditory processing without much interference from non-auditory activities. Generally, ERPs do not allow a clear identification of auditory activity in the two conditions: non-auditory response contributions to the ERP elicited by the action-sound event may contaminate the results.

The most widely used paradigm to investigate action-related auditory attenuation, the *contingent paradigm*, features three conditions (McCarthy & Donchin, 1976) to overcome the measurement problem described above, which are administered in separate experimental blocks:

1) In the Motor-Auditory condition, the participant voluntarily repeats an action from time-to-time according to the instructions. In this condition, each action leads to the presentation of a sound.

2) In the Auditory condition, a sound sequence is presented, and the participant listens to these without performing any actions.

3) In the Motor condition, the participant voluntarily repeats an action according to the instructions, which are the same as in the Motor-Auditory condition, but no sound is presented.

To estimate the auditory ERP contribution to the ERP recorded in the Motor-Auditory condition, the ERP recorded in the Motor condition is subtracted from that recorded in the Motor-Auditory condition. This *motor-corrected* waveform is compared to the one obtained in the Auditory condition.

Simple as it seems, this arrangement has a number of limitations.

First, because N1 amplitude increases with increasing inter-sound interval (Davis, Mast, Yoshie, & Zerlin, 1966), it is important to make sure that inter-sound intervals are not longer in the Auditory than in the Motor-Auditory condition, because this could result in an *apparent* N1-attenuation in the critical comparison. Because the inter-sound intervals in the Motor-Auditory condition are inherently variable, most studies (e.g. Schafer & Marcus, 1973; Martikainen et al., 2005; Baess, et al, 2009) present *replays* of the sound sequences generated by the same participant in the Motor-

Auditory experimental blocks. On one hand, this is an ideal choice, since the sequence of sounds (and inter-sound intervals) is the same in the two conditions. On the other hand, amplitude differences in the critical comparison could be attributed to the presentation order of the experimental blocks, because the Auditory (replay) blocks always follow the corresponding Motor-Auditory blocks. This confound can be reduced when multiple blocks are presented in an interwoven order.

Second, it needs to be decided what instructions the participants receive in the three conditions. Obviously, it is not possible to give participants the same task in the three conditions: in the Motor and Motor-Auditory conditions a sequence of actions is performed (e.g. press a button isochronously), but participants have to do something else in the Auditory condition. In most studies, the instruction in the Auditory condition is "passive listening", that is, participants do not perform any specific task regarding the sounds, only behave as requested by the experimenter to ensure good EEG-signal quality (minimize motion, reduce eye-movements, etc.). Since the task setting is different, it is difficult to attribute between-condition differences solely to action-related sensory attenuation, especially, because the N1 amplitude changes as a function of attentional focus (i.e. attended sounds elicit higher amplitude N1s than those not in the focus of attention, see, e.g. Hillyard, et al., 1973). In a recent design, Saupe, Widmann, Trujillo-Bareto & Schröger (2013) attempted to reduce between-condition differences by instructing participants to produce a sequence of button-presses with highly variable between-sound (i.e. between-action) intervals within a 1.8-5.0 s range in the Motor-Auditory condition. When listening to the replay, participants had to detect too short (< 1.8 s) or too long (> 5.0 s) between-sound intervals. This instruction ensures that participants attend to the same temporal feature of the sequence in both conditions, therefore, presumably, reduces between-condition differences. In our own experiments (see Study III and V, Horváth & Burgyán, 2013; Horváth, in press), we instructed participants to count the sounds produced by the actions in the Motor-Auditory condition, because only the first 60-70 actions produced a sound. Similarly, participants had to count the sounds in the Auditory condition (also a random number of sounds were presented, the replay of the first 50-60 sounds from the Motor-Auditory condition). The analyses showed that more miscounting errors were made in the Motor-Auditory condition, probably because the Motor-Auditory condition had been a dual-task, whereas the Auditory condition had been a single-task situation.

Third, when estimating the auditory ERP contribution to the Motor-Auditory ERP, it is assumed that the action-related (ERP) activity is the same when the actions consistently lead to the presentation of the sound (Motor-Auditory condition), and when they do not (Motor condition). Although this may be true, currently, there is no direct evidence for, or against this assumption. Studies reporting statistics on action timing do show between-condition differences, although the signs of these differences seem to be instruction-dependent: When participants counted sounds, Horváth & Burgyán (2013) and Horváth (in press) consistently found faster action pace in the Motor-Auditory than in the Auditory condition. When participants detected too short or too long inter-tone-intervals, a slower average pace in the Motor-Auditory than in the Auditory condition was found (Saupe et al., 2013).

In summary, the separation of the three conditions allows the attribution of *apparent* action-related ERP attenuations to a number of factors related to task-setting differences; that is, attributing such differences solely to action-sound contingency representations is generally not warranted.

Behavioral correlates of action-related auditory ERP attenuations

If auditory N1 and P2 attenuation reflects the cancellation of sound reafference, it seems plausible that this should have an effect on the behavioral indices of sound processing as well. Surprisingly, there are only few studies addressing this issue.

An intuitive first approach (based on the general interpretation of N1-attenuation) would suggest that N1-attenuation – as a reflection of attenuated sound-event detection or sound feature processing – would be manifested in worse perceptual performance in the detection of the given sound. I am unaware of any studies showing such effects in direct detection tasks. Our own unpublished experiments (Neszmélyi, 2014) showed that detection thresholds for self-generated pure tones presented in white noise were not substantially different from that of those only listened to.

The failure of the direct approach is easy to explain in a predictive framework. Although a predictive mechanism may allow one to reduce the sensory processing activity for a given sensory event, it would be highly dysfunctional if the prediction would result in behavioral performance loss when the predicted reafference is directly relevant to the ongoing task. That is, although the ERPs show a reduction of processing activity, this probably reflects an *efficiency gain*, and not a performance loss (i.e. less processing effort with similar performance).

More sophisticated experimental designs, however, showed effects resembling auditory ERP attenuation-effects: For example, some studies showed that self-generated sound were perceived as being softer. Because softer sounds elicit smaller N1s, such effects could be behavioral manifestations of action-related N1-attenuation. Sato (2008) found that loudness difference judgments of tone-pairs was influenced by the agent triggering the first tone: when the first tone was initiated by the participant‘s own or by the experimenter’s hand-movement, these sounds were judged to be softer than in those cases when no visible action, or robotic arm-movements triggered the first sound. Similarly, participants in the experiment by Desantis, Weiss, Schütz-Bosbach, and Waszak (2012) performed loudness difference judgments for sound pairs, in which the first sound was either believed to be self- or experimenter-initiated. It was found that the participants’ points of subjective loudness equality were lower when they believed that they initiated the sounds, suggesting that these sounds were perceived as being softer.

Although it is not impossible that these experiments reflect genuine internal forward modeling effects, the cause of these effects is difficult to determine unequivocally. Since in both studies participants were aware of the agent triggering the first sound, or were made to believe that a given agent was about to produce the first sound, these effects could also reflect top-down, voluntary attention set changes. Since voluntary attention sets (and the ways we represent actions in terms of their consequences, see below) systematically change with the (assumed) agent and our perceived relationship with it (see Colzato, de Bruijn, & Hommel, 2012; Colzato, Zech, Hommel, Verdonschot, Wildenberg, & Hsieh, S, 2012; Dolk, Hommel, Prinz, & Liepelt, , 2013), these results do not necessitate the involvement of a internal forward models.

Actions in action-related auditory attenuation research

Currently, action-related auditory attenuation is investigated using only two types of actions: vocalizations, and finger-movements resulting in button-presses. Although it is generally assumed that sensory attenuations measured in paradigms using these two types of actions reflect the same mechanisms, evidence for this is scarce. Some studies find similar differences in auditory attenuation patterns between normal participants and patients with schizophrenia for speech-production and button-press actions (i.e. that the magnitude of attenuation is reduced for the patients, Ford et al., 2007, and Ford et al., in press), but direct evidence for this is yet unavailable. Since speech is probably be *the* best example for a behavior producing consistent patterns of self-stimulation, it seems plausible that forward modeling would play a role during speech production, and speech-related N1-attenuation effects reported in the literature may well reflect the workings of an internal forward model. For settings with arbitrary, non-speech-related actions and contingent (speech or non-speech) sounds, which are associated only for short periods (typically for about 5-20 minutes during the experiments), this seems less plausible.

When one considers the methodological bases of assessing action-related sensory attenuation using these two actions, one finds exactly the opposite pattern: vocalization-based arrangements often yield results much more open to discussion and interpretation than finger-movement-based designs. Measuring auditory processing activity in the presence of concurrent speech-production is – to say the least – is difficult, and most experiments do not allow firm conclusions regarding the cause of the measured auditory processing differences. In the context of a contingent paradigm, it is difficult to create an Auditory condition which would perfectly match the sounds of the Motor-Auditory condition, because for speech, the sound source is within our own body. Because physical differences in the eliciting stimuli have a strong impact on the late auditory ERPs, results obtained in such designs may be biased, and are difficult to interpret. Creating a proper Motor condition is also difficult, because the action-related activity would not match that in the Motor-Auditory condition.

I am unaware of studies directly testing whether vocalization and finger-movement-based paradigms reflect the same mechanisms, that is, it is unclear whether the results obtained in relatively well-controlled finger-movement-based designs can be generalized to speech-production. On one hand, it would seem highly redundant if the highly similar event-related response effects occurring in similar action-stimulus contexts would be produced by different subsystems. On the other hand, speech may have highly specialized subsystems, which are not readily available for the processing of other types of actions and stimuli (see e.g. Alho, Rinne, Herron, & Woods, 2014). At this point, due to the lack of empirical evidence, this issue cannot be convincingly resolved.

The sensory preactivation account of auditory attenuation

As described earlier, N1-attenuation is generally interpreted as a general reflection of attenuated auditory processing. SanMiguel, Todd and Schröger (2013) speculated that it was the processing activity directly leading to a conscious detection and orientation towards a sound event that was attenuated, and not sound feature-specific processing activities. That is, the attenuation of the non-specific N1 subcomponent would reflect the expectation that “something would happen” for the given action. The prediction-based interpretational framework, however, does not (only) suggest that auditory processing activity will be *generally* attenuated, rather, it suggests that the processing of the action-contingent sounds will be specifically, selectively attenuated. A novel alterative account to the widely accepted forward modeling account of auditory action-related attenuation (the so-called *preactivation* hypothesis, Roussel Hughes, & Waszak, 2013) seems to relate directly to the question of prediction specificity.

The notion that the conditions of the *contingent paradigms* actually feature various differences in “temporal and event uncertainty” has been already suggested by McCarthy & Donchin (1976). In a recent review, Hughes, Desantis, & Waszak (2013a) summarized the types of predictions one can formulate. In their terms, most studies using the *contingent paradigm* feature differences in temporal prediction (knowledge when a sensory event will occur) and temporal control (control over the occurrence of the sensory event – this mostly implies temporal predicition as well). One could however also couple the actions to the identity of the self-induced tones (identity prediction), and instead of comparing the processing of self-induced sounds to that of the same sounds only listened to, one could also compare responses to sounds elicited in two self-induction conditions. For example, one may compare responses to action-*congruent* and action-*incongruent* sounds, that is, sounds which have been previously associated with the action or not. In such studies, the main question is whether the action-related predictability of the stimulation is stimulus specific or not.

Baess, Jacobsen, and Schröger (2008) investigated ERP attenuation to self-induced tones with predictable (always immediately after key-press) and unpredictable onsets (randomly in the 500-1000 ms following the key-press), and also predictable (1000 Hz) and unpredictable (400-1990 Hz) frequencies. They found that N1-attenuation was present in all cases, with stronger attenuation when frequency was predictable. Interestingly, P2-attenuations (which were not investigated) were not visible on the reported ERP waveforms, and the N1-attenuation effect was delayed in comparison to the N1. (Although no significant attenuation-differences related to the onset manipulation were found, it has to be noted that onset manipulation was confounded with temporal expectation effects –which are visible as slow ERP shifts already at the baseline).

In a different setting, a similar effect was demonstrated by Hughes, Desantis, & Waszak (2013b). In their study participants learned and maintained action-sound associations during the experiment. It was found that action-congruent sounds elicited lower amplitude N1s than action-incongruent ones (with no P2-effect visible in the reported waveforms).

SanMiguel, Widmann, Bendixen, Trujillo-Barreto, & Schröger, (2013) found that when a reliable contingent action-sound relationship was present (i.e., actions elicited in the same sound 88% of the time), rare sounds omissions lead to the elicitation of an ERP resembling part of the T-complex. This was not elicited when the actions lead to sounds only 50% of the time, thus suggesting that the action in itself resulted in auditory activation. Interestingly, this effect seemed to be sound-specific: if the action-induced sounds were randomly chosen on each button-press, the effect was no longer observable (SanMiguel, Saupe, & Schröger, 2013).

These results are compatible with the notion that contingent action-sound associations lead to the formation of *sensory templates*, which are activated when the actions are voluntarily produced. Although Baess et al. (2008) and SanMiguel, Widmann, et al. (2013) argued that these contingency-representations are internal forward models, the quickness of the acquisition suggests that these effects may well result from other, attention-related forms of action-effect representations.

A number of studies suggest that during task-performance, actions are represented by their sensory consequences in the cognitive system, and the activation of such representations, and attending to the task-relevant sensory consequence form an integral part of action preparation (Galazky, Schütze, Noesselt, Hopf, Heinze, & Schoenfeld, 2009; Brown, Friston, & Bestmann, 2011). The Theory of Event Coding (TEC, Hommel, Müsseler, Aschersleben, & Prinz, 2001) suggests that actions are encoded primarily (but not exclusively) in their distal, task-relevant consequences. When multiple action-consequences are available, then the one corresponding to the intentional reference frame is dominantly used (Sutter, Sülzenbrück, Rieger, & Müsseler, 2013). Intentional coding allows for considerable (but not unlimited) freedom in representing a task. For example, Hommel (1993) modified the Simon-paradigm (Simon & Ruddel, 1967), in which participants had to respond with left or right key-presses to the pitch of a target tone. The tone, however, was presented on the left or the right. Reaction times typically show a spatial interference: responses are slower if the response is to be given on the side opposite to the (task irrelevant) side of the tone. Hommel showed that if responses are coupled with an action-effect on the action-opposite side (a light emitting diode flash, i.e. pressing the left button results in a flash on the right and vice versa) the spatial interference could be reduced if the instructions emphasized the effect-side instead of the key-press side (for a summary see, Hommel, 2011).

It is important to note that although actions are encoded primarily in their task-relevant sensory consequences, irrelevant sensory consequences are also automatically (that is, without intention) acquired and represented together with the action, and when the action is executed, the associated irrelevant sensory representations are activated as well. It is hypothesized (Hommel, et al., 2001) that the automatic acquisition of task irrelevant effects makes it possible to initiate actions without a known task-relevant effect. Elsner and Hommel (2001) demonstrated that task-irrelevant action consequences (key-presses resulting in tones of different frequencies) are quickly coupled to the respective actions. Being exposed to such irrelevant action-consequences only about 100 times in about 9-10 minutes produced robust action-selection interference effects in the tasks following the initial exposure to these contingencies. That is, action-effect representations are formed without intention to do so, even for *artificial, arbitrary* action-effect contingencies in the typical timeframe used in experiments on auditory attenuation. Based on experiments using visual stimulation, Roussel et al. (2013) suggested that such learned action-effect associations lead to the *preactivation* of the sensory consequences of one's actions, which, in turn, lead to differences in stimulus detection performance, and the corresponding sensory ERPs (Roussel, Hughes, & Waszak, 2014). It seems plausible that sensory processing differences between action-congruent and -incongruent sounds (Hughes, et al., 2013b) reflect mainly action-effect associations acquired in such a way.

PART II.

The introduction of the coincidence paradigm - Experimental studies

Overview

The goal of the five studies presented in the thesis was to investigate the fundamental assumption of research on action-related auditory ERP attenuation, that is, that action-related auditory ERP attenuation reflected functions related to action-sound contingency representations.

Study I describes three experiments utilizing a novel experimental paradigm (the *coincidence paradigm)* which yielded results, which, at first sight, challenge contingency- representation-based explanations. To explain these results, a series of post-hoc hypotheses replacing or extending the forward-modeling account were suggested.

Study II investigated the hypothesis put forward in Study II, that the coincidence-related auditory ERP attenuation could be explained by well-known auditory selective attention-based mechanisms.

Study III investigated the possibility that action-related ERP attenuation was caused by a peripheral mechanism, the reduction of sound transmission efficiency brought about by the co-activation of the middle-ear stapedius muscle with the action-related effector.

Study IV investigated the hypothesis suggested in Study II, that coincidence-related auditory ERP attenuation was caused by a generalized expectation that button-presses (but not –releases) would produce an sensory effect.

Study V investigated whether the coincidence-related auditory ERP attenuation was caused by establishing mechanical contact with an object.

Although the studies use similar methodology, the description of the methods are presented separately (and redundantly) for each study to support readability.

Study I.
A sufficient condition for action-related auditory ERP attenuation[2]

Summary

Study I demonstrated in three experiments using a novel coincidence paradigm that actual contingency between actions and sounds is not a necessary condition for N1-suppression. Participants performed time interval production tasks: they pressed a key to set the boundaries of time-intervals. Concurrently, but independently from key-presses, a sequence of pure tones with random onset-to-onset intervals was presented. Tones coinciding with key-presses elicited suppressed N1(m) and P2(m), suggesting that action-stimulus contiguity (temporal proximity) is sufficient to suppress sensory processing related to the detection of auditory events.

Introduction

The generally accepted interpretation of studies showing action-related N1 (and P2) attenuation is that exposure to a contingent action-stimulus relationship leads to the formation of an action-sound contingency representation – an internal forward model. Upon performing the action, the model produces sensory predictions, which allow the adjustment of the auditory system to accommodate the incoming sound. This is manifested in the attenuation of the sound-related N1-response.

At the time this study was conducted, the studies interpreting auditory N1-attenuation in the framework of internal forward modeling exclusively used *contingent* stimulation: actions always brought about a sound event. In these studies, action-contingent stimulation also involved a consistent temporal relationship between action and stimulus (i.e. stimuli were delivered at least within a couple hundred ms after the action). Therefore, it seems possible that the necessary condition for auditory N1 suppression is not *contingency*, but *temporal contiguity*, that is, the temporal proximity of an action and a sound. That auditory processing may be affected by concurrent, but not causally related motor activity is not without support: Makeig, Müller and Rockstroh (1996) found that the amplitude and phase of the auditory steady state response in the EEG was perturbed by concurrent, voluntary finger movements. Hazemann, Audin, & Lille (1975) presented a sound sequence with random inter-stimulus intervals, and instructed participants to produce an even-paced key-press sequence. They found that the amplitude of the N1 and P2 ERP waveforms elicited by sounds close to key-presses was smaller than for sounds far from key-presses. Whereas Hazemann and colleagues did not directly remove key-press-related ERPs from the sound-locked waveform, the contributions of these ERPs to the N1- and P2- effects

---

[2] Based on Horváth, J., Maess, B., Baess, P., Tóth, A. (2012) Action-sound coincidences suppress evoked responses of the human auditory cortex in EEG and MEG. *Journal of Cognitive Neuroscience, 24* (9), 1919-1931. doi:10.1162/jocn_a_00215

were probably low due to the randomness of key-press-stimulus separation (in a 0-220 ms range).

The goal of the present study was to investigate whether key-press-tone *contiguity* without a key-press-tone *contingency* was sufficient to produce an N1-suppression effect. We utilized a *coincidence paradigm*: participants pressed a button to set boundaries in a time interval production task, while a concurrent, but temporally independent sound sequence was presented with random inter-sound-intervals. This arrangement leads to the occurrence of all relevant events (action only, sound only, and combined action-sound events) in the same experimental condition, which eliminates most of the confounds related to between-condition differences present in contingent designs (see above). No block-order related differences can occur, there are no task differences, and expectations regarding the occurrence of a sound upon pressing the button should be similar for all button presses (as it will be discussed in Study II, if one knows the presentation frequency distribution of the tones, one can actually adopt a strategy to produce actions which are more likely or less likely to coincide with a tone – as Study II. shows, even if one has the explicit task to try such a strategy, this theoretical possibility does not substantially alter coincidence rates).

The main question was whether key-press-sound *coincidences* resulted in attenuated auditory processing as reflected by the N1(m) event-related response. Because there was no contingent action-tone relationship in this setting, a potential N1-attenuation effect could not be attributed to the cognitive system capturing a causal action-tone-relationship in the form of a forward model. We recorded EEG in two experiments (Experiments 1 & 2) with different interval production instructions, and recorded MEG in Experiment 3 (with the same experimental setting as in Experiment 1).

Methods

Participants

Fourteen paid volunteers (6 women, age: 19-24 years, two left-handed) participated in Experiment 1, thirteen (8 women, age: 19-24 years, three left-handed) in Experiment 2, and twenty (10 women, age 23-31, all right handed) in Experiment 3. In all three experiments, participants gave written informed consent after the experimental procedures were explained to them. All participants reported normal hearing status and had no history of neurological disorders.

Stimuli and procedure

In all experiments participants performed time interval production tasks. In Experiment 1 and 3 participants were instructed to produce a sequence of key-presses in which between-key-press intervals showed a uniform distribution between 2 and 6 seconds within each 5-minute-long experimental block. In Experiment 2, on the other hand, a regular, even-paced sequence with a key-press every 4 s was required. In all three experiments, the experimental session started with a training phase, during which participants performed the task with on-line visual feedback: a computer screen showed a histogram of their between-key-press intervals, which was updated after each key-press. During the experimental phase, feedback was provided only at the end of each experimental block.

In Experiments 1 and 2, each participant held a rod-mounted key in their dominant hand and pressed the key with the thumb; in Experiment 3, the key was mounted on a response box, which was positioned under the dominant hand, and participants used their index finger to press the key.

In all three experiments a series of 50 ms long (including 10 ms linear rise and 10 ms linear fall times), 1000 Hz sinusoid tones were presented. Tone intensity was individually adjusted to 60 dB sensation level (SL – above hearing level) in Experiments 1 and 3, and to 50 dB (SL) in Experiment 2. The tones were delivered through headphones (HD-600, Sennheiser, Wedemark, Germany) in Experiments 1 and 2, and through tubal insert phones (TIP-300, Nicolet Biomedical, Madison, WI, U.S.A.) in Experiment 3.

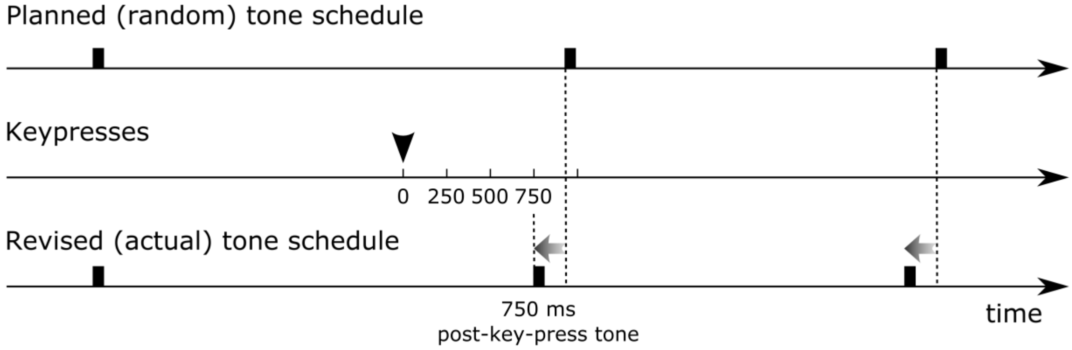

Figure 1.1. Example for the manipulation of the tone presentation in the experiments. When a key-press occurred, the scheduled sounds were shifted towards the key-press so that to the next tone was delivered with a delay of integer multiples of 250 ms. In this example, the next tone would be delivered between 750 and 1000 ms after the key-press according to the planned schedule. This is revised by shifting the tone schedule following the key-press so that the next tone is delivered exactly 750 ms after the key-press (*750 ms post-key-press tone*).

In all three experiments, the schedule of tone presentation was pre-generated for each participant so that the onset-to-onset intervals were random in the range of 2-6 s (with uniform distribution). The experiments were divided into fourteen experimental blocks with 72 tones presented in each of them (1008 tones in total). Between blocks short breaks were taken as needed, with a longer break around the middle of the session (after the seventh block). Key-press-tone coincidences were created through the following manipulation (see Figure 1.1): At every key-press the preplanned tone presentation schedule was revised: The schedule was shifted so that the next tone was presented either right after the key-press (0 ms) or with a delay of integer multiples of

250 ms. That is, if the next tone was to be presented between 0 and 249 ms following the key-press, it was presented right away; if it was to be presented in 250-499 ms, it was presented 250 ms after the key-press, and so forth. If there were more key-presses before a tone, the manipulation was carried out in reference to the last key-press only. The result of this manipulation was that all tones preceded immediately by a key-press were shifted similarly, ensuring that the distribution of the intervals separating these tones from the previously presented tones was the same. In contrast, sounds not preceded by a key-press were not shifted at all; therefore the tone-to-tone intervals preceding these tones were longer. Because N1 amplitude is known to increase with increasing tone-to-tone interval (Näätänen and Picton, 1987), a comparison between shifted and unshifted tones would be confounded by the systematic tone-to-tone interval differences. Therefore, only (shifted) sounds preceded by at least a key-press were included in the analyses.

For coincidences (i.e. when a tone was presented right after a key-press), there was a short delay between the key-presses and the tone due to the necessary processing time after key-presses: this was 4.3 ± 0.1 ms (mean ± standard error, SD) in Experiments 1 and 2; and 9.3 ± 0.1 ms in Experiment 3. These delays were taken into account during the analyses, but for convenience, in the following, we will not include these short delays into the references to key-press-tone intervals, and refer to the corresponding events only as coincidences, 250 ms post-key-press tones, 500 ms post-key-press tones, and so on. Also, due to a programming error, for some of the coincidences a further, additional delay occurred: in Experiment 1 this additional delay was 9.1 ± 0.1 ms, and affected 33 ± 8 % of the cases; in Experiment 2, it was 9.0 ± 0.1 ms, and affected 44 ± 7 % of the cases; in Experiment 3 this was 8.4 ± 0.1 ms, and affected 31 ± 7 % of the cases. Coincidences with such unwanted, additional tone-delays were discarded from the event-related response analyses.

EEG recording and processing - Experiments 1 & 2

In Experiments 1 and 2 participants sat in comfortable chair in a noise-attenuated chamber. The EEG was recorded with a Synamp 2 amplifier (Compumedics Neuroscan, Victoria, Australia) from Ag/AgCl electrodes placed at the Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz. C4, T4, T5, P3, Pz, P4, T6, O1, O2 (10-20 system, Jasper, 1958) sites and the left and right mastoids (Lm, and Rm respectively). Because the auditory N1 elicited by pure tones often shows a polarity inversion between electrodes placed at the two sides of the Sylvian fissure when a nose-reference is used (Vaughan & Ritter, 1970), the reference electrode was placed on the tip of the nose. The horizontal electro-oculogram (EOG) was recorded by a bipolar electrode-setup placed near the outer canthi of the two eyes; the vertical EOG was recorded by electrodes placed above and below the right eye. Sampling rate was 1000 Hz, and on-line low-pass filtering of 200 Hz was used. The continuous recording was bandpass filtered off-line (0.1-20 Hz). Epochs of 600 ms duration including a 200 ms pre-event interval were extracted for various events described below in the event-related potential and field analyses section. Epochs with a signal range exceeding 100 μV on any channel were rejected from further processing.

MEG recording and processing - Experiment 3

In Experiment 3, the magnetoencephalogram was recorded with a 306-channel (204 orthogonal planar gradiometers and 102 magnetometers at 102 locations) Neuromag Vectorview MEG system (Elekta Oy, Helsinki, Finland) in supine position in an electromagnetically shielded room (Vacuumschmelze, Hanau, Germany). Horizontal and vertical EOG was recorded with a bipolar electrode setup from the outer canthi of the two eyes and from above and below the left eye respectively. Sampling rate was 1000 Hz, and an on-line low-pass filtering at 330 Hz was used. Five head position indicator coils (three positioned at forehead and two behind the ears) were used to continuously monitor head movements. The analysis of the head positions was based on 500 ms windows of data shifted by 250 ms intervals. Head-movements were always less than or equal to 4 mm for all participants, therefore, we did not apply head movement correction. The Signal Space Separation method (Taulu, Kajola, Simola, 2004) was used for external interference suppression, for the interpolation of bad channels, and to recompute the MEG data for an identical head position across all blocks. The continuous recording was offline bandpass filtered (0.8-16 Hz). Epochs of 350 ms duration, including a 100 ms pre-event interval were extracted, corresponding to events described below in the event-related potential and field analyses section. Epochs with a signal range exceeding 200 pT/m (gradiometer), 4 pT (magnetometer), or 80μV (HEOG and VEOG) were excluded from the analyses.

Event-related potential and field analyses

In all three experiments epochs corresponding to the following events were extracted from the EEG or MEG recordings: *coincidences* (i.e. tones immediately presented after a key-press)*; 250 ms post-key-press tones* (i.e. tones following key-presses by 250 ms)*, 500 ms post-key-presses tones* (i.e. tones following key-presses by 500 ms)*, 750 ms post-key-press tones* (i.e. tones following key-presses by 750 ms)*,* and *1000+ ms post-key-press tones* (i.e. tones following key-presses by at least 1000 ms). For all these events no other event occurred between the initial key-press and the last sampling point of the epoch.

In order to subtract the contribution of the action-(motor)-related activity from the tone-locked event-related responses (see below), a number of key-press-locked responses were also extracted: epochs corresponding to *key-presses* at least one second away from any other event, and epochs following such key-presses at integer multiples of 250 ms with no actual events in them (*250 ms, 500 ms, 750 ms, and 1000+ ms post-key-press epochs).* The zero time points of these post-key-press epochs were at 250, 500, 750, and at least 1000 ms following the key-press.

To estimate the auditory activity for coincidences, the event-related response elicited by key-presses was subtracted from that elicited by coincidences (in the following this is referred to as *corrected coincidence tone response*). To estimate the auditory activity in responses to 250, 500, 750, and 1000 + ms post-key-press tones, the corresponding post-key-press epochs were subtracted, respectively. These are termed *corrected 250 ms, 500 ms, 750 ms, and 1000+ ms post-key-press tone response,* respectively.

Two lines of analyses were conducted: In the first analysis, the corrected coincidence tone response was compared to the corrected 1000+ ms post-key-press tone response. Second, a trend analysis was calculated for the corrected coincidence, 250, 500, 750, and 1000+ ms post-key-press tone responses.

For the ERPs, individual N1 and P2 amplitudes were measured as the average signal in 20 ms long intervals centered at the group-average peak latency of the waveform. The amplitudes at the F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 leads were submitted to repeated-measures analyses of variance (ANOVA) with Sound, Laterality (3, z, 4) and Anterior-Posterior (AntPos; F, C, P) factors. Greenhouse-Geisser corrections were calculated when appropriate; in such cases uncorrected degrees of freedom, ε-values and corrected p-values are reported. Interactions involving the two-level Stimulus factor were explored further through pair-wise Student's t-tests. For the N1 amplitudes a further Sound × Laterality (left, right) repeated-measures ANOVA was calculated for the mastoid signals as well. To assess whether the N1- and P2 amplitudes were different functions of key-press-tone separations, a further Component (N1, P2) × Key-press-tone separation (0,250, 50, 750, 1000+ ms) ANOVA was conducted over the amplitudes measured at Cz. Because in this analysis the shapes of the functions were investigated, the amplitudes were z-transformed (producing distributions with a mean of 0 and a standard deviation of 1), separately for each component, each pooling the amplitudes for different key-press-tone separations. Whereas this transformation eliminated the component main effect, the key-press-tone separation effect was preserved for both components. In this analysis, the difference between the shapes of the amplitude functions is indicated by a significant Component × Key-press-tone separation interaction.

For the N1m ERFs, the measured variable was the source strength of single dipoles individually fitted in each hemisphere (as described below). Source strengths were analyzed through pair-wise Student's t-tests, and Sound × Hemisphere repeated-measures ANOVAs. Because the dipole fitting approach was not successful for the P2m, root mean squared (RMS) ERF amplitudes (calculated over all magnetometers in a 20 ms interval centered at the group-average N1m and P2m latencies) were also analyzed in pair-wise Student's t-tests, and trend analyses. The similarity of the N1m and P2m RMS amplitudes as functions of key-press-tone separations were assessed as in the analysis of the ERPs described above.

Results

Participants complied with the instructions in all three Experiments (Figure 1.2). Mean coincidence rates (i.e. the proportion of tones coinciding with a key-press [out of the 1008 tones delivered in total]; with standard deviations) were 5 ± 1 % in Experiment 1; 6 ± 1 % in Experiment 2; and 5 ± 1 % in Experiment 3.

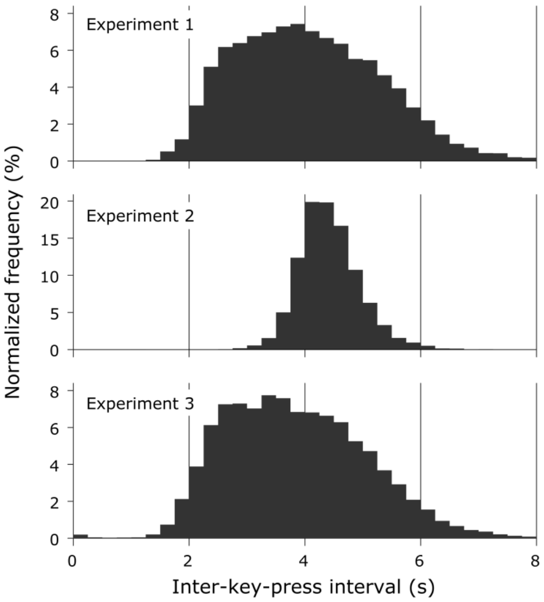

Figure 1.2. Histograms of all inter-key-press intervals from all participants in the three experiments.

Event-related potentials

In Experiment 1 and 2, tones elicited a clear succession of N1, and P2 waveforms (Figure 1.3).

Experiment 1

N1

The N1 peaked at 98 ms in the corrected coincidence ERP, and at 101 ms in the corrected 1000+ ms post-key-press tone ERP at Cz in the group-average waveform. The ANOVA of the N1 amplitudes measured in the 89-109 ms range showed a Stimulus main effect: F(1,13)=4.73, p<.05 indicating a lower (less negative) N1 amplitude for the coincidence; an AntPos main effect: F(2,26)=32.24, ε=0.80, p<.001; a Laterality main effect: F(2,26)=24.56, ε=0.88, p<.001; Stimulus × AntPos interaction: F(2,26)=8.42, ε=0.65, p<.01; and a Stimulus × Laterality interaction: F(2,26)=6.20, ε=0.97, p<.01. Student's *t* tests conducted between the post-key-press-minus-coincidence amplitudes at the F, C, and P leads (averaged over the Laterality levels) indicated that the N1-attenuation effect was stronger at central and parietal than at frontal sites (t[13]>2.97, p <.05). Student's *t* tests conducted between the post-key-press-minus-coincidence amplitudes on the left, middle, and right leads (averaged over the levels of the AntPos factor) indicated that the N1-attenuation effect was stronger at central than at lateral sites (t[13]>2.62, p <.05).

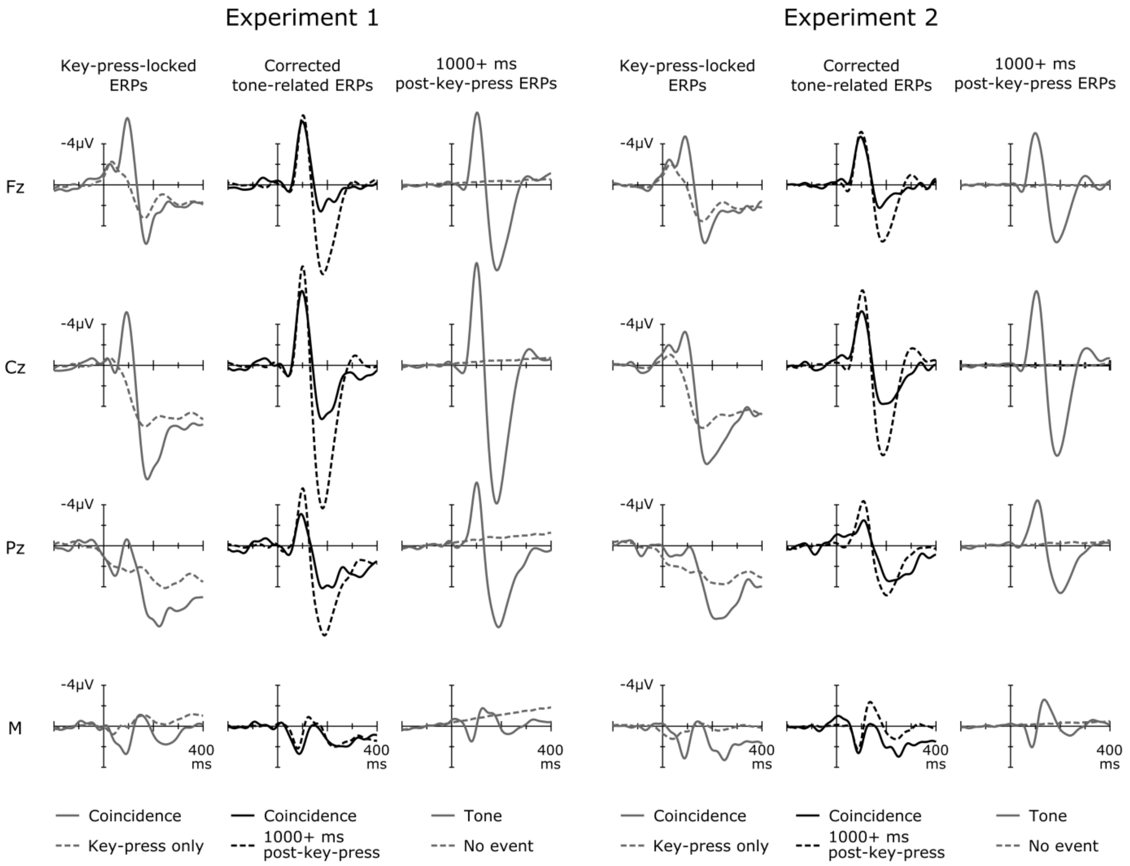

Figure 1.3. Group-mean event-related potentials elicited in Experiment 1 (left) and 2 (right) at the Fz, Cz, Pz, and the average of the mastoid leads. For each experiment, the key-press-locked ERPs are presented in the left column (for tone-key-press coincidences, and for key-presses separated from any other event by at least 1 s); in the right column ERPs to tones following key-presses by at least 1000 ms, and the average of epochs following key-presses by at least 1000 ms with no actual events in them. In the middle column, the corrected tone-related ERPs are shown: The *corrected coincidence tone ERP* is the coincidence-minus-key-press difference waveform of the ERPs presented in the left column, whereas the *corrected 1000+ ms post-key-press tone ERP* is the difference of the waveforms presented in the right column.

At the mastoids a significant Stimulus main effect was found: F(1,13)=8.74, p<.05, showing that the corrected coincidence ERP was *more positive* than the corrected 1000+ ms post-key-press tone ERP (i.e. for the coincidence the polarity-reversed N1 amplitude was higher).

The trend analysis (see Figure 1.6, top left panel) of the N1 amplitudes at Cz showed a significant linear trend: F(1,13)=14.89, p<.001, indicating an amplitude increase with growing key-press-tone separation.

P2

At the Cz lead, in the corrected coincidence waveform, the P2 peaked at 178 ms, whereas in the corrected 1000+ ms post-key-press tone ERP it peaked at 181 ms. The ANOVA of the P2 amplitudes measured in the 169-189 ms range showed a Stimulus main effect: F(1,13)=32.39, p<.001 indicating a lower (less positive) P2 amplitude for the coincidence; an AntPos main effect: F(2,26)=30.76, ε=0.70, p<.001; a Laterality main effect: F(2,26)=17.13, ε=0.73, p<.001; a Stimulus × AntPos interaction: F(2,26)=20.66, ε=0.93, p<.001; a Stimulus × Laterality interaction: F(2,26)=13.88, ε=0.98, p<.001; and an AntPos × Laterality interaction: F(4,52)=5.30, ε=0.66, p<.01. Student's *t* tests conducted between the post-key-press-minus-coincidence amplitudes at the F, C, and P leads (averaged over the Laterality levels) indicated that the P2-attenuation effect was the strongest at the central, and weakest at the parietal sites, with significant differences between each pair of sites (t[13]>2.55, p <.05). Student's *t* tests conducted between the post-key-press-minus-coincidence amplitudes on the left, middle, and right leads (averaged over the levels of the AntPos factor) indicated that the P2-attenuation effect was stronger at central than at lateral sites (t[13]>3.68, p <.01).

For the P2 amplitudes measured at Cz (see Figure 1.6, top right panel), significant linear and quadratic trends were found: F(1,13)=58.11, p<.001; and F(1,13)=15.53, p<.001, respectively. These indicate an amplitude increase with growing key-press-tone separation.

N1 and P2 amplitudes as the function of key-press-tone separation

The Component × Key-press-tone separation ANOVA of the z-transformed amplitudes showed a significant interaction: F(4,52)=3.61, ε=0.72, p<.05, indicating that the two component amplitudes behave differently as the function of key-press-tone separation,

Experiment 2

N1

The N1 peaked at 101 ms in the corrected coincidence ERP, and at 104 ms in the corrected 1000+ ms post-key-press tone ERP at Cz in the group-average waveform. The ANOVA of the N1 amplitudes measured in the 92-112 ms range showed a Stimulus main effect: F(1,12)=4.79, p<.05 indicating a lower (less negative) N1 amplitude for the coincidence; an AntPos main effect: F(2,24)=15.90, ε=0.69, p<.001; a Laterality main effect: F(2,24)=23.70, ε=1.00, p<.001; and a Stimulus × AntPos interaction: F(2,24)=6.13, ε=0.65, p<.05. Student's *t* tests conducted between the post-key-press-minus-coincidence amplitudes at the F, C, and P leads (averaged over the Laterality levels) indicated that the N1-attenuation effect was stronger at central and parietal than at frontal sites (t[12]>2.53, p<.05). No significant effects were found at the mastoids.

The trend analysis of the N1 amplitudes at Cz (see Figure 1.6, top left panel) showed a significant linear trend: F(1,12)=4.80, p<.05, indicating an amplitude increase with growing key-press-tone separation.

P2

At the Cz lead, in the corrected coincidence waveform, the P2 peaked at 186 ms, whereas in the corrected 1000+ ms post-key-press tone ERP it peaked at 187 ms. The ANOVA of the P2 amplitudes measured in the 177-197 ms range showed a Stimulus main effect: F(1,12)=11.71, p<.01 indicating a lower (less positive) P2 amplitude for the coincidence; an AntPos main effect: F(2,24)=14.62, ε=0.89, p<.001; a Laterality main effect: F(2,24)=5.96, ε=0.98, p<.01; a Stimulus × AntPos interaction: F(2,24)=25.01, ε=0.82, p<.001; an AntPos × Laterality interaction: F(4,48)=4.70, ε=0.55, p<.05; and a Stimulus × AntPos × Laterality interaction: F(4,48)=3.27, ε=0.69, p<.05.

To resolve the three-way interaction, Student's *t*-tests were conducted between the post-key-press-minus-coincidence amplitudes measured at the left, midline, and right electrodes at each levels of the AntPos factor. The attenuation-effect was maximal at the middle for all three AntPos levels. For the frontal leads, the attenuation-effect was stronger on the left and midline than on the right (t[12]>2.37, p <.05); at the central leads a significant difference in attenuation was only found between the right and midline electrodes (t[12]=2.66, p<.05). No significant attenuation-difference was found at the parietal sites. Student's *t* tests conducted between the post-key-press-minus-coincidence amplitudes at the F, C, and P leads (averaged over the Laterality levels) indicated that the P2-attenuation effect was the strongest at the central, and weakest at the parietal sites, with significant differences between each pair of sites (t[12]>2.96, p<.05).

For the P2 amplitudes measured at Cz (see Figure 1.6, top right panel) significant linear and quadratic trends were found: F(1,12)=17.99, p<.001; and F(1,12)=11.15, p<.01, respectively, indicating an amplitude increase with growing key-press-tone separation.

N1 and P2 amplitudes as the function of key-press-tone separation

The Component × Key-press-tone separation ANOVA of the z-transformed amplitudes showed no significant interaction.

Event-related fields – Experiment 3

Group-mean magnetic N1 and P2 (N1m and P2m) distributions as well as event-related magnetometer and RMS amplitude signals are presented in Figure 1.4. N1m was observed for all participants at 97 ± 4 ms after stimulus onset in the corrected 1000+ ms post-key-press tone ERF response. For eighteen participants, the ERFs were dipolar over both hemispheres - for these participants two-dipole models were fitted using a spherical volume conductor model and data from all sensors; for two participants, a dipolar event-related field was observable over the right hemisphere only. For these two participants a single dipole model was fitted using data from the right hemisphere sensors only. The goodness of fit values were in the 84 - 99% range (median: 96%, 1st quartile: 93%, 3rd quartile: 97%; mean: 94%, SD: 5%). Each dipole was located in one of the supra-temporal cortices (as assessed on each participant's magnetic resonance image; see Figure 1.5). Using these dipole positions and orientations, dipole magnitudes were fitted at the same latency to the corrected coincidence tone ERF response, and the corrected 250, 500, and 750 ms post-key-press tone ERF responses. Student's t-tests showed that source strength was smaller for the corrected coincidence tone response than for the corrected 1000+ ms post-key-press tone response on both sides (t[17]=3.54, p<.01 on the left, and t[19]=2.82, p<.05 on the right).

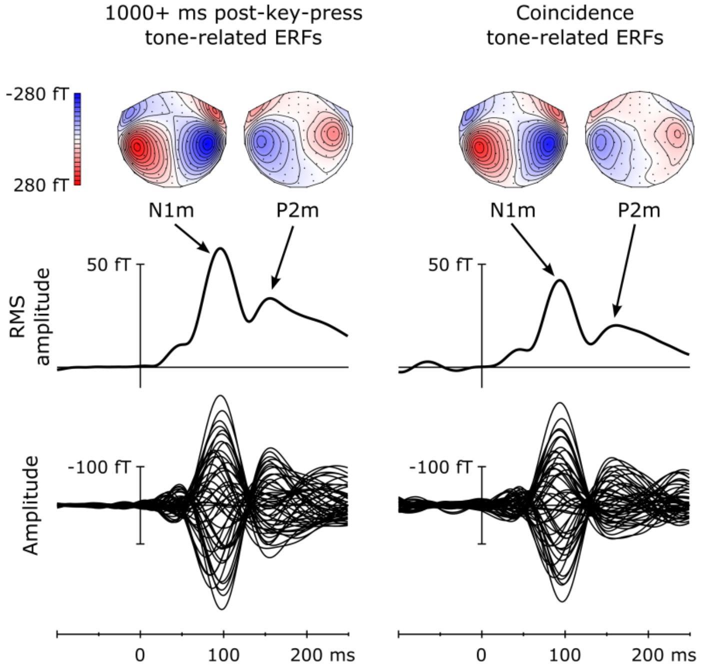

Figure 1.4. Group-mean event-related fields elicited in Experiment 3 for the *corrected 1000+ ms post-key-press tone* (left column), and for the *corrected coincidence tone* (right column). The bottom panel shows the amplitudes measured at the temporal magnetometers. The middle panel shows the root-mean-squared (RMS) amplitudes (calculated over all magnetometers). The average RMS amplitude in the 100 ms pre-stimulus interval (baseline) was subtracted from the RMS amplitudes. The top panel shows the ERF distributions in the N1m and P2m time-range.

The Stimulus (corrected coincidence vs. corrected 1000+ ms post-key-press) × Hemisphere (left, right) ANOVA of the source strengths for the participants with a dipolar ERF on both sides showed a main effect of stimulus type: F(1,17)=11.38, p<.01 indicating a smaller source strength for the corrected coincidence tone response; and a Stimulus × Hemisphere interaction: F(1,17)=7.48, p<.05, showing that the source strength differences between the two stimuli were larger on the left than on the right side.

The trend analyses of the source strengths (see Figure 1.6, bottom left panel) showed significant linear (F[1,17]=15.72, p<.001) and quadratic (F[1,17]=6.54, p<.05) trends in left, as well as in the right hemisphere (F[1,19]=12.48, p<.001 – linear and F[1,19]=4.85, p<.05 - quadratic). This indicates that N1m source strength increases with increasing key-press-tone separation.

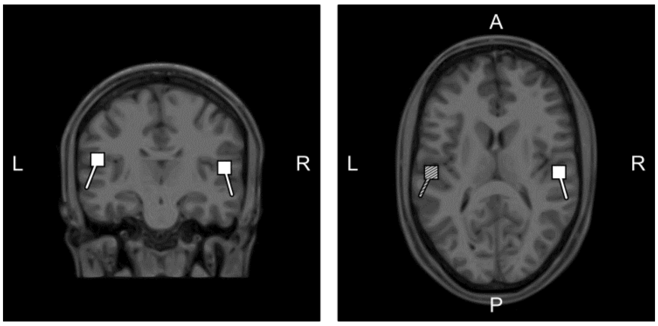

Figure 1.5. Equivalent current dipole positions (squares) and orientiations (bars) fitted at the N1m, superimposed on the T1-weighted magnetic resonance image (coronal- left, axial – right) of one representative participant. The axial slice is positioned at the height of the right dipole. The shaded square shows the vertical projection of the left dipole, which was about 6 mm above the slice. Slice orientations are indicated by letters: A-anterior, P-posterior, L-left, R-right.

The group average RMS amplitude peaked at 96 ms (N1m) and 156 ms (P2m) in the corrected 1000+ ms post-key-press tone ERF response. The RMS amplitude in the N1m and P2m time ranges (Figure 1.4, middle) were significantly lower in the corrected coincidence tone ERF, than in the corrected 1000+ ms post-key-press tone ERF response (t[19] = 7.82, p<.001 for N1m, and t[19] = 6.60, p<.001 for P2m). The trend analyses of the RMS amplitudes (Figure 1.6, middle panels) showed significant linear (F[1,19]=61.61, p<.001) and quadratic (F[1,19]=7.13, p<.01) trends for N1m. For the P2m, significant linear (F[1,19]=31.24, p<.001), quadratic (F[1,19]=9.91, p<.01) and cubic (F[1,19]=4.84, p<.05) trends were found. This indicates that RMS amplitude in the N1m and P2m time ranges increased with increasing key-press-tone separation.

The Component × Key-press-tone separation ANOVA of the z-transformed RMS amplitudes showed no significant interaction.

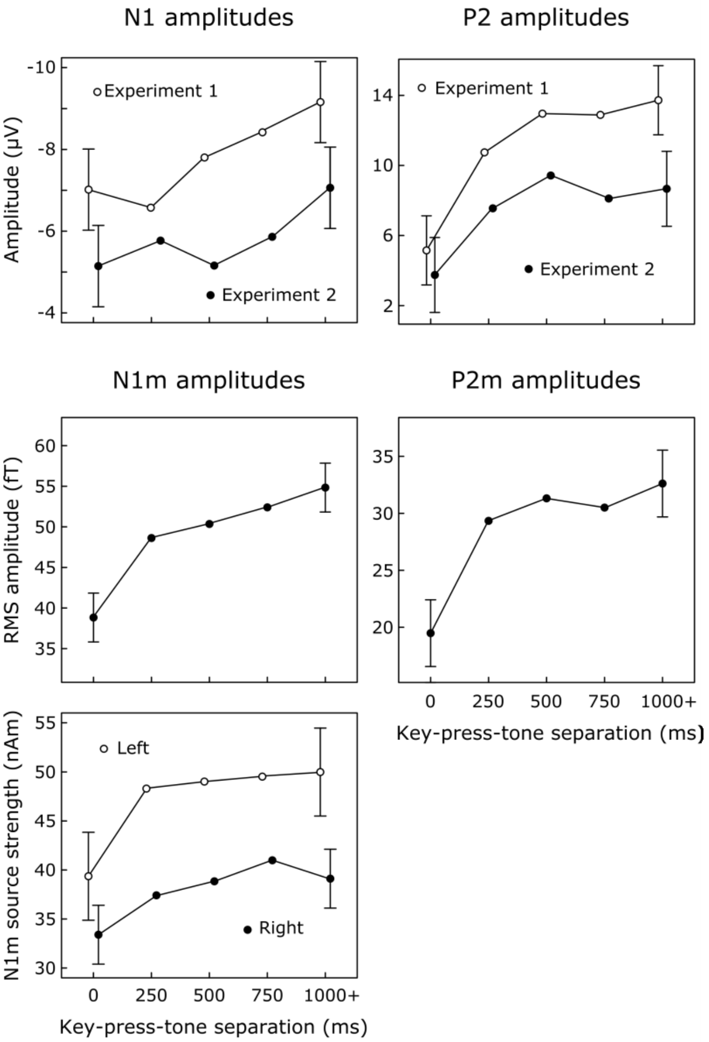

Figure 1.6. Group-mean corrected N1 and P2 amplitudes elicited at the Cz electrode in Experiments 1 and 2 (top, left and right panels); N1m and P2m RMS amplitudes (middle left and middle right panel, respectively), and N1m source strengths (bottom, left panel) in Experiment 3 for tones following key-presses by 0 (coincidences), 250, 500, 750 and at least 1000 ms. The whiskers represent 95% confidence intervals for the pair-wise comparison of coincidence- and 1000+ ms post-key-press amplitudes (calculated according to Masson & Loftus, 2003).

Discussion

The results of the three experiments show that the N1(m) and P2(m) auditory event-related responses are consistently attenuated when the eliciting sound coincides with an action, even when no contingent action-stimulus relationship exists. In contrast with previous research, the present results are not contaminated by between-block differences or condition order because the relevant key-press-, sound- and coincidence events occurred within the same experimental blocks.

Whereas the N1m-attenuation observed in Experiment 3 shows that auditory processing was attenuated for tones coinciding with key-presses, the topography of the ERP attenuation effect in Experiment 1 and 2 was more posterior than expected (i.e.

larger at the parietal than at the frontal leads). Moreover, key-press-tone coincidences even increased the positive aspect of the N1 at the mastoids in Experiment 1. This indicates that beside the attenuation of auditory sensory processing, further processing changes take place when an action-sound coincidence occurs. This may be the attenuation of the widely distributed non-specific N1 component, which, in turn, would make the polarity reversal of the stimulus-specific N1 more obvious, but it is also possible that coincidences result in a further, parietally positive ERP component overlapping the N1.

Whereas P2(m) was attenuated in all three experiments, the two methods showed somewhat different aspects of the P2. P2m ERF peaked earlier, and its sensitivity to the key-press-tone separation was not different from that of the N1m ERF. The P2 ERP amplitude on the other hand, showed a different dependence on key-press-tone separation than the N1 (in Experiment 1 where amplitudes were higher due to their relative loudness). This suggests that the N1 and P2 ERP waveforms include subcomponents which reflect functionally different aspects of processing (see also Ford, et al., 2001, and Knolle et al., 2012)

Taken together with the results of Hazemann et al. (1975) and Makeig et al. (1996) the present study provides strong evidence that performing an action leads to an attenuation of concurrent auditory processing, that is, action-sound *contiguity* is a *sufficient* condition for N1-attenuation. The attenuation effects in the present study are very similar to those routinely found in paradigms in which actions and sounds have a contingent relationship. Whereas previous studies assumed that *contingency* was a *necessary* condition for N1-attenuation, the present design provides a baseline condition for the assessment of potential contingency-related N1-attenuation effects; and the results show the necessity to directly test whether action-stimulus contingency and its forward modeling contributes to N1-attenuation or not. It is important to emphasize, that based on the present results, it cannot be determined, whether action-sound contiguity is the sole cause of auditory attenuation or not. It seems possible that when a contingent action-stimulus relationship exists, this might be represented by a forward model, which could bring further attenuation about. In comparison to the present experiments, in which actions had no auditory consequences, presenting action-contingent stimulation is likely to result in an explicit expectancy of the given contingent sensory event, which might contribute to N1-attenuation as well. Indeed, some studies show attenuation effects which would be difficult to explain solely on the basis of action-stimulus contiguity. For example, Baess et al. (2011) found that N1-attenuation is stronger when action-independent sounds are mixed into a self-induced (contingent) action-sound sequence.

Discussion - novel hypotheses

To accommodate the N1 and P2 attenuation effects found in Study I, three interpretations - post-hoc hypotheses - were formulated:

1) The first interpretation suggests that N1-suppression (in the experiments of Study I, as well as other) does not reflect the workings of a forward model, rather that it results from a *dynamic change in the distribution of attentional resources*. As described in the Introduction of the present thesis, auditory N1 is enhanced when sounds are in the focus of attention (e.g. Hillyard et al., 1973) and it is attenuated when attentional focusing is disrupted (Horváth & Winkler, 2010). It seems possible that pressing a button or performing an action draws attention away from task-irrelevant auditory stimulation for a short period of time, which results in attenuated N1 for tones close to key-presses. This explanation is on a par with that offered by Makeig et al. (1996) for the auditory steady-state response.

2) The second interpretation suggests that N1-suppression does not *result from* the cancellation of sensory reafference, rather that it reflects a process *subserving* the formation of a forward model or other contingency-representation. Since contiguity is one of the cues that may allow the inference of a causal relationship between events (Hume, 1739/1896), the detection of action-contiguous sound events may be necessary for the formation of an action-sound contingency-representation. N1-suppression might simply reflect a process which “flags” such sounds, and thereby provides a signal which could serve as a basis for the formation of an action-sound contingency-representation.

3) A third type of interpretation suggests that despite the absence of a contingent action-stimulus relationship, the suppression effect nonetheless reflects the cancellation of sensory reafference, that is, it still results from *the workings of an internal forward model*. Whereas forward modeling allowed *more efficient* interactions with the environment in most previous studies, this interpretation suggests that N1-suppression reflects the workings of a *dysfunctional* contingency-representation in the experiments of Study I, (after all, such a model would produce invalid predictions about the occurrence of self-initiated sounds). In the following, this line of thought is elaborated, which delineates some key questions regarding this interpretational framework.

Assuming that there is an internal forward model representing a key-press-sound contingency in the experiments of Study I, there are two questions which should be answered: 1) how is this representation created, and 2) how (and why) does such a representation get preserved?

First, it seems possible that the hypothetical contingency representation is not built up at the beginning of, or over the course of the experiment (see also Lange, 2011), but it *already exists*: it is a general “expectation” that our actions should generate some kind of a sensory event in the environment. Whereas it is an intriguing possibility that N1-suppression reflects an innate readiness for capturing contingent action-effect relationships during interactions with the environment, in the present study such a “readiness” could also be brought about by extensive training through the wide-spread use of key-press-based interfaces in everyday devices. Long-term training creates strong associations between actions and their sensory consequences, which influence perception even if the actions do not take place in their usual context: Repp & Knoblich (2007) showed that when pianists performed movements which would generate a rising or descending tone-pair on a piano, this induced a corresponding bias in the perception of an ambiguous pitch-change, whereas for non-pianists no bias was observable. In the present case, the long-term use of key-press-based interfaces might give rise to a general action-effect association in which the effect can be a large class of sensory events (including a tone). The generality of the effect would also explain how it is possible that an arbitrary contingency between a key-press and an artificial tone is represented by an internal forward model similarly to that hypothesized to exist between speech production and speech sounds.

A second possibility is that a contingency representation is built up (rapidly) during the experiment, but the build-up of the representation does not depend on *actual* action-stimulus contingency, rather, it is based on the instances when the action and stimulus events were temporally contiguous. That is, coincidences give rise to an "*illusory contingency*" in the present case despite the absence of an actual contingent relationship. It seems even possible that the cognitive system might not only take "real" coincidences as evidence for the establishment of such a representation, but also stimulus events which follow actions within a *sensitive period*. Elsner & Hommel (2004) found that contiguity played a role in forming action-effect associations even if (contingent) task-irrelevant sounds followed actions by 1 s. If stimuli following actions within 1 s were interpreted as evidence for a contingent action-effect relationship, then in the present experiments the relevant coincidence rate would be around 25% (whereas the "real" coincidence rate was 5%). Note nonetheless that the present results show a decrease of the suppression effect with growing key-press-tone separation over 1 s, which suggests that the duration of such a sensitive period may be much shorter (on the order of couple hundred ms).

The hypotheses outlined above suggest ways in which a contingency representation might be created despite the absence of *actual* contingency. These hypotheses however do not reflect on why or how such representations are *preserved* in these situations. Adaptation to changed contingencies in the sensory-motor system is usually investigated in paradigms in which an established action-stimulus contingency is abruptly changed to a different contingency. Adaptation to the new contingency takes place on multiple time scales, and is based on learning from prediction errors (for a summary see Shadmehr et al., 2010). On this basis, two speculations on the preservation of the hypothetical key-press-sound contingency representation can be put forward: First, similarly to the acquisition of a contingency representation, the characteristic time for changing such a representation may be simply too long compared to the typical duration of an experiment. That is, if it takes long-term training to build-up such representations, then changing them might take long time as well. Second, it is possible that representations are changed when the action starts to lead to *novel* consequences, however, when the associated effect is simply absent, the representation may remain unchanged, because a prediction error cannot be calculated when no effects are present. To put it differently, changes may be induced by the *interference* between the associated and actual consequences, whereas the *lack of a consequence* might not produce such interference, thereby not affecting the model itself.

Study II. Attention-related explanations of the coincidence effect[3]

Summary

The role of auditory selective attention in the action-sound coincidence-related auditory ERP attenuation was investigated in two experiments by presenting tones separated by random intervals. In the Homogeneous Condition of Experiment 1 and in Experiment 2 the same tone was repeated, whereas in the Mixed Condition of Experiment 1 tones with five different frequencies were presented. Participants performed a time interval production task by marking intervals with key-presses in Experiment 1, and tried to produce key-press-tone coincidences in Experiment 2. Although the auditory ERPs were attenuated for coincidences, no modulation by the multiplicity of tone-frequencies in Experiment 1, or by the task-relevancy of tones and coincidences in Experiment 2 was found. This suggests that coincidence-related ERP attenuation cannot be fully explained by voluntary attentional mechanisms.

Introduction

The goal of Study II was to investigate the mechanisms behind the coincidence effect found in Study I. Specifically, the hypotheses that the coincidence effect was caused by a dynamic change in the distribution of attentional resources was investigated. Two hypotheses based on well-known voluntary attentional effects on auditory processing were tested in variations of the coincidence paradigm.

Because the auditory N1 is enhanced when the eliciting sounds are attended (Hillyard, et al. 1973, Schwent, Hillyard, Galambos, 1976, Hansen, Hillyard, 1980, Rif, et al., 1991; Okamoto et al., 2007), and attenuated when a sound-focused attention set is disrupted (Horváth & Winkler, 2010; Horváth 2014a and 2014b), it was hypothesized that the coincidence effect may be caused by a dynamic change in the distribution of attentional resources. That is, the N1 and P2 difference between tones coinciding with an action and tones separated by longer intervals from preceding actions may be caused by a difference in the amount of attention allocated to these tones. In the experiments demonstrating the coincidence effect, tones were presented relatively infrequently (with 1.5-6 s inter-stimulus interval) and they were task-irrelevant. Participants made key-presses in a time-interval reproduction task, every 4 seconds on average. Due to the scarcity of stimulation, attending the tones may allow one to maintain vigilance in this situation, and performing the action may lead to a short-term redirection of attention from the tones to the key-press. This change in the distribution of attention is made visible by the impacted processing of tones presented in close temporal proximity to key-presses, which is manifested in attenuated N1 in comparison to tones separated by longer time intervals from the actions. That is, pressing the key “removes” the

---

[3] Based on Horváth, J. (2013). Attenuation of auditory ERPs to action-sound coincidences is not explained by voluntary allocation of attention. *Psychophysiology, 50*(3), 266–273. doi:10.1111/psyp.12009

attentional N1 enhancement from tones in close temporal proximity to the key-press, whereas this enhancement is unaffected by key-presses occurring temporally far away.

Evidence for the attentional enhancement of auditory N1 comes from paradigms in which participants directed attention towards the sounds because of their task-relevancy, that is, attending the sounds was voluntary. In such paradigms, it was found that this type of attentional effort influenced auditory processing and evoked auditory responses in two ways (Okamoto et al., 2007; Kauramäki et al., 2007): by a "gain" and a "sharpening" mechanism. The "gain" mechanism generally improves processing when attention is directed towards an auditory channel, whereas the "sharpening" mechanism (demonstrated for frequency) allows one to selectively enhance the processing of sounds with a given feature-level while suppressing others. In the present context, the hypothetical key-press-related attenuation of the auditory ERPs may reflect the workings of the "gain", as well as the "sharpening" functions: One may generally increase auditory "gain" to maintain vigilance in a situation which is impoverished in sensory input. Because in the studies of Hazemann, et al. (1975) and Horváth et al. (2012) the same tone was presented over and over again, one may also assume that through the tone repetition a selective attention set to the given tone (and tone frequency) might have been established ("sharpening").

To test the hypothetical involvement of the "sharpening" mechanism, tones with five different frequencies were presented in randomly mixed and homogenous sequences in a coincidence paradigm (Experiment 1). If attentional "sharpening" allowed by the constant repetition of a single frequency played a role in the coincidence-related attenuation of auditory ERPs, then attenuation should be absent or less in the mixed than in the homogeneous sequences.

In Experiment 2, the tones were made task-relevant by instructing participants to press the key so that key-presses coincided with tones as often as possible (a "coincidence game"), motivating them by a small bonus payment for each "hit". In this arrangement, attention should not be diverted from the tones by key-presses, since tones occurring at the time of the key-press have high task- (and motivational) relevance. If attention (either through a "gain" or a "sharpening" mechanism) played a role in coincidence-related ERP attenuation, then its magnitude should be smaller when the "coincidence game" was played in contrast to when a time-interval reproduction task was performed in which tones were task-irrelevant.

Methods

Participants

Experiment 1.

15 young adult volunteers (20-26 years, 23 years average age; five women, one left-handed) participated in Experiment 1 for monetary compensation. To motivate participants to attend both sessions, a bonus was paid for completing both sessions. Data

from an additional participant was discarded because of the completion of only the first session of the experiment. All of them reported normal hearing status. Before the experimental sessions, participants gave written informed consent after the experimental procedures were explained to them.

Experiment 2.

14 volunteers from the group in Experiment 1 participated in Experiment 2 for monetary compensation. Data from one participant was discarded because of extensive eye-movement artifacts in the EEG. Before the experiment, participants gave written informed consent after the experimental procedures were explained to them.

Stimuli and procedures

Experiment 1.

Experiment 1 comprised two sessions, with 15 experimental blocks presented in each session. Blocks were separated by short breaks as needed, with a longer (~10 min) break around the middle of each session. The two sessions were administered within 5 weeks. During the experimental sessions, participants were sitting in a comfortable armchair in a well-lit, sound-proofed room. To minimize eye-movements, participants were instructed to rest their gaze at a fixation puppet placed in front of them during the experimental blocks.

Participants performed a time interval production task: they were required to press a rod-mounted key held in their dominant hand to mark time intervals, so that the histogram of the between-key-press-intervals would show a uniform distribution in the 2-6 s range, with no intervals outside this range. The experimental sessions started with a training phase in which participants learned the task with on-line visual feedback: the histogram of the produced intervals was updated on a screen in front of them after each key-press. During the experiment, this feedback was provided only at the end of each experimental block.

During task-performance, a sequence of tones was presented through headphones (HD-600, Sennheiser, Wedemark, Germany). The duration of the tones was 50 ms, including 10 ms rise and 10 ms linear fall times. Tone intensity was individually adjusted to 50 dB sensation level (above hearing threshold level). In each experimental block 75 tones were presented with onset-to-onset inter-stimulus intervals randomly sampled from a 2-6 s uniform distribution (the average block duration was 5 min). There were five tone frequencies: 988, 1109, 1245, 1397 and 1568 Hz. The choice of frequencies was based on the fact that Okamoto et al. (2007) found that the attentional “sharpening” effect on N1m significantly increased when the separation of frequencies interfering with a 1000 Hz tone decreased from ± 80 Hz to ± 40 Hz and lower separations. Also, the equivalent rectangular bandwidths of auditory filters is between 131 and 198 Hz for the given frequency range (Glasberg & Moore, 1990), so this range of frequencies should provide sufficient frequency-separation to make the hypothesized attentional “sharpening” effect visible. In the Mixed condition each tone was delivered 15 times in random order in each block. In the Homogenous condition, only one of the frequencies was presented in each block (75 times). There were 15 Mixed and 15 Homogenous (3 with each of the five frequencies) blocks presented in the two sessions. The blocks were presented in an interwoven order: “MHHMMHHM…” or “HMMHHMMH…” where “M”s denote Mixed, and “H”s denote Homogeneous blocks. The order of the frequencies between Homogenous blocks was random with the constraint that no frequency-repetitions between consecutive blocks were allowed.

The pre-generated, random tone presentation schedule was manipulated on-line during the experiment to produce key-press-tone coincidences (see Horváth et al., 2012). When a key-press occurred, the tone schedule was revised: tone presentation times following the key-press were shifted uniformly towards the moment of key-press so that the next tone to be delivered was delivered right away (if it was scheduled within 250 ms of the key-press) or after the integer multiple of 250 ms preceding its originally scheduled presentation time (e.g. if a tone was scheduled to be delivered 983 ms after the key-press, it was re-scheduled to 750 ms following the key-press). This adjustment was made only for the last key-press preceding a tone.

Due to the constraints of the stimulation equipment, for coincidences, there was 8 ± 2 ms (mean ± standard error) delay between key-presses and tone-presentations; also due to a programming error, for a number of coincidences (36 ± 6% of the coincidence trials in Experiment 1, and 39 ± 5% in Experiment 2) this delay was 18 ± 2 ms. These delays were taken into account in ERP processing (see below).

Experiment 2.

Experiment 2 was conducted within 4 months after Experiment 1 was completed. In Experiment 2, the stimulation arrangement was the same as in the Homogenous Condition of Experiment 1. The only difference was the task: In Experiment 2, participants performed a “coincidence game”: they were instructed to press the key so that key-presses would coincide with (“hit”) a tone as often as possible. For each coincidence, a small bonus was paid (which resulted in an accumulated premium of about +50-80% of the participation fee). To make the key-pressing activity similar to that in Experiment 1, we also introduced the following constraints in the instructions: a maximum of 75 key-presses could be made in a block, and key-presses must not have occurred closer than 2 s, or with a delay exceeding 10 s; also failure to comply with these constraints would result in the loss of the bonuses accumulated in the previous and the current block. Whereas there are some strategies which could help one to achieve a higher-than-random coincidence-rate (for example, not pressing the key if a tone was presented in the last two seconds), due the random nature of tone presentation, this is a game of chance.

EEG-recording and analysis

Experiment 1.

The EEG was recorded by a Synamp2 amplifier (Compumedics Neuroscan, Victoria, Australia), with 1000 Hz sampling rate and on-line low-pass filtering of 200 Hz, from Ag/AgCl electrodes placed at the Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz. C4, T4, T5, P3, Pz, P4, T6, O1, O2 (10-20 system, Jasper, 1958) sites and the left and right mastoids (Lm, and Rm respectively). The reference electrode was placed at the tip of the nose. Horizontal EOG was recorded between the outer canthi of the two eyes by a bipolar setup, similarly to the vertical EOG, which was recorded from electrodes placed above and below the right eye.

The EEG was off-line low-pass filtered (30 Hz). Epochs corresponding to coincidences, and tones following key-presses by at least 1000 ms (1000 + ms post-key-press tones) were extracted. Because tones following key-presses were shifted towards the key-press in time, the inter-tone interval is shorter for these tones than for those preceded immediately by another tone (and not a key-press). Because inter-tone interval affects auditory ERPs (see Näätänen & Picton, 1987), only tone-events following key-presses were included in the analyses. To estimate the auditory activity for coincidences and the 1000+ ms post-key-press tones, the temporally corresponding key-press-related ERP was subtracted from them: For the coincidence ERP, epochs corresponding to key-presses without any other events within ±1 sec were extracted. For the 1000+ ms post-key-press tone ERP, epochs following key-presses by at least by 1000 ms with integer multiples of 250 ms with no actual events in them were extracted (1000+ ms post-key-press epochs). The variability in the key-press-tone delay for coincidences (as described above) was taken into account by shifting the key-press-related ERPs used for correction by 8 ms or 18 ms with random epoch selection in proportion to the occurrence of the two delays. That is, for example, if the key-press-tone delay for coincidences was 18 ms in 30% of the coincidences (and 8 ms for the rest), then 30% of the key-press-related epochs were shifted by 18 ms, the rest by 8 ms before averaging.

The duration of the extracted epochs was 700 ms, including a 200 ms pre-event interval. Amplitude calculations were referred to this pre-event interval. Epochs within the first 10 s of each block, as well as those with signal range exceeding 100 µV on any channel were discarded from the analyses.

Individual N1 and P2 amplitudes were measured as average signals in 20 ms long intervals centered at the group-average peak latencies of the action-corrected tone-related waveforms. The amplitudes at the Fz, Cz, Pz and Oz leads were submitted to repeated-measures analyses of variance (ANOVA) with Condition (Homogeneous vs. Mixed), Tone (corrected coincidence vs. 1000+ ms post-key-press tone) and Electrode (Fz, Cz, Pz, Oz) factors (Oz was calculated as the average signal of the O1 and O2 leads). For N1, a separate ANOVA for the average of the mastoid signals was also conducted with Condition and Tone factors. Partial eta-squared measures are reported. Greenhouse-Geisser corrections were calculated when appropriate; in such cases uncorrected degrees of freedom, ε-values and corrected p-values are reported. Interactions involving the two-level Tone factor were explored further through pair-wise Student's t-tests. All significant effects are reported.

Experiment 2.

In Experiment 2, EEG-recording and processing was identical to that in Experiment 1. N1 amplitude was analyzed in a repeated-measures ANOVA with Tone (corrected coincidence vs. corrected 1000 + ms post-key-press tone) and Electrode (Fz, Cz, Pz, Oz) factors. To compare N1-attenuation magnitudes between experiments, N1-amplitudes (elicited in the Homogeneous conditions) were normalized by dividing the individual amplitudes by the group mean N1-amplitude measured for the corrected 1000 + ms post-key-press tones in the same experiment. The normalized N1 amplitudes were submitted to a repeated-measures ANOVA with Experiment (1 or 2), Tone, and Electrode factors. In this analysis significant effects involving the Experiment and Tone factors were of interest. A further repeated-measures Experiment × Tone × Electrode ANOVA was calculated for the amplitudes measured in the time ranges of the P3b waveforms observable in Experiment 2.

Results

Behavioral measure

Participants complied with the instructions in both experiments (see Figure 2.1). The proportion of tones coinciding with key-presses was 5.0 ± 0.8 % in Experiment 1, and 5.5 ± 0.8 % in Experiment 2. This difference did not reach significance (paired t-test calculated for those participating in both experiments, t[12] = 2.09, p =.06).

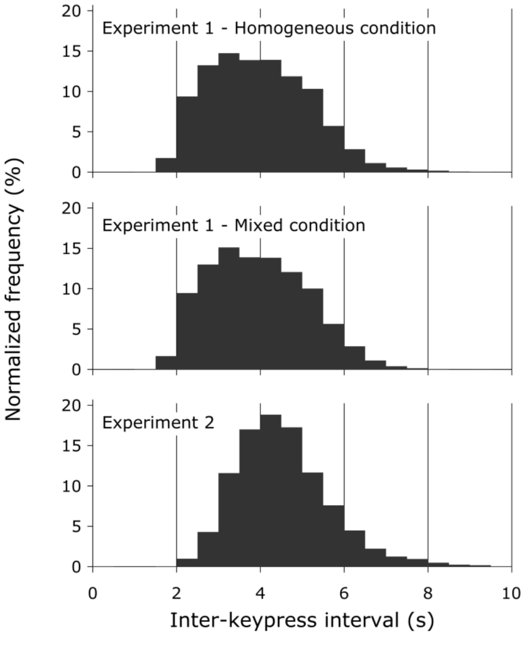

Figure 2.1. Histograms of all inter-key-press intervals from all participants in the Homogeneous and Mixed Conditions of Experiment1, and in Experiment 2.

ERPs

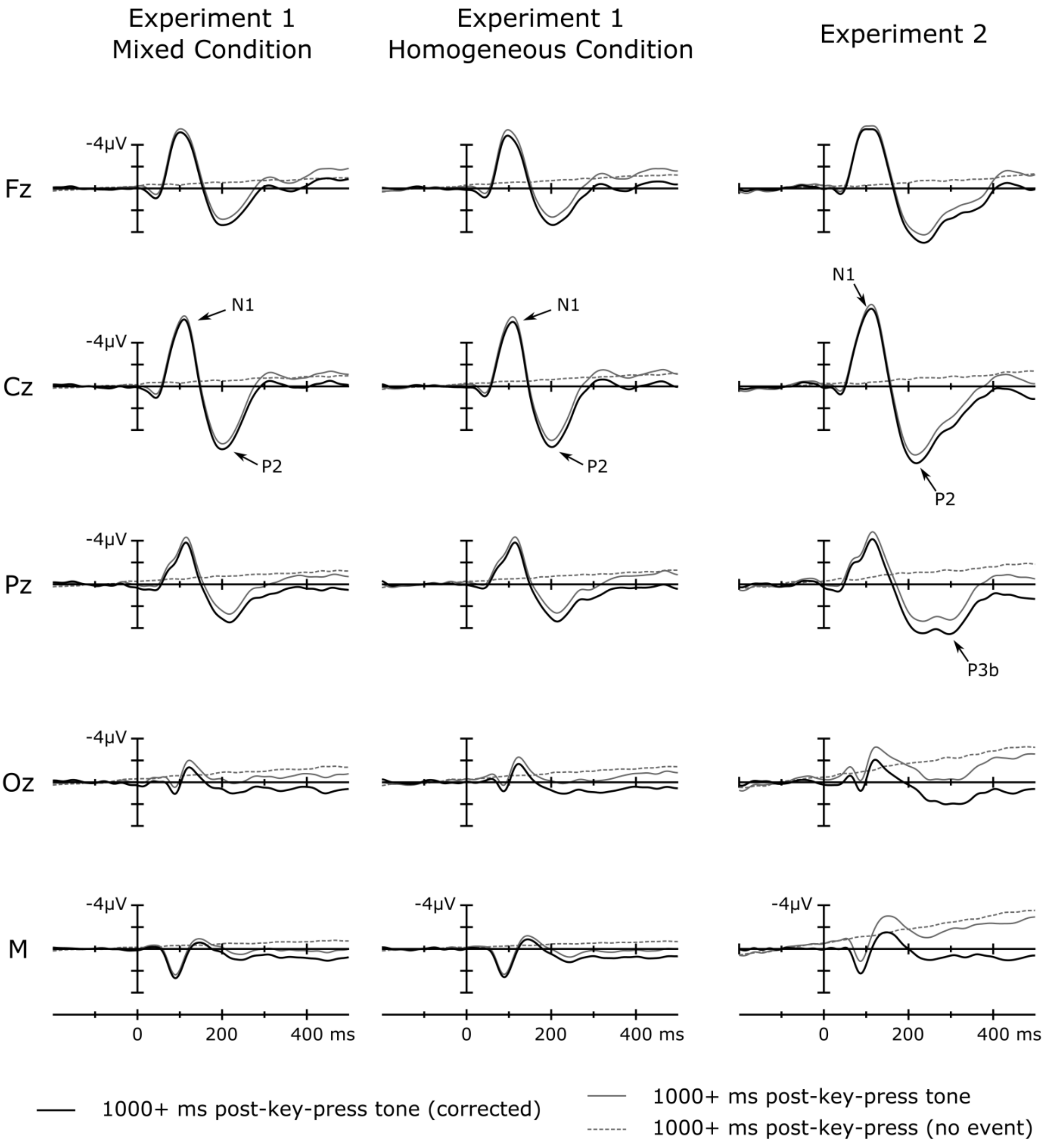

Figure 2.2. Group-average event related potentials elicited by tones separated by at least 1000 ms from preceding key-presses, and the corresponding average epochs with no actual events, and their differences (corrected 1000+ ms post-key-press tones), measured in the two experiments at the Fz, Cz, Pz, the average of the O1 and O2 (Oz), and the average of the mastoid (M) signals. Tone onset is at the crossing of the axes.

The group average ERPs recorded in the two experiments are presented in Figures 2.2, 2.3 and 2.4. Figure 2.2 shows that in both experiments a clear linear trend was present after more than 1000 ms following a key-press. The corrected 1000+ ms post-key-press tone waveform was obtained by subtracting this trend from the 1000+ ms post-key-press tone ERP. Figure 2.3 shows the derivation of the corrected coincidence waveform as the coincidence-minus-key-press-related waveform, and Figure 2.4 shows the corrected waveforms.

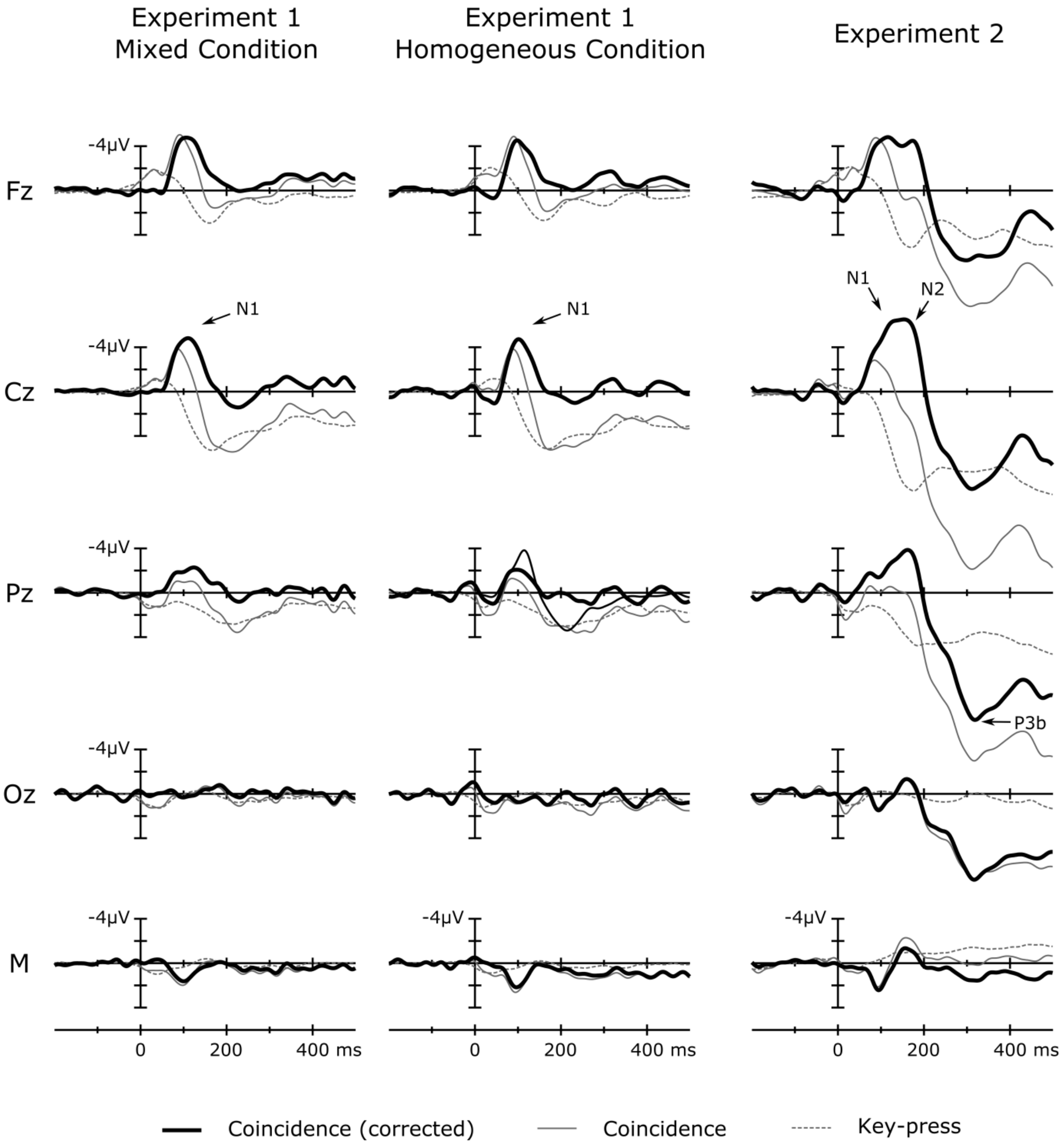

Figure 2.3. Group-average event related potentials elicited by tone-key-press coincidences and the corresponding key-presses, and their differences (corrected coincidence), measured in the two experiments, at the Fz, Cz, Pz, the average of the O1 and O2 (Oz), and the average of the mastoid (M) signals. Tone onset is at the crossing of the axes.

ERPs – Experiment 1.

In Experiment 1, the corrected 1000+ ms post-key-press tone-related N1 waveform peaked at 110 ms in the Mixed, and at 109 ms in the Homogeneous condition; the P2 peaked at 200 and 202 ms, respectively (Figure 2.4, left and center columns). The corrected coincidence waveform showed the same ERP pattern. The ANOVA of the N1 amplitudes at the midline showed a Tone main effect: F(1,14) = 6.06, $\eta^2_p = 0.30$, p<.05, indicating lower (less negative) amplitude for coincidences; and an Electrode main effect: F(3,42) = 85.99, ε = 0.49, $\eta^2_p = 0.86$, p<.001. The ANOVA of the N1 amplitudes at the mastoids showed no significant effects. Although no significant Tone × Electrode interaction was found, the topographical distributions (Figure 2.5, top and middle rows) show that the coincidence effect was numerically more posterior than the tone-related N1.

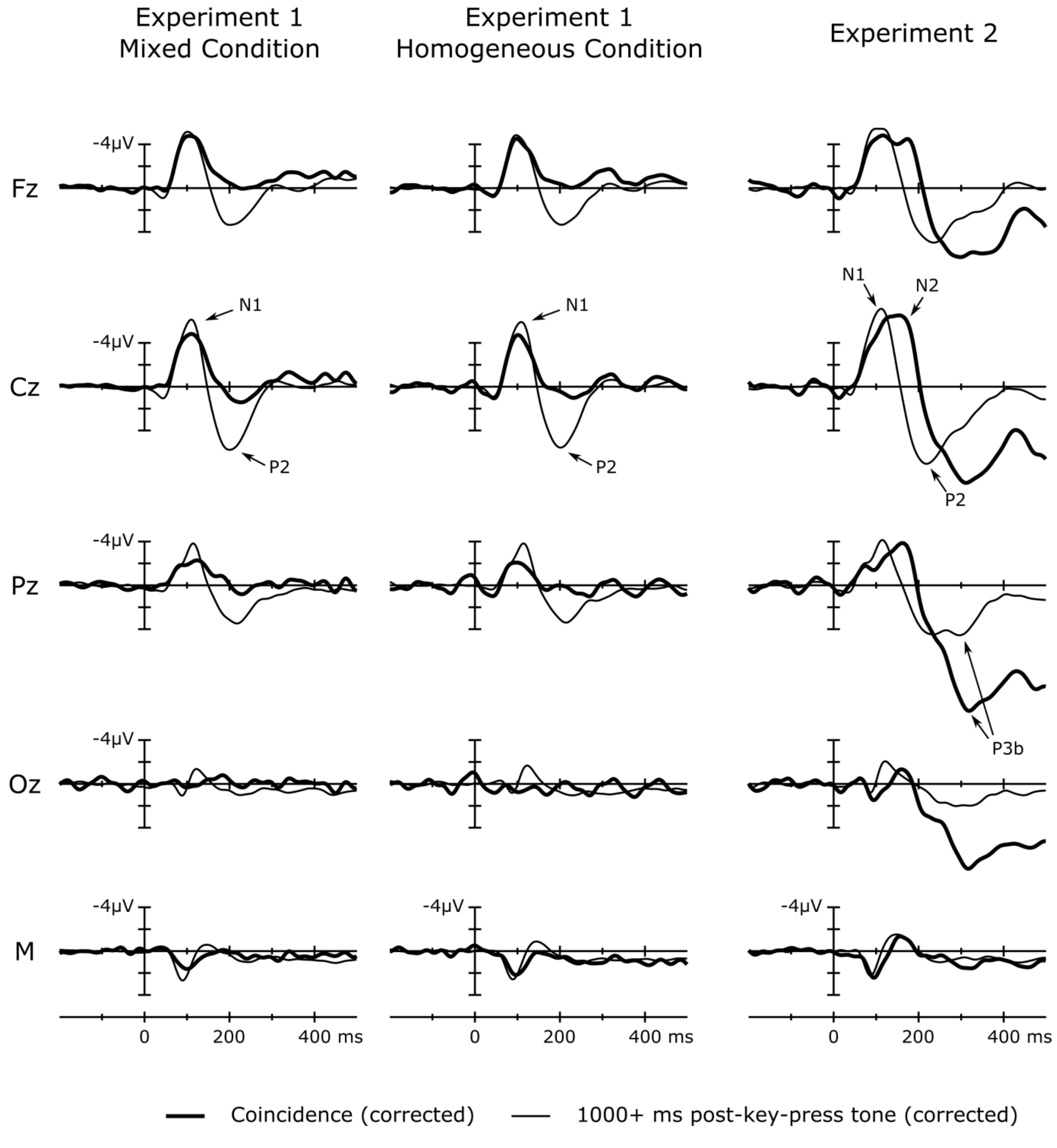

Figure 2.4. Group-average corrected coincidence and corrected 1000+ ms post-key-press tone ERP waveforms measured in the two experiments, at the Fz, Cz, Pz, the average of the O1 and O2 (Oz), and the average of the mastoid (M) signals. Tone onset is at the crossing of the axes.

The ANOVA of the P2 amplitudes showed a Tone main effect: F(1,14) = 35.99, $\eta^2_p = 0.72$, p<.001, indicating lower (less positive) amplitudes for coincidences; an Electrode main effect: F(3,42) = 12.32, $\varepsilon = 0.46$, $\eta^2_p = 0.47$, p<.01; and a Tone × Electrode interaction: F(3,42) = 31.21, $\varepsilon = 0.50$, $\eta^2_p = 0.69$, p<.001. Pair-wise t-tests showed that the 1000+ ms post-key-press tone-minus-coincidence difference was larger at Cz, and smaller at Oz than at Fz or Pz (t-scores > 3.91, p-scores < .01).

ERPs – Experiment 2.

In Experiment 2, the corrected 1000+ ms post-key-press tone-related N1 peaked at 112 ms and was followed by a P2 at 218 ms (Figure 2.2, right column). In the corrected coincidence-waveform the N1 was followed by a partially overlapping, centrally peaking negativity (N2) at around 190 ms. Because P2 was overlapped by the N2, only N1 amplitudes were analyzed. The ANOVA of the N1 amplitudes at the midline showed a Tone main effect: F(1,12) = 5.11, $\eta^2_p = 0.30$, p<.05, indicating lower (less negative) amplitudes for coincidences; and an Electrode main effect: F(3,36) = 61.75, $\varepsilon = 0.56$, $\eta^2_p = 0.84$, p<.001. Although no significant Tone × Electrode interaction was found, the topographical distributions (Figure 2.5, bottom row) show that the coincidence effect was numerically more posterior than the tone-related N1. The ANOVA of the N1 amplitudes at the mastoids showed that the reversed N1 amplitude was higher (more positive) for coincidences: F(1,12) = 4.94, $\eta^2_p = 0.29$, p<.05.

ERPs – Between-experiment comparisons.

Since significant N1-attenuation was found in both experiments, the question in the following analysis of the normalized amplitudes was whether interactions between the Experiment and Tone factors indicating a difference in the magnitude (ratio) of N1-attenuation would be found. The ANOVA of the normalized N1 amplitudes showed a Tone main effect: F(1,12) = 6.96, $\eta^2_p = 0.37$, p<.05, an Electrode main effect: F(3,36) = 71.48, $\varepsilon = 0.53$, $\eta^2_p = 0.86$, p<.001, and a Tone × Electrode interaction: F(3,36) = 3.73, $\varepsilon = 0.58$, $\eta^2_p = 0.24$, p<.05. Pair-wise t-tests showed that the normalized 1000+ ms post-key-press tone-minus-coincidence difference was larger at Pz and Oz than at Fz (t-scores > 2.21, p-scores < .05). The mean N1-attenuation-ratio was 33 ± 12 % and 35 ± 12 % in Experiment 1 and 2, respectively. This analysis confirms that the coincidence effect (as presented in Figure 2.5) is more posterior than the tone-related N1.

In contrast with Experiment 1, in Experiment 2, a parietally distributed positivity, identified as a P3b waveform (see e.g. Polich, 2007) was observable, peaking at 296 ms and 319 ms for the corrected 1000+ ms post-key-press tone and corrected coincidence waveform, respectively (Figure 2.4). The ANOVA of the amplitudes in the P3b intervals (see Table 2.1) showed a significant three-way interaction, which was followed up by separate ANOVAs in the two Experiments. In Experiment 1, only an Electrode main effect: F(3,36) = 5.95, $\varepsilon = 0.51$, $\eta^2_p = 0.33$, p<.05; in Experiment 2, however, a Tone main effect: F(1,12) = 28.43, $\eta^2_p = 0.70$, p<.001, showing higher (more positive) amplitudes for the coincidence; and an Electrode main effect: $F(3,36) = 8.69$, $\varepsilon = 0.46$, $\eta^2_p = 0.42$, $p<.01$ was found.

Effect	**Result**
Experiment	$F(1,12) = 47.80$, $\eta^2_p = 0.80$, $p<.001$
Tone	$F(1,12) = 5.52$, $\eta^2_p = 0.31$, $p<.05$
Electrode	$F(3,36) = 10.84$, $\varepsilon = 0.58$, $\eta^2_p = 0.47$, $p<.001$
Experiment × Tone	$F(1,12) = 36.79$, $\eta^2_p = 0.75$, $p<.001$
Experiment × Electrode	$F(3,36) = 5.51$, $\varepsilon = 0.44$, $\eta^2_p = 0.31$, $p<.05$
Experiment × Tone × Electrode	$F(3,36) = 4.52$, $\varepsilon = 0.44$, $\eta^2_p = 0.27$, $p<.05$

Table 2.1. The significant effects in the Experiment × Tone × Electrode ANOVA of the amplitudes in the P3b time intervals (20 ms long windows centered at the P3b peak amplitudes for the corrected 1000+ ms post-key-press tone-, and coincidence-related group-averaged ERP waveforms measured in Experiment 2).

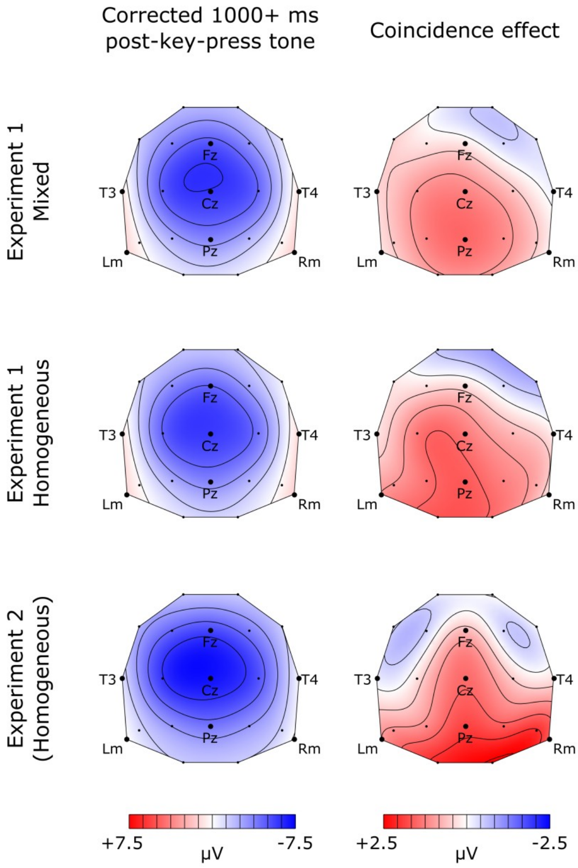

Figure 2.5. Group-average topographical distributions of the corrected 1000+ ms post-key-press tone ERPs (left column), and the corresponding coincidence-minus-tone differences (coincidence-effect, right column) in the N1 time interval in the Mixed (top row) and Homogeneous (middle row) conditions of Experiment1, and Experiment 2 (bottom row). The topographies were interpolated as described by Perrin, Pernier, Bertrand, & Echallier (1989, 1990), using a spline order of 4, a Legendre-polynomial order of 50, and no smoothing.

Discussion

The results showed attenuated auditory ERP responses to tone-action coincidences in comparison with tones separated by at least 1 s from preceding key-presses in both experiments, which confirms previous studies (Hazemann et al., 1975; and Study I, that is, Horváth et al., 2012). This attenuation, however, was not significantly modulated by the multiplicity of tone-frequencies in Experiment 1, or by the task-relevancy of tones and coincidences in Experiment 2. Although the present experimental design cannot prove the absence of voluntary attentional effects, the contribution of these to the attenuation of the auditory N1 (and P2 in Exp. 1) is unlikely to be substantial (as suggested by Ford et al., 2001). That is, mechanisms of voluntary attention cannot explain the action-sound coincidence-related attenuation of the auditory ERPs.

When tones were task-irrelevant (Experiment 1) only the N1 and P2 waveforms were elicited. In Experiment 2, tones elicited an additional P3b, attributable to the task-relevancy of the tones. For coincidences, the P3b was enhanced, and it was preceded by an N2 waveform, which probably reflects the significance of this event in task- and motivational terms (see e.g. Ritter, Simson, Vaughan, Friedman, 1979; Polich, 2007). The emergence of these ERPs, however, made it impossible to assess P2-modulation in Experiment 2.

Because the present results suggest that the contribution of voluntary attention is not substantial, the attenuation-effect should be explained by other mechanisms. One hypothesis put forward in Study I, was that N1 (and possibly P2) attenuation is caused by processes initiating the formation of action-sound contingency-representations which are triggered when action-sound coincidences occur. Some of these hypothetical processes may be reflected in the present results (as well as in the results of Study I.), by the topography of the coincidence-effect in the N1 time range, which showed that the effect was larger at posterior than at frontal sites (note that this difference was statistically significant only in the analysis of the normalized amplitudes involving data from both experiments). The coincidence-related enhancement of the positive aspect of N1 at the mastoids in Experiment 2, may also suggest that N1-attenuation does not reflect a “pure“, genuine auditory N1-effect, but, in part, may result from the emergence of a different (positive) ERP component in the same time range as the N1. It has to be noted, however, that the enhancement at the mastoids may also signal a stronger attentional focus on the auditory channel at the moment of the key-press, which is offset by the coincidence-related activity at the midline sites.

Whereas an attenuation of the auditory N1 subcomponent (as evidenced by the attenuation of its magnetic reflection) takes place in both types of paradigms (see e.g. Martikainen et al., 2005 and Horváth et al., 2012), the posterior ERP topography in the N1 time range seems to be a distinct feature of the coincidence-related effect. Due to the nature of the paradigm, one may speculate that this may be an ERP signature of processes initiating the formation of action-sound contingency-representations or binding actions and sensory events together (see Hughes, Desantis & Waszak, 2013a).

Other studies on the potential involvement of attention in action-related auditory ERP attenuation

Since Study II. was conducted, other studies also investigated whether the allocation of attention caused the N1 and P2 attenuations in contingent paradigms.

Timm, SanMiguel, Saupe, and Schröger (2013) investigated this hypothesis in a paradigm with self-initiated tones by instructing participants to count tones, key-presses or independent visual stimuli, thus manipulating the focus of attention, and found that these manipulations did not influence the magnitude of attenuation. One may suggest that these task-relevancy manipulations were ineffective in modulating the action-related ERP attenuation because the participants' primary task was to perform actions according to some rules. Because violating the rules resulted in somewhat adverse effects (loss of bonuses in Study II, and repeating the given block in the study by Timm and colleagues), one may speculate that participants were always inclined to make sure that the primary time interval production tasks were properly tended to, which may have limited the efficiency of the task-relevancy manipulations.

Saupe et al. (2013), on the other hand, found evidence suggesting that although the allocation of attention may not explain the phenomenon in its entirety, attention effects may contribute to N1 and P2 attenuation. In a contingent setting, they instructed participants to produce a sequence of button-presses with highly variable between-sound (i.e. between-action) intervals within a 1.8-5.0 s range in the Motor-Auditory condition. They administered two types of Auditory conditions: In the Active Listening condition participants had to detect too short (< 1.8 s) or too long (> 5.0 s) between-sound intervals in the replay of the sound sequence produced in the Motor-Auditory condition, whereas no task was given in the Passive Listening condition. Comparisons of the corrected Motor-Auditory ERPs to those obtained in the two Listening conditions showed that N1 was attenuated. The difference was larger for the comparison with the Active Listening condition, suggesting that attentional difference may contribute to the attenuation effect. Importantly, however a topographical difference in the attention-, and self-induction-related effects was found, which suggested that the origins of the effects were different: self-induction mainly affected the T-complex, while attention mainly affected the supratemporal N1 subcomponent.

In retrospect, (based on the findings of Study IV - Horváth, 2013b, Saupe et al., 2013; SanMiguel, Todd & Schröger, 2013) Tb-attenuation can be also observed in Experiment 1 (Fig. 6). A repeated measures Condition (Homogeneous vs. Mixed) × Tone (corrected coincidence vs. 1000+ ms post-key-press tone) × Electrode (T3, T4)

ANOVA of the amplitudes measured in a 20 ms long windows centered at the Tb peaks of the group average waveforms at T3 and T4 (140 ms for the Homogeneous and 138 ms for the Mixed conditions), showed only a Tone main effect: $F(1,14) = 25.74$, $\eta^2_p = 0.65$, $p<.001$, indicating significant coincidence-related attenuation. In Experiment 2, probably due to the overlap of the N2, such an effect was not observable.

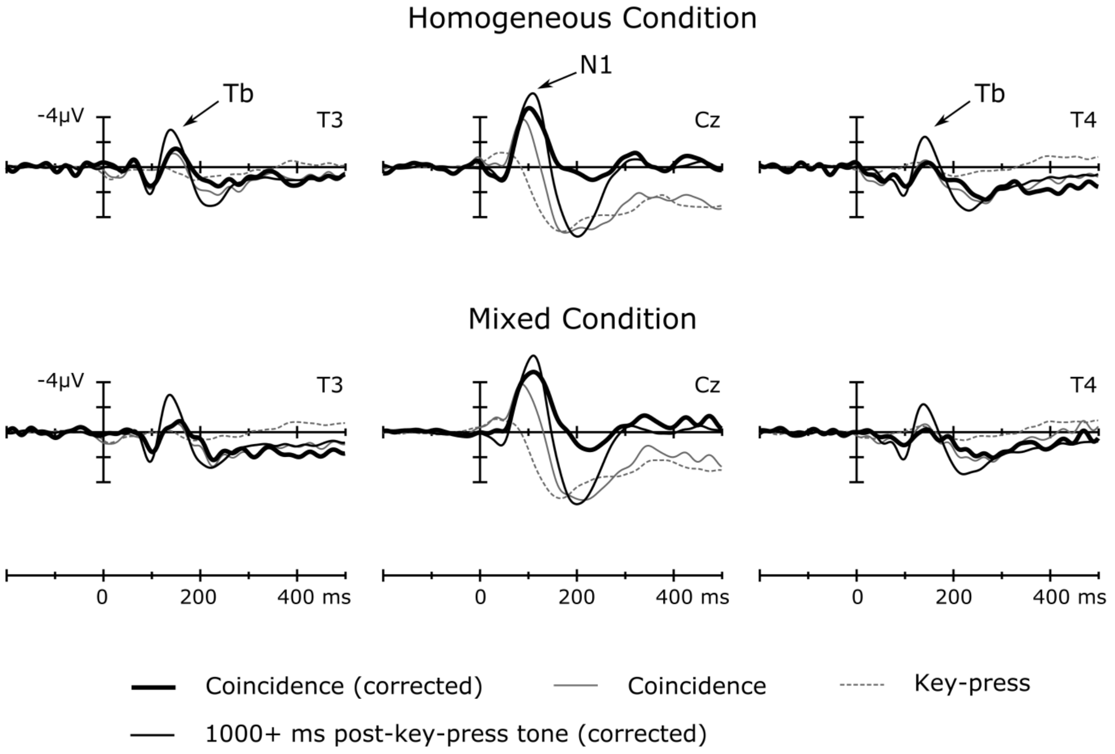

Figure 2.6. Group-average ERPs measured at the T3, Cz and T4 leads, to key-presses, coincidences, corrected coincidence ERPs , and corrected 1000+ ms post-key-press tone ERPs in the Homogeneous (top row) and the Mixed (bottom row) conditions of Experiment 1. The Tb and vertex N1 peaks are indicated by labeled arrows in the top row.

Looking back at Study I, Experiment 2 yielded similar results: Tb-attenuation can be also observed in Experiment 2 of Study I, but not in Experiment 1 of Study I (Figure 2.7). A repeated measures Tone (corrected coincidence vs. 1000+ ms post-key-press tone) × Electrode (T3, T4) ANOVA of the amplitudes measured in a 20 ms long windows centered at the Tb peaks of the group average waveforms at T3 and T4 (both at 133 ms), showed a Tone main effect: $F(1,12) = 7.79$, $\eta^2_p = 0.39$, $p = .016$, indicating significant coincidence-related attenuation, and an Electrode main effect: $F(1,12) = 8.35$, $\eta^2_p = 0.41$, $p = .014$, indicating higher (more negative) Tb-amplitudes at T3. Because the two ERP experiments of Study I differ in a number of ways, it would be difficult to attribute the observable Tb differences to any specific factor. Nonetheless, one may speculate that this may have to do with the difference in the time interval production tasks: In Experiment 1 a uniform 2-6 s interval distribution had to be produced, whereas in Experiment 2, the intervals had to be regular (4s). This would suggest that Tb-attenuation may be related to the production of regularly paced sound sequences (see also Study IV, in which a regular sequence had to be produced). Although this is an interesting possibility, the study by Saupe et al. (2013) does not fit this pattern: in that study, the sequence to be produced was also similarly variable as in Experiment 1 of Study I.

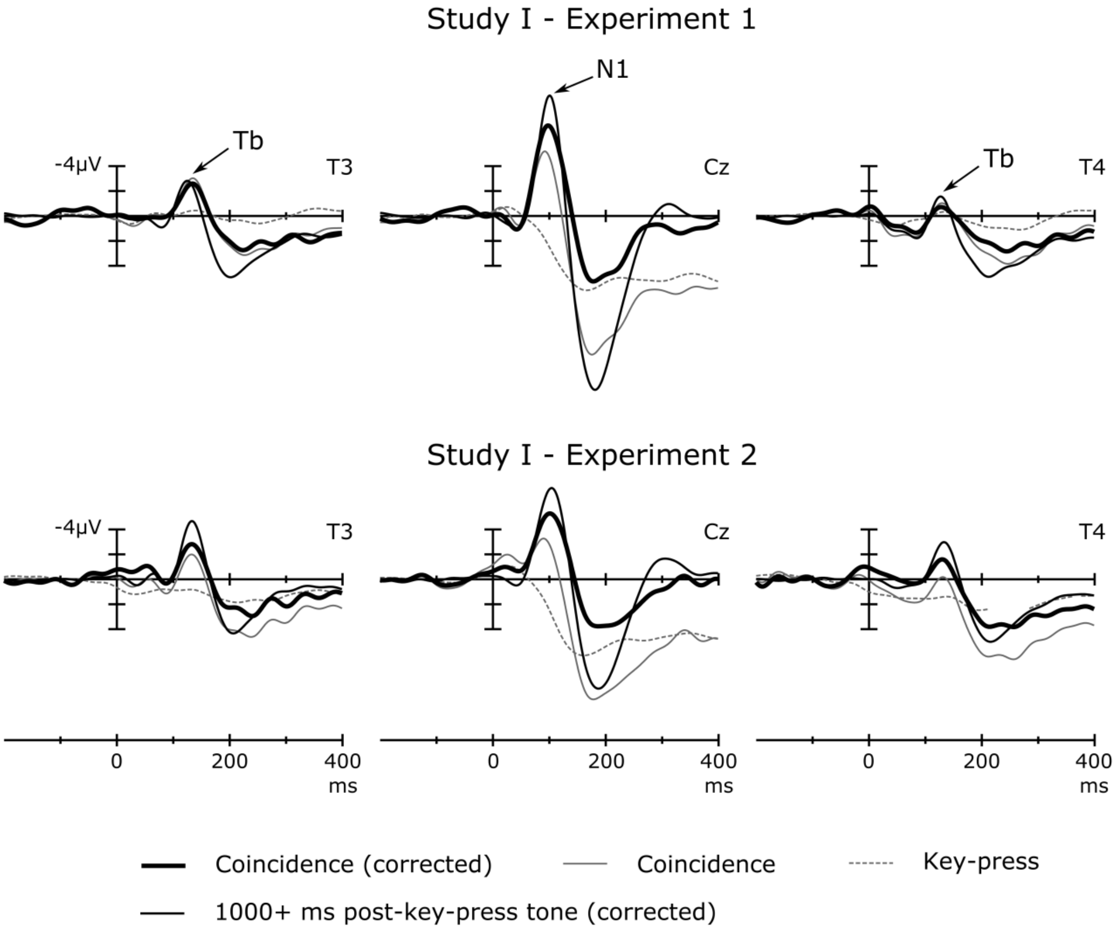

Figure 2.7. Group-average ERPs measured at the T3, Cz and T4 leads, to key-presses, coincidences, corrected coincidence ERPs , and corrected 1000+ ms post-key-press tone ERPs in Experiement 1 (top row) and 2 (bottom row) in Study I. The Tb and vertex N1 peaks are indicated by labeled arrows in the top row.

In summary, although it is difficult to reject the hypothesis that N1-attenuation is caused by attention-effects, the experiments explicitly testing this issue found either no evidence for it, or were able to dissociate attention- and action-related attenuation ERP contributions on the basis of topographical differences between the Tb- and vertex N1-effect. The fact that Tb-attenuation was observable in the coincidence-, as well as in contingent paradigm (Saupe et al., 2013) supports the notion that the two ERP-effects reflect (at least in part) similar processes.

Study III. The potential role of a peripheral process in action-related auditory attenuation [4]

Summary

Although action-related ERP attenuation could be caused by central processes, it may also reflect a peripheral mechanism: the co-activation of the stapedius muscle with the task-relevant effector, which reduces signal transmission efficiency in the middle ear, reducing the effective intensity of concurrently presented tones, which, in turn, elicit lower-amplitude auditory ERPs. Because stapedius muscle contraction attenuates frequencies below 2 kHz, no attenuation should occur at frequencies above 2 kHz. A self-induced tone paradigm was administered with 0.5, 2.0, and 8.0 kHz pure tones. Self-induced tones elicited attenuated N1 and P2 ERPs, but the magnitude of attenuation was not affected by tone-frequency. This result does not support the hypothesis that ERP attenuation to self-induced tones are caused by stapedius muscle contractions.

Introduction

The results of Study I also open up the possibility that the general assumption that action-related auditory ERP attenuation reflects a central mechanism may be unfounded. The available evidence does not rule out the potential involvement of a peripheral mechanism. The goal of the current study was to test the viability of this hypothesis.

It is well-known that the N1 waveform is sensitive to the physical characteristics of the eliciting sounds. Importantly, N1 amplitude decreases with decreasing sound intensity (Picton, Woods, Baribeau-Braun, & Healey, 1977; Näätänen & Picton, 1987), therefore an attenuated N1 response could reflect a mechanism changing auditory signal transmission efficiency at the periphery. One such mechanism in humans is the contraction of the middle-ear stapedius muscle, which leads to the effective attenuation of incoming sound intensity. Although stapedius muscle contraction can be triggered by the presentation of high intensity sounds through the stapedius reflex (see e.g. Mukerji, Windsor & Lee, 2010), stapedius muscle activation can be observed when vocalization, chewing or even bodily movements are initiated (i.e. the contraction starts 50-100 ms before the actual movement, Carmel & Starr, 1963, Salomon & Starr, 1963; Simmons, 1964), and it may be also under voluntary control in some individuals (Liberman & Guinan, 1998). Middle-ear muscle activity also seems to occur in conjunction with other motor events during sleep (Slegel, Benson, Zarcone, & Schubert, 1991). The transmission decrease caused by the acoustic activation of the stapedius muscle is about 10 dB (Rabinowitz, 1977), but much larger decreases may also occur (Pang & Peake,

---

[4] Based on Horváth, J., Burgyán, A. (2013): No evidence for peripheral mechanism attenuating auditory ERPs to self-induced tones. *Psychophysiology 50*, 563-569. doi:10.1111/psyp.12041

1987; Simmons, 1964). The contraction of the stapedius muscle primarily attenuates frequencies lower than 2 kHz (Borg & Zakrisson, 1974; Rabinowitz, 1977; Zakrisson & Borg, 1974). It has been suggested that the function of stapedius muscle contraction is to reduce the level of physiological noise resulting from our own movements (especially those generated in or near to our head, e.g. speech, chewing, head-movements, see Simmons, 1964), and to prevent the upward spread of masking (Liberman & Guinan, 1998; Pang & Guinan, 1997), that is, it may allow one to counteract the nonlinearly growing masking of higher frequencies by low frequency sounds (for a summary, see Moore, 2012).

Based on these findings, one may hypothesize that N1 attenuation is brought about by the co-activation of the stapedius muscle with the task-relevant effector, which leads to the effective presentation of a softer tone. Though this hypothesis has been mentioned in the literature (Ford, et al., 2001), its implications, have not been empirically tested. An overview of the literature on N1-attenuation in self-initiation- or coincidence-based paradigms shows that most studies used spectrally rich sounds (clicks, noise, or speech) or pure tones in the 0.5-2.0 kHz range, which matches the frequency range directly affected by the stapedius muscle contraction. If N1 attenuation was brought about by co-activation of the stapedius muscle with the task-relevant effector, then the magnitude of N1-attenuation should diminish for frequencies above 2 kHz, because of the decreased effect of stapedius muscle contraction on transmission for these frequencies. In the present study this implication of the peripheral hypothesis was tested by administering a self-induced tone paradigm with 500, 2000, and 8000 Hz pure tones.

Methods

Participants

14 young adult volunteers (six women, aged 18-25 years, mean 21 years, thirteen right-, one left-handed) participated in the experiment for monetary compensation. Data from two participants were not used in the analyses due to the exceeding number of ocular artifacts contaminating the EEG recordings (over 40% of the epochs in some conditions). All participants reported normal hearing and no history of neurological disorders. Before the experiment, participants gave written informed consent after the experimental procedures were explained to them.

Stimuli and procedures

Participants were sitting in an armchair in a sound-proofed room during the experiment. There were three types of conditions: In the Motor-Auditory condition, participants were instructed to press a rod-mounted button held in their dominant hand at regular 3 s intervals, and more importantly, count the tones triggered by the button-presses. Button-presses resulted in the presentation of a tone for the first 61 to 70 button-presses (randomly with uniform distribution). The randomness of the time-point at which the button “no longer produced tones” allowed monitoring whether participants attended the tones. Due to the constraints of the stimulation equipment, there was a 7 ms delay between the button-press and tone onset, which was taken into account during EEG processing. At the end of the block, participants reported the number of tones, and received feedback about accuracy. In the Motor condition, participants were instructed to press the button at regular 3 s intervals until instructed to stop. In the Auditory condition, the first 51-60 tones (random uniform distribution) from the sequence delivered in the preceding Motor-Auditory condition was replayed. Participants were instructed to count the tones, and report the number of tones at the end of the block. Due to a programming error, tone onset-to-onset intervals were 10 ms longer than in the Motor-Auditory condition, which, compared to the average onset-to-onset interval of 3 s, is unlikely to significantly affect the present results. The presentation of the conditions was organized into block-triplets. Each triplet started with a Motor-Auditory, followed by a Motor-, and concluded by the Auditory condition block. The experiment started with a practice triplet, in which the tone was a 50 ms long white-noise burst (including 10-10 ms linear rise and fall times). In the following experimental phase six triplets with 50 ms long pure tones (including 10-10 ms linear rise and fall times) of 500, 2000 and 8000 Hz frequency were administered. Each frequency was presented in two triplets. The order of triplets was randomized so that each type of triplet was delivered once in the first half and once in the second half of the experiment. The sound level was individually adjusted to 60 dB sensation level (above hearing threshold level) using the 2000 Hz tone. To compensate for the higher normal hearing threshold level (Suzuki, & Takeshima, 2004), the sound level of the 8000 Hz tone was set 10 dB higher than for the other two tones.

Analysis of the behavioral data

Between-key-press-intervals were analyzed in repeated-measures analysis of variance (ANOVA) using condition (Motor-Auditory vs. Motor) and frequency (500, 2000 and 8000 Hz) factors. Only the first fifty events in each experimental block were used in the analyses. The number of counting errors were analyzed in a repeated-measures condition (Motor-Auditory vs. Auditory) × frequency ANOVA. For all ANOVAs generalized eta-squared effect sizes (Olejnik and Algina, 2003; Bakeman, 2005) are reported; also Greenhouse-Geisser correction was applied to correct for potential violations of the sphericity assumption (in all such cases the corrected p-value, ε and the uncorrected degrees of freedom are reported). The alpha-level was set to 0.05. For all analyses all significant effects are reported.

EEG-recording and analysis.

The EEG was recorded with Ag/AgCl electrodes mounted on an elastic cap (EASYCAP, Herrsching, Germany) according to the 10% system (Nuwer, Comi, Emerson, Fuglsang-Frederiksen, Guerit, Hinrichs et al., 1998) with a sampling rate of 1000 Hz and on-line, 200 Hz low-pass filtering. Additional electrodes were placed at the mastoids. The reference was placed on the tip of the nose, the ground on the forehead. Horizontal EOG was obtained by an electrode placed under the right eye.

Vertical EOG was calculated off-line by subtracting the horizontal EOG signal from the signal recorded at Fp2. The EEG was off-line re-referenced to average reference, 20 Hz low-pass filtered, and segmented into epochs of 600 ms corresponding to tone-onsets (Motor-Auditory and Auditory conditions) and time-points when a tone onset would have been if key-presses would have triggered a tone (Motor conditions), including a 200 ms interval preceding the onset of the tone (or the time-point when a tone would have occurred, see above). Amplitude calculations were referred to the average signal measured in this (baseline) interval for each channel. Only the first fifty epochs of each experimental block were used in the analyses. Epochs with a signal range exceeding 100 µV on any EEG of EOG channels were discarded.

The epochs were individually averaged for each type of condition and tone-frequency. Individual N1 and P2 amplitudes were calculated as average signals in 20 ms long intervals centered at the N1 and P2 peaks of the group average waveforms measured at Fz and Cz in the Auditory conditions. For each frequency, the tone-related ERPs elicited by the self-induced tones were estimated by subtracting the ERPs obtained in the Motor condition from the ERP obtained in the Motor-Auditory conditions, that is, Motor-Auditory-minus-Motor difference ERPs were calculated. Because N1 and P2 typically peak at Fz and Cz, respectively, and often show reversed peak amplitudes at the mastoids when an average reference is used, N1 and P2 amplitudes were measured as Fz-minus-average-mastoid, and Cz-minus-average-mastoid signals, respectively. This also maximizes the chance to detect potential between-frequency differences in the magnitude of attenuation. These were normalized by the amplitudes measured in the corresponding Auditory conditions (i.e., the individual amplitudes for each frequency were divided by the group-mean amplitudes from the corresponding Auditory condition), and submitted to a repeated-measures ANOVA using stimulation (self-induced vs. replayed) and frequency (500 Hz, 2000 Hz and 8000 Hz) factors. Normalization was necessary, since N1 and P2 amplitudes could be different for tones with different frequencies. In these analyses different attenuation-ratios for different frequencies would be signaled by Stimulation × Frequency interactions. Because it could be a priori assumed that the attenuation-ratio would be more modest in the 8000 Hz condition than in the others, planned comparisons of the attenuation-ratios between 8000 Hz and the other two frequencies were also conducted (paired Student's t-tests; we also report JZS Bayes factors - BF_{01} – based on the default Bayesian t-test, as proposed by Rouder, Speckman, Sun, Morey, & Iverson, 2009, to quantify evidence in favor of the null hypothesis). To test whether N1- and P2-attenuation was significant at each frequency, amplitudes in the auditory and motor-auditory conditions were compared by Student's t-tests as well. To test whether the topographical distributions of the attenuation effects differed from the topographical distributions of the corresponding N1 and P2 ERPs elicited in the auditory conditions, stimulation (motor-auditory, auditory) × electrode (Fz, Cz, Pz, Oz) ANOVAs were conducted with the amplitudes vector-normalized following the method described by McCarthy & Wood (1985) for each frequency. In these analyses, a significant interaction would mean that the attenuation effect was not a "pure" modulation of the underlying ERP, rather, that it was brought about by the emergence of a different ERP component. The choice of midline electrodes is motivated by the fact that Horváth (2013, and in press) found that action-tone coincidence related ERP attenuation in the N1 time-interval was (in part) due to the emergence of a posterior, midline ERP effect.

Results

Behavioral measures

Counting performance is presented in Figure 3.1. The condition × frequency ANOVA of the number of erroneous counts showed that participants made more errors in the motor-auditory than in the auditory conditions (significant condition main effect): $F(1,11)=7.86$, $\eta_G^2=0.06$, $p<0.05$.

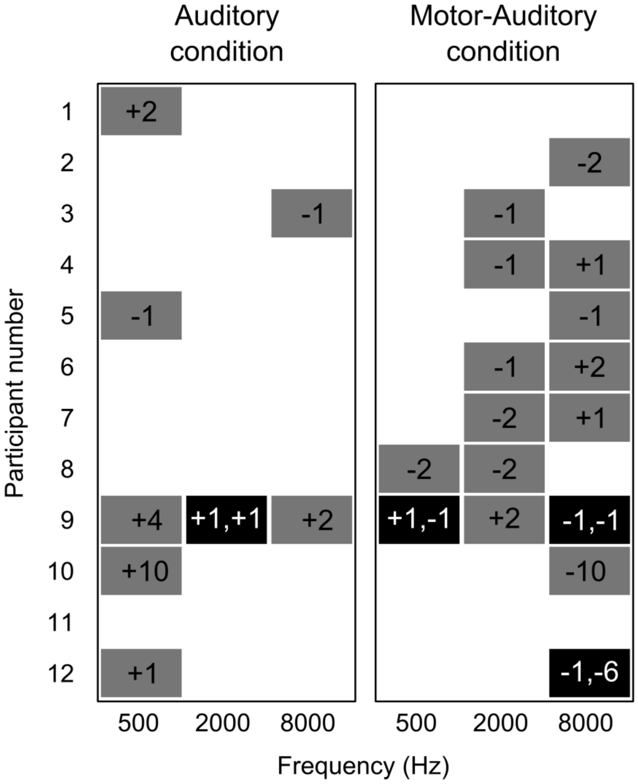

Figure 3.1. Counting performance for each participant, in each Auditory and Motor-Auditory condition with each tone-frequency. Gray fields indicate that the given participant miscounted in one of the two blocks of the given condition, whereas black fields indicate miscounting in two blocks. The positive or negative numbers in these fields indicate the difference between the reported and the correct number of tones.

The condition × frequency ANOVA showed that between-key-press intervals were shorter in the motor-auditory (2874 ± 157 ms, mean ± standard deviation) than in the motor (3098 ± 239 ms) conditions (significant condition main effect): $F(1,11)=16.50$, $\eta_G^2=0.19$, $p<0.01$.

ERPs

Figure 3.2 shows the ERPs elicited at selected midline electrodes and the average of the mastoid signals for all tone frequencies and conditions. The topographies of the N1 and P2 waveforms in the tone-related ERPs, as well as the corresponding (reversed) attenuation-effects are shown in Figure 3.3.

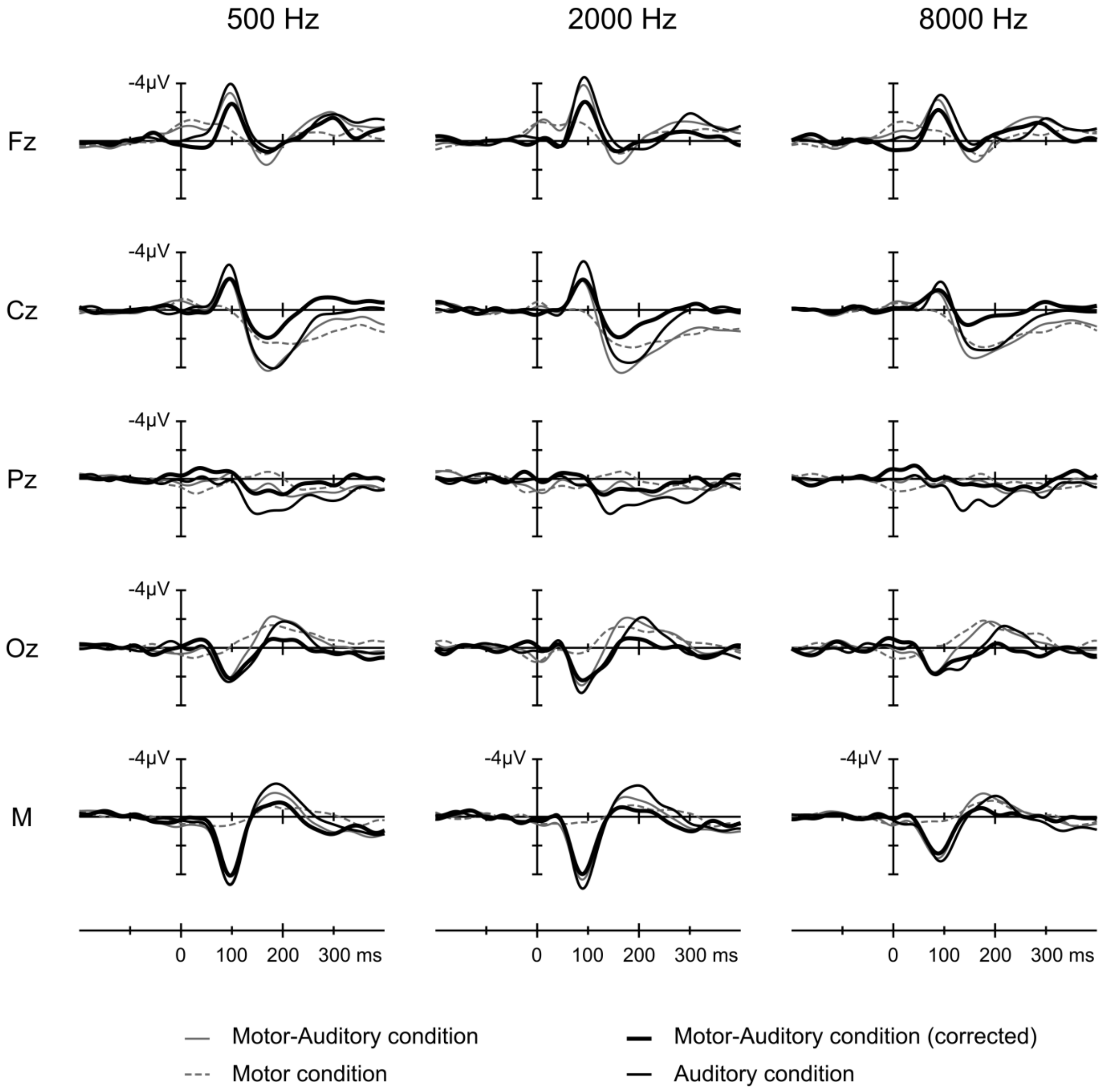

Figure 3.2. Group-average ERP waveforms for the three frequencies in the Auditory-, Motor-, and Motor-Auditory conditions, and the corresponding Motor-Auditory-minus-Motor difference waveform (corrected Motor-Auditory waveform) measured at Fz, Cz, Pz, Oz leads, and the average of the left and right mastoids signals (M). Tone onset is at the crossing of the axes.

N1 peaked at 97, 93, and at 93 ms at Fz following the onsets of the 500, 2000, and 8000 Hz tones, respectively. The ANOVA of the normalized amplitudes showed a significant stimulation main effect only: $F(1,11)=65.58$, $\eta_G^2=0.13$, $p<0.001$ (for the frequency main effect: F(1,11) = 0.15,ε = 0.86, p = 0.83; and for the interaction: F(2,22) = 0.14, ε = 0.98, p = 0.87). The attenuation ratio was significantly different from zero at each frequency: It was 24 ± 24 %, t(11) = 3.44, $p<0.01$, BF_{01} = 0.104; 29 ± 17 %, t(11) = 5.85, $p<0.001$, BF_{01} = 0.003; and 27 ± 26 %, t(11) = 3.68, $p<0.01$, BF_{01} = 0.073, respectively for the 500, 2000 and 8000 Hz tones. The planned direct comparisons of the attenuation-ratios showed no significant differences: t(11) = 0.15, p = 0.89, BF_{01} = 4.61, for the 500 vs. 8000 Hz comparison; and t(11) = 0.63, p = 0.54, BF_{01} = 3.86, for the 2000 vs. 8000 Hz comparison. The stimulation (motor-auditory, auditory) × electrode (Fz, Cz, Pz, Oz) ANOVAs of the vector-normalized amplitudes conducted separately for each frequency showed no significant interactions.

P2 peaked at 180, 181 and at 180 ms at Cz following the onsets of the 500, 2000, and 8000 Hz tones, respectively. The ANOVA of the normalized amplitudes showed only a significant stimulation main effect: F(1,11)=17.49, η_G^2=0.30, $p<0.01$ (for the frequency main effect: F(1,11) = 0.28, ε = 0.71, p = 0.68; for the interaction: F(2,22) = 0.52, ε = 0.80, p = 0.56). Attenuation ratios were significantly different from zero at each frequency: they were 57 ± 53 %, t(11) = 3.73, $p<0.01$, BF_{01} = 0.067; 63 ± 65 % t(11) = 3.35, $p<0.01$, BF_{01} = 0.121; and 71 ± 62 %, t(11) = 4.02, $p<0.01$, BF_{01} = 0.043, respectively, for the 500, 2000 and 8000 Hz tones. The planned direct comparisons of the attenuation-ratios showed no significant differences: t(11) = 1.07, p = 0.31, BF_{01} = 2.76, for the 500 vs. 8000 Hz comparison; and t(11) = 0.70, p = 0.50, BF_{01} = 3.72, for the 2000 vs. 8000 Hz comparison. The stimulation (motor-auditory, auditory) × electrode (Fz, Cz, Pz, Oz) ANOVAs of the vector-normalized amplitudes conducted separately for each frequency showed no significant interactions.

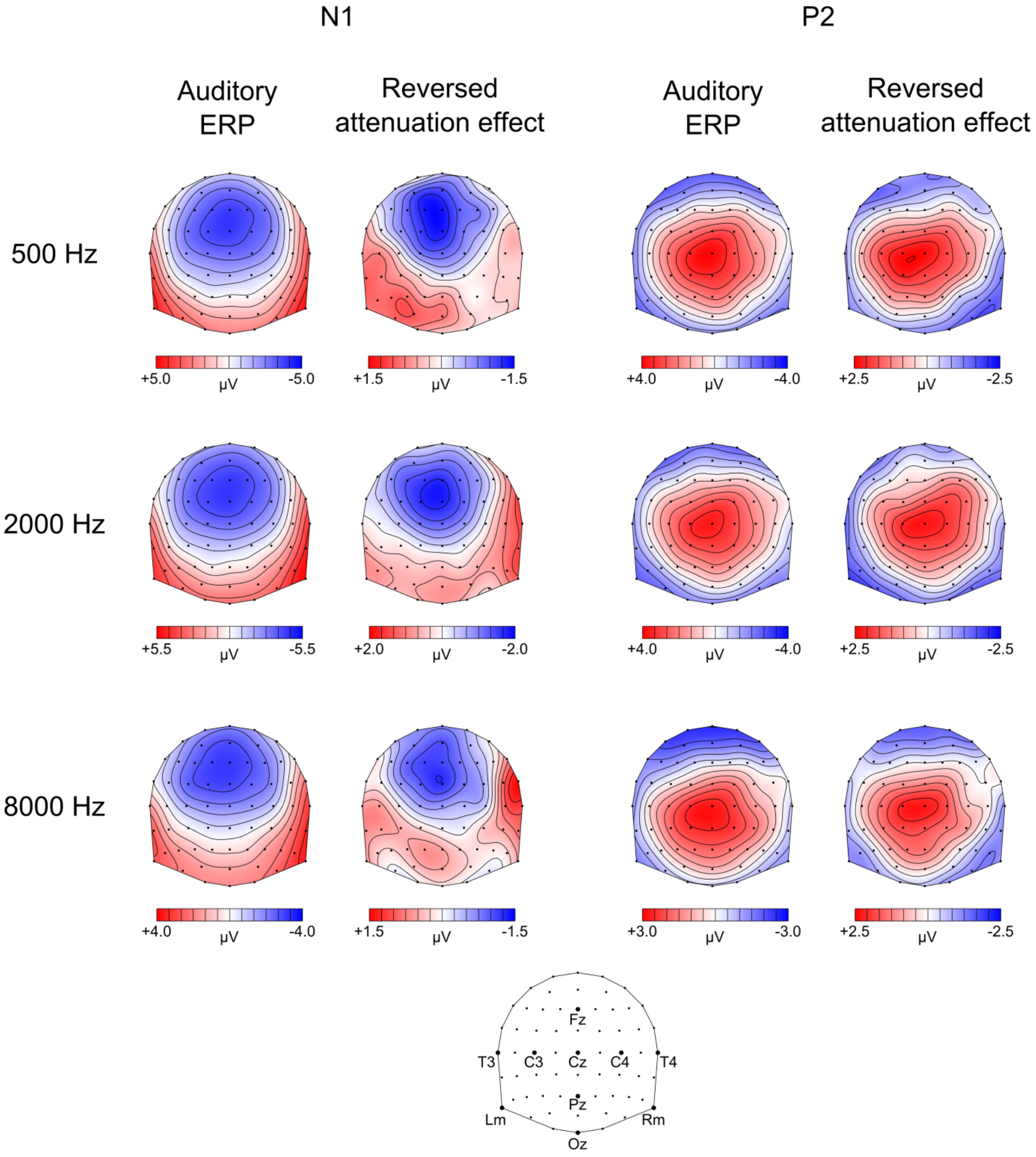

Figure 3.3. Group-average topographical distributions of the N1 and P2 ERPs elicited in the Auditory conditions (average signals in 20 ms long windows centered at the group-average peak latencies), and the corresponding reversed attenuation effects for the three tone frequencies. The signal range differs between the individual topographical distributions in order to emphasize similarities or differences in shape.

Discussion

Self-induced tones elicited attenuated N1 and P2 ERPs, which confirms previous results (e.g. Aliu et al., 2009; Baess et al., 2011; Ford et al., 2007), and extends the range of pure tone frequencies at which N1/P2 suppression was observed. The results of the topographical analyses are compatible with the interpretation that the ERP-effects were modulations of the N1 and P2 waveforms (i.e. they were not caused by the emergence of other components). Importantly, no significant difference in the magnitude of attenuation as a function of tone-frequency was found. The direct comparisons gave some evidence for the equality of the attenuation effects (i.e. JZS Bayes factors larger than 3 in favor of the respective null-hypotheses, see Rouder, et al., 2009, except for the P2 attenuation-ratios in the 0.5 vs. 8 kHz comparison, where P2-attenuation was numerically stronger in the 8 kHz condition). This suggests that N1 and P2 attenuations across tone-frequencies do not decrease or change substantially in the 0.5-8.0 kHz range. The results do not support the hypothesis that auditory ERP attenuation for self-induced tones is caused by the co-activation of the stapedius muscle with the key-press-related movement, and are compatible with the notion that these N1/P2 attenuations are caused by non-frequency-specific, possibly central mechanisms.

In the motor-auditory and the auditory conditions participants performed a tone counting task, and participants had to keep a steady key-pressing pace in the motor and the motor-auditory conditions. In contrast to previous studies (in which no task was given in the auditory condition) this arrangement allowed us to make sure that participants attended the tones in the auditory condition in a qualitatively similar fashion to that in the motor-auditory condition. On the other hand, however, counting performance was lower in the motor-auditory than in the auditory condition. This difference in performance may signal that less attentional resources were allocated to the tones in the motor-auditory condition. Because N1 elicited by sounds is enhanced when attention is directed towards the sound channel (see e.g. Hillyard et al., 1973), it could be argued that N1 and P2 may be elicited with higher amplitude in the auditory than in the motor-auditory condition, because more attention is directed towards the tones in the auditory condition.

The present results do not allow direct conclusions on whether the ERP attenuation observable for sounds induced by speech-producing movements is caused by peripheral or central processes. Because it is only an assumption that the mechanism underlying auditory ERP attenuation is the same for speech- and non-speech-related actions, and because the peripheral hypothesis seems more plausible for speech-producing movements, it is possible that ERP attenuation for speech-producing actions is based (in part) on the co-activation of the stapedius muscle. On the other hand, if these central processes are triggered even for the arbitrary association of a button-press and a tone, then such processes probably play a substantial role when it comes to highly familiar patterns of action-sound correspondence like in speech.

In summary, the present results do not support the hypothesis that auditory ERP attenuation for self-induced tones is caused by the co-activation of the stapedius muscle with the tone-inducing movement.

Study IV. Exploring the role of actions in the coincidence effect[5]

Summary

Studies demonstrating action-related auditory (ERP) attenuation in a non-speech-production context used exclusively key-presses as actions. It was hypothesized that attenuation might be caused by robust key-press-effect associations formed by long-term interactions with everyday devices. Key-pressing may be special because most everyday devices are designed to comply with the perceived affordance that a button should be pressed to trigger an effect. Therefore, key-presses would attenuate auditory processing, but key-releases would not. In a coincidence arrangement, participants marked time intervals by pressing or releasing a key. Tones coinciding with a key-press or –release elicited similarly attenuated Tb, vertex N1, and P2 ERPs, suggesting that coincidence-related auditory attenuation was not brought about by special key-press-effect associations. Whereas Tb and P2 attenuations were pure amplitude modulations, vertex N1 was attenuated (partly) by an overlapping coincidence-related ERP.

Introduction

One of the hypotheses introduced in the discussion of Study I suggested that despite the lack of a causal action-tone relationship the coincidence paradigm, the action-related auditory ERP attenuation measured for non-speech-related actions may nonetheless reflect the cancellation of sensory re-afference. Because in all studies investigating this phenomenon, actions were key- or bar-presses, the attenuation effects may have resulted from an action-sound contingency representation brought about by extensive training through the widespread use of key-press-based interfaces in everyday devices. In other words, key-presses may be similar to speech producing actions in that we have extensive experience with situations in which key-presses consistently cause sensory effects. Key-presses or button-presses trigger a multitude of sensory event types in everyday situations (e.g., turning on the lights, typing, etc.), which may lead to a general expectation that key-presses should result in some kind of a sensory event. Such expectations may be robust in the sense that it may not be easy to extinguish (or deactivate) them even if a key-press mostly fails to produce an effect in a given situation (as in a coincidence paradigm). Such an expectation could be encoded in an internal forward model which represents an abstract effect: the occurrence, but not every specific detail of the corresponding sensory event (see also Desantis, Hughes & Waszak, 2012; Hughes, Desantis & Waszak, 2013a). Therefore, attenuated auditory responses may not reflect that the neural system captured the action-sound contingencies in the matter of minutes during the experiments, rather, they may reflect

---

[5] Based on Horváth, J. (2013). Action-sound coincidence-related attenuation of auditory ERPs is not modulated by affordance compatibility. *Biological Psychology 93,* 81-87. doi:10.1016/j.biopsycho.2012.12.008

the workings of an already existing forward model. (Although Aliu and colleagues [2009] suggested that action-sound contingency representations are built up rapidly in the matter of minutes, a closer inspection of their results precludes firm conclusions regarding this matter). This hypothesis fits the results from experiments with both contingent and non-contingent action-sound arrangements; and it also explains why seemingly arbitrary actions (key-presses) lead to similar patterns of auditory attenuation to those found with speech-producing actions. That action-sound associations established by long-term training influence perception has been demonstrated by Repp and Knoblich (2007). They showed that performing patterns of finger-movements which would result in ascending or descending tone pairs on a piano induced a corresponding bias in the perception of an ambiguous pitch change for pianists but failed to induce a bias in non-pianists. The goal of the present study was to investigate whether the coincidence-related attenuation of auditory event-related potentials (ERPs) demonstrated in previous studies was brought about by such an association.

In order to test this hypothesis, an asymmetry in the association of sensory effects to key-presses and key-releases in everyday devices was exploited: Although there are devices in which the intended effect is brought about by releasing an already pressed key, in most everyday devices the intended effect(s) result from key-*presses*. Such a design allows for easy operability, because it is compatible with the users' expectations and the button's perceived affordance (i.e., that buttons should be pressed, Norman, 1989; McGrenere and Ho, 2000; e.g. when typing, a character is presented when the key is pressed, not when it is released). In the present study it was hypothesized that the prevalence of devices compatible with this perceived affordance may lead to an asymmetry in the cognitive representation of action-effect associations, and this asymmetry may be reflected in the magnitude of coincidence-related attenuation of the auditory ERPs. That is, if the coincidence-effect was brought about by an association between key-presses and a wide class of sensory events established by long-term practice, then the attenuation of auditory responses should be stronger for key-press- than for key-release events. To test this hypothesis, the coincidence paradigm was modified: instead of instantaneous key-press-release actions, participants were required to produce separate key-press and key-release events, that is, after the key was pressed, they had to keep on pressing they key for longer periods, and then release it. Concurrently, but independently from the key-events, a sequence of tones with random inter-tone-intervals was presented. Key-press-tone coincidences were expected to result in attenuated auditory N1 and P2 ERP responses, whereas key-release-tone coincidences were expected to result in no (or much reduced) ERP attenuations. The electrode layout also allowed the separation and assessment of the coincidence effect on the Tb component of the T-complex waveform.

Methods

Participants.

20 young adult volunteers (8 women, aged 18-24 years, mean 21 years, all right-handed) recruited through a student part-time job agency participated in the experiment for monetary compensation. All reported normal hearing status, and gave written informed consent before the experiment, after the experimental procedures were explained to them.

Stimuli and procedures.

Participants were sitting in an armchair in a well-lit, sound-attenuated room. The experiment consisted of 14 experimental blocks, separated by short breaks as needed, with a longer (10 min.) break at the middle of the session. In each block, 75 sinusoid tones of 1000 Hz frequency, 50 ms duration (including 10-10 ms linear rise and fall times), and 50 dB sensation level (above hearing threshold level) intensity were presented through headphones (HD-600, Sennheiser, Wedemark, Germany) with onset-to-onset intervals sampled randomly from a 2-6 s uniform distribution. The duration of the blocks was around 5 minutes.

Participants were instructed to press and release a rod-mounted push-button held in their right (dominant) hand so that the time intervals between key-presses and key-releases (and key-releases and key-presses) would be 4 s. That is, they had to press the key, hold it down, wait, release the key, wait, press the key, hold it down, and so on during the experimental blocks. The task was practiced with on-line feedback before the experiment: after each key-press or -release event, a histogram of the between-key-event-intervals with 1 s bins centered at 4 s was updated on a screen in front of them. During the experiment, the same type of histogram was presented only at the end of each block as a performance feedback. To minimize eye-movements, participants were instructed to keep looking at a fixation puppet placed in front of them.

The objective of the paradigm was to allow the comparison of auditory ERPs to tones coinciding with key-events, and ERPs to tones far away from key-events. To produce coincidences, the pre-generated tone presentation schedule was revised after each key-press or key-release event (see also Study I., Horváth et al., 2012), by shifting the tone schedule uniformly towards the key-event so that the next tone to be presented was delivered immediately if it was to be presented within the next 300 ms, or at the smallest integer multiple of 300 ms if it was scheduled after that. For example, if a tone was to be presented 675 ms after a key-event, it was presented at 600 ms instead. This adjustment was only performed in reference to the last key-event preceding a tone (that is, if multiple key-events preceded a tone, re-scheduling was only done in reference to the last event). Adjusting the timing of tones not immediately following a key-event was necessary to keep the distribution of inter-tone-intervals the same between coincidences and non-coinciding tones following key-events (see below). Due to hardware limitations, there was an 11 ms delay between key-events and sound onsets, which was taken into account during ERP processing.

EEG-recording and analysis

The EEG was recorded with a Synamps2 amplifier (Compumedics, Victoria, Australia) with 1000 Hz sampling rate, and 200 Hz on-line lowpass filtering, using Ag/AgCl electrodes mounted on an elastic cap (EASYCAP, Herrsching, Germany) according to the 10% system (Nuwer, et al., 1998). Electrodes were also placed at the two mastoids. The reference electrode was placed on the tip of the nose. A further electrode was placed under the right eye to record the horizontal EOG. The vertical EOG was obtained off-line by subtracting the horizontal EOG signal from the signal recorded at Fp2.

The EEG was off-line 1-20 Hz bandpass filtered (Kaiser-windowed sinc finite impulse response filter, beta of 4.53, 5859 coefficients; transition bandwidth 0.5 Hz, stopband attenuation at least 50 dB). 600 ms long epochs including a 200 ms pre-event interval were extracted and averaged for coincidences; and tones preceded by key-presses or key-releases by at least 1200 ms (referred to as 1200+ ms post-key-event-tones). Only tones preceded by a key-event were included, because the manipulation of the tone presentation schedule reduces the inter-tone-interval for tones preceded by key-events in comparison to those preceded only by another tone. Because the N1 waveform is known to be reduced when inter-tone-intervals are reduced (Näätänen & Picton, 1987), including tones preceded only by other tones in the comparison with the coincidence ERPs would have confounded the results (even if the inter-tone-interval was only reduced by 1-300 ms).

The objective of the analyses was to compare auditory ERPs elicited by coincidences and the 1200+ ms post-key-event tones. Whereas for the 1200+ ms post-key-event tones the contribution of the key-event-related ERP is negligible with the filtering used (1 Hz high-pass cutoff frequency, for comparison see Study I, Horváth et al., 2012, and Study II, Horváth, 2013a), the auditory ERP contribution within the coincidence-related ERP has to be estimated. For this reason, epochs corresponding to key-events separated by at least 1 s from any other events were extracted, averaged and subtracted from the averaged coincidence ERPs. In the following, this difference waveform is referred to as corrected coincidence waveform. Only epochs with no events in the last 400 ms of the epoch were included. Epochs within the first 10 s of each block, as well as those with a signal range exceeding 100 µV on any channels were discarded.

Individual ERP amplitudes were measured as the average signal in a 20 ms window centered at the peaks of group-average ERP waveforms. ERP amplitudes were submitted to Tone (corrected coincidence vs. 1200+ ms post-key-event tone) × Key-Event Type (press or release) × Electrode repeated measures analyses of variance (ANOVA). For the Tb waveform T3, C3, Cz, C4 and T4; for N1 and P2 Fz, Cz, Pz, and Oz were used as Electrode levels. Greenhouse-Geisser correction was applied whenever the effect structure allowed a potential violation of the sphericity assumption (ε-values and uncorrected degrees of freedom are reported). Partial eta-squared are reported. In these analyses all significant effects are reported. To determine whether the topographical distributions of the found coincidence-related ERP-effects differed from that of the ERPs elicited by the 1200+ ms post-key-event tones, the same types of ANOVAs were conducted with amplitudes normalized separately for the levels of the Tone factor. In these analyses a significant interaction including the Electrode factor indicates a difference in the shape of the distributions (McCarthy & Wood, 1985), which suggests that the effect is not a simple modulation of the tone-related ERP but an additional ERP waveform.

Results

Behavioral measures

Participants complied with the instruction. Mean individual key-event separations were between 3.60 and 4.36 s. The group-average key-press-key-release and key-release-key-press intervals did not differ significantly (paired t-test; means and standard deviations: 4.01 ± 0.61 and 3.97 ± 0.57 s, respectively.) Tones coincided 2.4 ± 0.7 % of the time with key-presses, and 2.5 ± 0.6 % with key-releases (not significant difference).

ERPs

Group-average ERP waveforms at midline electrodes for tone-key-press- and tone-key-release coincidences and the corresponding key-press and key-release ERPs and the corrected coincidence waveforms are presented in Figure 4.1. In Figure 4.2, the corrected coincidence and the 1200+ ms post-key-event tone-related ERPs are presented for midline and lateral electrodes. In Figure 4.3, the topographical distributions of the vertex N1, Tb and P2 ERPs, and the corresponding attenuation-effects are presented.

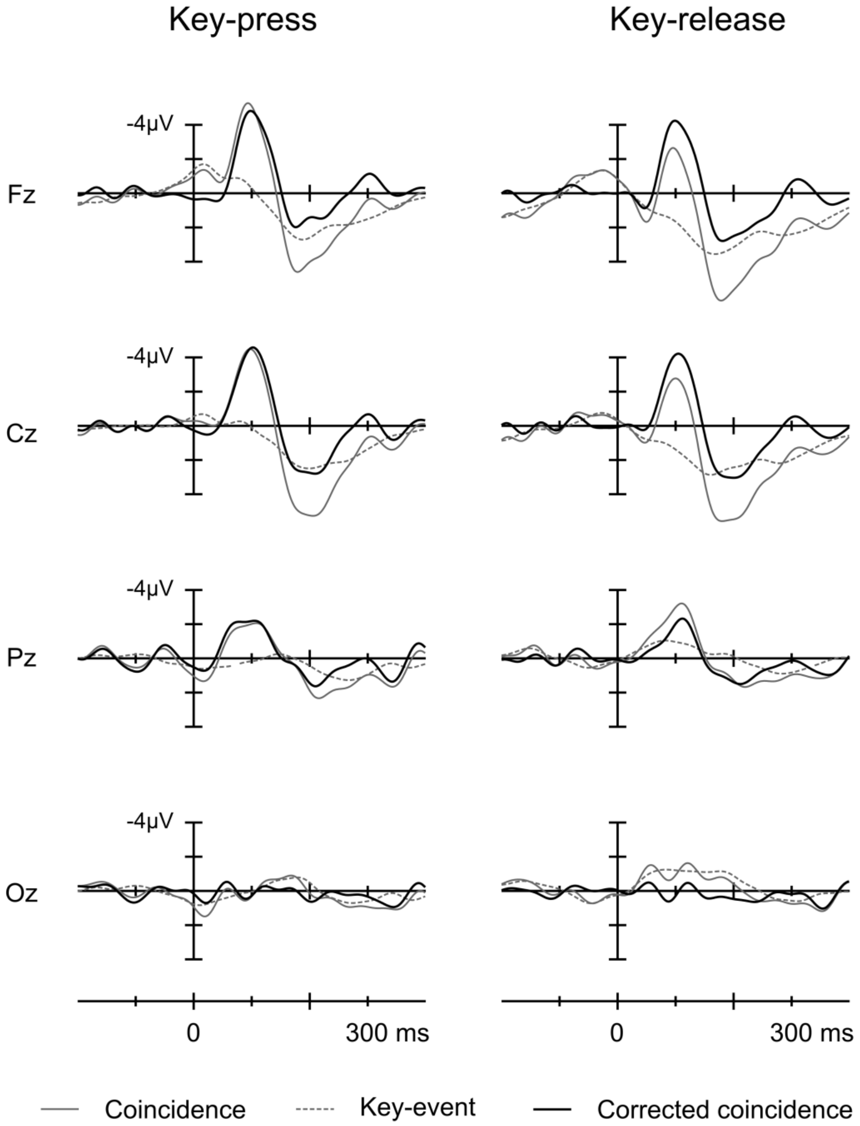

Figure 4.1. Group-average ERP waveforms at the Fz, Cz, Pz, Oz leads, elicited by key-event-tone coincidences, and key-events separated by at least 1 s from any other event, and the corresponding coincidence-minus-key-event difference waveforms (corrected coincidence) for key-press (left) and key-release (right) events. Tone onset (for coincidences) is at the crossing of the axes. For key-events the crossing of the axes is at the time point at which the tone onset would have been if the key-event would have been a coincidence.

For the corrected 1200+ ms post-key-event tones the vertex N1 peaked at 109 and 108 ms; the P2 at 189 and 191 ms, respectively for key-presses and key-releases; and Tb peaked at 136 ms for tones.

The analysis of the vertex N1 amplitudes showed a Tone main effect: $F(1,19)=10.35$, $\eta_p^2=0.35$, $p<.01$, indicating an attenuated N1 response to coincidences; an Electrode main effect: $F(3,57)=23.70$, $\varepsilon=0.49$, $\eta_p^2=0.55$, $p<.001$; and a Tone × Electrode interaction: $F(3,57)=5.99$, $\varepsilon=0.55$, $\eta_p^2=0.24$, $p<.01$. No significant effects involving the Key-Event Type were found. Pair-wise t-tests of the tone-minus-coincidence difference amplitudes between electrodes showed that the difference was smaller at Fz than at any other electrode (t-scores>2.72, p-values <.05). The analysis of the normalized amplitudes showed a significant Tone × Electrode interaction: F(3,57)=18.91, ε=0.52, η_p^2=0.50, p<.001, which suggests that the coincidence-related difference is not (only) due to a modulation of the vertex N1 ERP, but to the emergence of an additional ERP component.

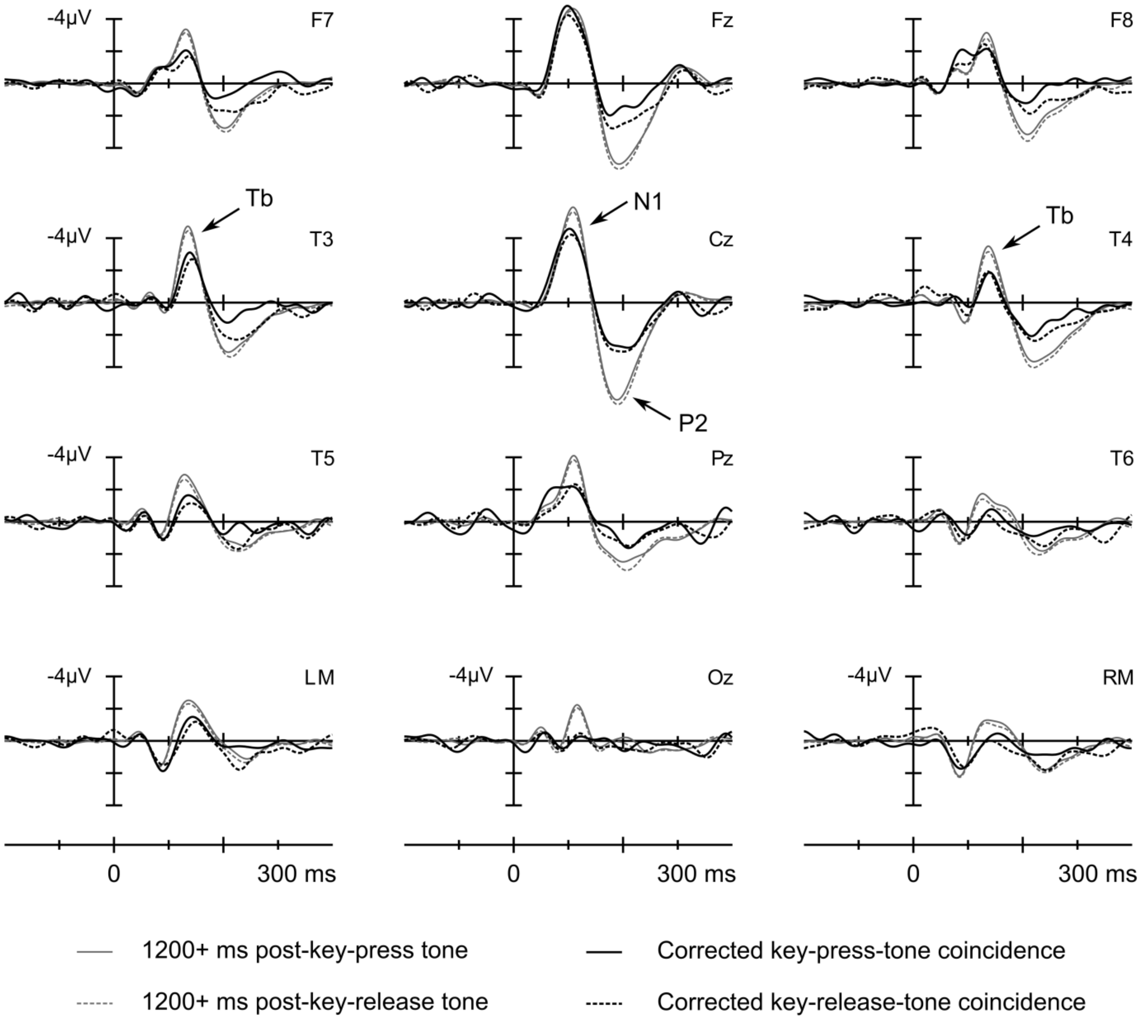

Figure 4.2. Group-average corrected coincidence and 1200+ ms post-key-event tone waveforms for key-press and key-release events measured at Fz, Cz, Pz, Oz, F7, T3, T5, F8, T4, T6 and the left and right mastoid (LM and RM respectively) leads. Tone onset is at the crossing of the axes. The peaks of the Tb waveform are marked with arrows at T3 and T4, the vertex N1 and P2 are marked at Cz.

The analysis of the Tb amplitudes showed a Tone main effect: F(1,19)=6.49, η_p^2=0.25, p<.05, indicating attenuated amplitudes for coincidences, an Electrode main effect: F(4,76)=8.06, ε=0.56, η_p^2=0.30, p<.001; and a Tone × Electrode interaction: F(4,76)=6.91, ε=0.57, η_p^2=0.27, p<.01. No significant effects involving the Key-Event Type were found. Pair-wise t-tests of the tone-minus-coincidence difference amplitudes between electrodes showed that the difference was larger at T3 than at C3, Cz, or C4; larger at T4 than at C3 or Cz; and at Cz it was smaller than at C3 or C4 (t-scores>2.16, p-values <.05). The analysis of the normalized amplitudes showed no significant interaction, which is compatible with the interpretation that the observed coincidence-effect is a modulation of the Tb.

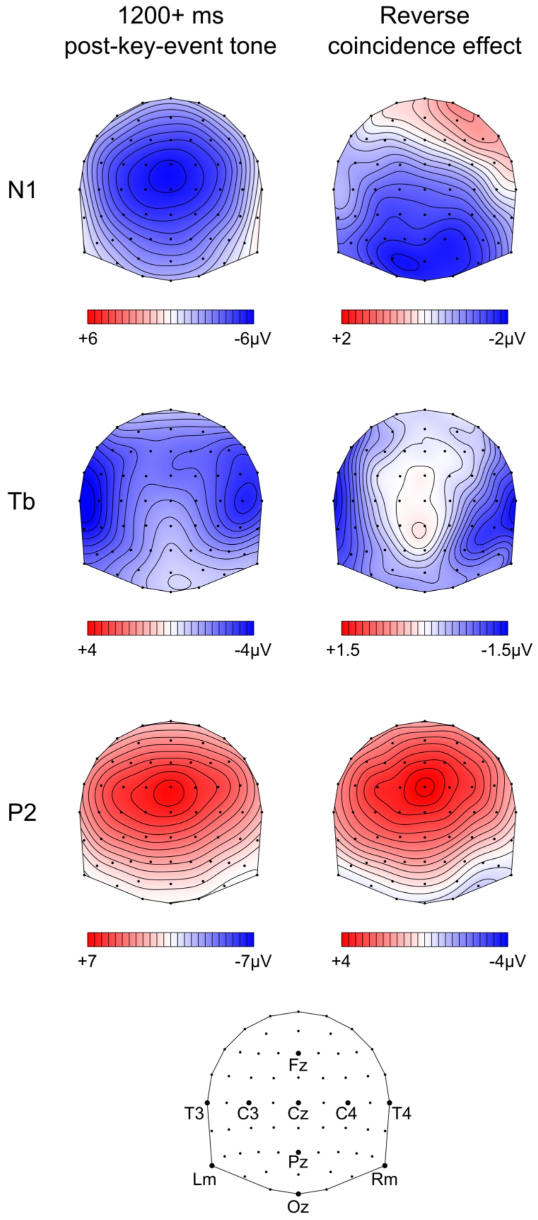

Figure 4.3. Group-average topographical distributions of the vertex N1 (top), Tb (middle), and P2 (bottom) ERP waveforms elicited by the 1200+ ms post-key-event tones (left column; average signals in 20 ms long windows centered at the group-average peak latencies), and the corresponding 1200+ ms post-key-event-tone-minus-coincidence difference topographies (i.e. the reversed coincidence effect; right column). The signal range differs between the individual topographical distributions in order to emphasize similarities or differences in shape.

The analysis of the P2 amplitudes showed a Stimulus main effect: F(1,19)=13.43, η_p^2=0.41, p<.01, indicating attenuated amplitude for coincidences, an Electrode main effect: F(3,57)=40.79, ε=0.55, η_p^2=0.68, p<.001, and a Stimulus × Electrode interaction: F(3,57)=22.17, ε=0.61, η_p^2=0.54, p<.001. No significant effects involving the Key-Event Type were found. Pair-wise t-tests of the tone-minus-coincidence difference amplitudes between electrodes showed that the difference significantly increased from the posterior electrodes to Fz and Cz, which did not significantly differ (t-scores>2.29, p-values <.05). The analysis of the normalized amplitudes showed no significant interaction, which is compatible with the interpretation that the observed coincidence-effect is a modulation of the P2.

Discussion

Tones coinciding with key-press or key-release events elicited attenuated vertex N1, Tb, and P2 ERPs, which confirms and extends the results of previous studies (Hazemann et al., 1975; and Study I and II, i.e. Horváth et al., 2012; Horváth, 2013a). The type of the action (key-press or –release), however, did not modulate the magnitude of attenuation for any of the components.

The topographical analyses showed no significant differences between the attenuation effects and the waveforms elicited by 1200+ ms post-key-event tones for the Tb and P2 ERPs, which suggests that Tb and P2 attenuations are pure modulation-effects. The attenuation of the Tb component, which is thought to be generated in secondary auditory cortices (e.g. Scherg & von Cramon, 1985, 1986; Ponton et al., 2002), confirms that actions coinciding with a tone influence a sensory processing stage (see also Study I, Horváth et al., 2012). For the vertex N1, however, a topographical difference was found, which suggests that the observed attenuation is the result of (partial) overlap by a different, coincidence-related, positive ERP component, which might not originate from auditory areas. This is on-a-par with the findings of Study I and II (Horváth et al.,2012; and Horváth, 2013a), in which the distribution of the attenuation-effect was more posterior than that of the vertex N1. At this point no functional description of the hypothetical posterior component in the vertex N1 time range can be given. One may speculate that because action-effect coincidence is one of the cues indicating a contingent relationship (Hume, 1739/1896), this ERP might reflect a process contributing to the maintenance or formation of a contingency-representation.

Although a direct comparison of the present results with ERP-attenuation patterns in a contingent stimulation protocol (i.e. in which actions lead to the presentation of sounds) was not possible, some of the studies using such protocols reported finding P2- (e.g. Knolle et al., 2012) and Tb-modulations (SanMiguel, Todd & Schröger, 2013). Note that in many studies only N1-effects were reported, but P2-attenuations are observable in the ERP figures (see e.g. Schäfer & Marcus, 1973; Baess et al., 2011). The posterior ERP effect in the vertex N1 time-range, on the other hand, seems to differ from the more fronto-central N1-effects found in such studies.

That the coincidence-related ERP attenuation was not modulated by the type of action in the present arrangement can be explained in various ways. The most far-reaching explanation is that the coincidence effect is not brought about by contingency-representations (forward models) at all, but it is caused by other mechanisms. A further possibility is that although there are such robust representations, the associations of key-presses and key-releases to sensory events are equally strong, because the asymmetry in the role of key-presses and key-releases in everyday devices is not sufficient to cause substantial long-term training differences. That is, even if infrequently used, devices coupling the intended effects with key-releases may still provide enough training to establish a contingency representation on-a-par with that of key-press-driven ones. It is seems also possible that action-effect associations may not only be general with respect to the effects (as described in the Introduction), but also with respect to the actions: key-presses and key-releases may be just two instances of "manual actions" associated to sensory effects. Both types of actions may simply be represented as manual actions because affordances (Gibson, 1979) are also determined by context: a released key can only be pressed, and a pressed key can only be released. Also, the abstraction of the effect described in the Introduction (that an internal forward model may represent only the occurrence, but not the specifics of the corresponding sensory event), may also be applicable to the action itself: a forward model might also encode only partial information about an action (for example, when it is going to occur) and map this to a sensory event.

In summary, the present results confirm previous findings of action-tone coincidence-related attenuation of the auditory vertex N1 and P2 ERP responses, and extend those with the observation of Tb attenuation. The manipulation of the action type was not effective in influencing the coincidence effect, which may be due to a pre-existing forward model representing a generalized action-effect association both in terms of the action and the corresponding effect, but it also leaves open the possibility that the coincidence effect may be caused by other mechanisms.

Study V. The role of mechanical impact in action-related auditory attenuation [6]

Summary

In the present study, the role of establishing a mechanical contact with an object was investigated in action-related auditory attenuation. In three experiments, participants performed a time-interval production task. In each experiment, in one condition the action involved touching an object, but no mechanical contact was made in the other. The estimated tone-related ERP contributions to the action-tone coincidence ERP waveforms (calculated by subtracting the action-related ERP from the coincidence ERP) were more attenuated when the action involved moving the finger and making a mechanical contact at the end of the movement. However, when participants kept their finger on a piezoelectric element, and applied pressure impulses without moving their finger, the action did not result in stronger attenuation of the tone-related auditory ERP estimates. Although these results may suggest that auditory ERP attenuation is stronger for actions resulting in mechanical impact, they also imply that mechanical impact may confound and lead to the overestimation of auditory ERP attenuation in such paradigms because the impact may result in faint, but audible sounds.

Introduction

Actions are seldom performed for their own sake. Most of our actions are goal-directed, and result in direct interactions with the environment. One common feature of all studies investigating action-related auditory attenuation with non-speech actions is that the actions are not simply movements, but *interactions with objects* (typically button-presses). The success of the action depends not only on our intention and capability to act, but on the physical properties of the manipulated objects and our knowledge of these properties. For example, too short or too weak movements may not result in a successful button-press (although these movements are still actions). The goal of the present study was to investigate whether the mechanical interaction with an external object played a role in the auditory ERP attenuations reported in the literature.

The hypothesis is an alternative specification of the rather general notion of “attending the action” put forward in Study I (Horváth et al, 2012). The present study investigated whether finger-movement-related auditory attenuation was caused by the allocation of input attention to the tactile stimulation occurring when a finger-movement was performed. Because previous studies used exclusively button-presses or -releases as actions, and participants’ primary task was to perform a sequence of actions corresponding to a pre-defined timing schedule, it was hypothesized that these settings might compel participants to attend the tactile feedback when the action was performed

---

[6] Based on Horváth, J. (in press) The role of mechanical impact in action-related auditory attenuation. *Cognitive, Affective, & Behavioral Neuroscience* doi: 10.3758/s13415-014-0283-x

to support maintaining the optimal interaction with device, or as an integral part of action preparation (Galazky et al., 2009; Brown et al., 2011).

Experiment 1 and 2 were conducted to test this hypothesis. In both experiments participants produced a sequence of finger-movements which either resulted in mechanical contact with an object or not. Experiment 1 used a coincidence paradigm. In Experiment 2, a contingent arrangement was administered, in which the given actions always resulted in tones. It was hypothesized that if mechanical interaction played a role in action-related auditory attenuation, then the magnitude of N1- and P2-attenuation would be reduced when no mechanical interaction took place. As described in the Methods of Experiment 1, establishing a mechanical contact with an external object might have caused faint, but audible transient sounds, so in order to better specify the nature of the mechanical contact-related effects found in these two experiments, in Experiment 3, in a coincidence arrangement, participants performed actions which did not produce such transient sounds: participants maintained continuous contact with an object and applied force impulses from time-to-time, or produced finger-movements without mechanical contact (Figure 5.1). As usual in the literature, in all three experiments, the contribution of the auditory ERP to the action-tone coincidence-related ERP was estimated by subtracting the action-related ERP registered when the action did not co-occur with a tone.

Experiment 1- Materials & Methods

Participants

19 young adults (nine women, aged 18-25 years, mean 21 years, 17 right-, two left-handed) participated in Experiment 1 for monetary compensation. They reported normal hearing and no history of neurological disorders, and gave written informed consent after the experimental procedures were explained to them.

Stimuli and procedures

Participants were seated in an armchair in a sound-proofed room during the experiment. In front of their dominant hand, a frame with an infrared light beam was positioned (Figure 5.1, left and middle column). Participants were instructed to interrupt the light beam with their index finger with a quick downward-upward movement (i.e. the resting state was above the light beam) at regular, 4 s intervals. The frame was positioned so that the light beam was close to the lowest point of the finger movement. To minimize eye-movements, participants were instructed to look at a fixation puppet placed in front of them during the task. Before the experiment, participants practiced the finger movements and their pacing in 2-3 blocks with 30 actions. In these blocks, a histogram of the between-action intervals was presented on a screen with 1 s bins, which was updated for each action. During the experiment, such feedback was only provided at the end of each block.

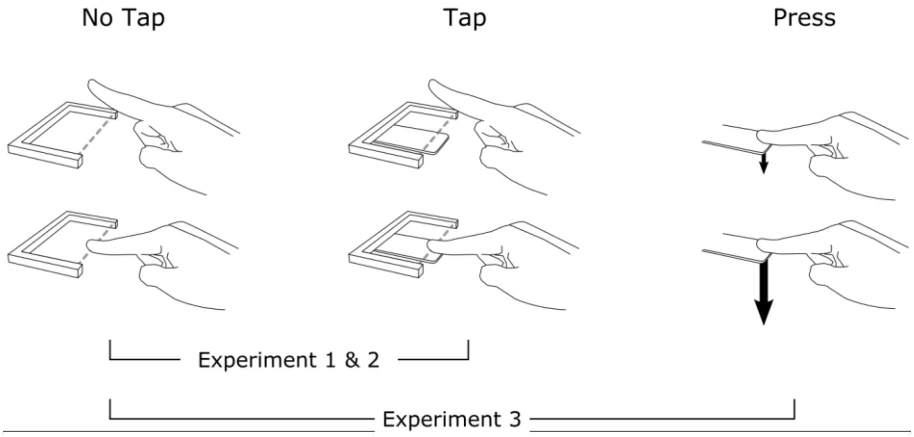

Figure 5.1. Action arrangements in the No Tap (left), Tap (middle), and Press (right) conditions. The top row shows the rest-, the bottom row the action phase. Participants interrupted the infrared (invisible) light beam by making a sharp downward-upward finger-movement in the No Tap and Tap conditions. In the Tap condition, the finger tapped on a plate positioned under the light beam. In the Press condition, participants maintained contact with a piezoelectric element and applied a short force impulse as an action. In Experiments 1 and 2 the No Tap and Tap conditions, in Experiment 3, the No Tap and Press conditions were administered.

There were two conditions: in the Tap condition, a plate was placed under the light beam, so that participants' fingers tapped on this plate when the light beam was interrupted. In the No Tap condition, the plate was removed, so no mechanical contact was made. Each condition was administered in seven blocks, each lasting 5 minutes. The blocks were presented in an interwoven order ("TNNTTNNT..."), with 9 participants starting with the Tap (denoted by "T"), 10 with the No Tap ("N") condition.

In both conditions, 1000 Hz pure tones of 50 ms duration (including 10-10 ms linear rise and fall times) were presented with pre-generated, random inter-tone-intervals sampled from a 2-6 s uniform distribution through open headphones (HD-600, Sennheiser, Wedemark, Germany). Tone intensity was individually adjusted to 60 dB above hearing threshold level. Similarly to that described in Horváth et al. (2012), the tone presentation schedule was adjusted when an action occurred: The tones following the action were shifted uniformly towards the action (i.e. earlier): if the next tone would have been delivered within 300 ms according to the original schedule, it was delivered immediately (an action-tone coincidence); if it was to be delivered between 300 and 600 ms, it was re-scheduled to 300 ms, and so on. Due to the constraints of the stimulation equipment, a constant action-tone timing difference of 7 ms was always present – this was taken into account during EEG processing. If multiple actions preceded a tone, the adjustment was made only in reference to the last action. Due to this adjustment, the inter-tone-interval was shorter for tones preceded by actions than for tones preceded by another tone. Because shorter inter-tone-intervals result in lower N1 amplitudes (Näätänen & Picton, 1987), only ERPs to tones immediately preceded by actions (and not by tones) were analyzed in this experiment.

Tapping on the plate inevitably causes sound, which may be audible. To give a rough estimate of the loudness of these sounds, after the experiment, the intensity of the presented tone (adjusted for an individual with 0 dB SPL hearing threshold level at 1000 Hz), and the sounds caused by tapping were recorded in a setup closely matching the original with microphones set in an artificial head (HSU III.2, without torso, Head Acoustics, Germany). To capture the power of the transient signals resulting from tapping, root-mean-squared (RMS) signal energy was calculated in 20 ms sliding windows across the whole recording. Although forceful hits on the plate (produced with the intention to make a sound as loud as possible by moving the whole hand from the wrist and hitting the plate with the index finger) could produce measurement maxima reaching -15 dB in comparison to the tone, the typical maximal power for taps (measured in a sequence of 40 taps delivered at a rate of once per second) was around -35 dB (standard deviation, SD: ±3 dB) in comparison to the tone. Note that these are rough estimates because the actual loudness also depends on a number of factors, most importantly on the individual threshold level for the tone used in the experiment and the individual movement characteristics.

EEG-recording and analysis

The EEG was recorded with 1000 Hz sampling rate and on-line, 200 Hz low-pass filtering (Synamp 2, Compumedics, Victoria, Australia) by 61 Ag/AgCl electrodes mounted on an elastic cap (EASYCAP, Herrsching, Germany) according to the 10% system (Nuwer et al., 1998). Additional electrodes were placed at the mastoids. The reference was placed on the tip of the nose, the ground on the forehead. Horizontal EOG was obtained by a bipolar setting between electrodes placed close to the outer canthi of the two eyes. Vertical EOG was calculated off-line by subtracting the signal of an electrode placed under the right eye from the signal recorded at Fp2. The EEG was off-line re-referenced to average reference, 1-20 Hz band-pass filtered, and segmented into epochs of 600 ms corresponding to events described below, including a 200 ms long interval preceding the given event. Amplitude calculations were referred to the average signal measured in this interval. Epochs with a signal range exceeding 100 μV on any channel were discarded. In some blocks some actions were not registered because of improper hand positioning or technical problems. When such problems were detected on-line, the affected block was re-run with a different pre-generated tone-schedule. To minimize the influence of potentially undetected actions on the ERPs, time intervals in which between-action-intervals exceeded 6.5 s were rejected from analyses.

The auditory ERPs elicited by action-tone coincidences were estimated by subtracting the ERPs elicited by actions from the ERPs elicited by coincidences (only actions separated from any other events by at least 1 s were used for this purpose). To assess whether the auditory ERP was attenuated, this estimate (*corrected coincidence ERP*) was compared to the *tone ERP*, that is, the ERP elicited by tones following an action by at least 1200 ms (and no action following it within 400 ms). Peak latencies and topographical extrema of the N1, and P2 auditory ERPs were assessed in the group average ERPs to such tones. ERP amplitudes were then calculated as average signals in

20 ms long windows centered at these time points at the electrode with the maximal signal for N1 and P2. Amplitudes were analyzed in Condition (Tap vs. No Tap) × Event (corrected coincidence vs. tone) repeated-measures analyses of variance (ANOVAs). The hypothesis-relevant question was whether a significant interaction was present, that is, whether the tone-minus-corrected-coincidence differences were modulated by the Condition factor. The power of the experiment to reveal this effect in the N1 and P2 time ranges was assessed as suggested by Cohen (1988), post-hoc. Effect size (d) was determined by dividing the mean between-condition attenuation difference by its standard deviation, for a two-sided arrangement. To assess whether attenuation was present despite the modulation by the Condition factor, corrected coincidence and tone ERPs elicited in the No Tap condition were compared by paired, two-tailed Student's t-tests.

Previous studies (Horváth et al., 2012, Horváth, 2013a) suggested that the topography of the coincidence-related N1-attenuation effect may be more posterior than that of the tone-related N1, suggesting that N1 attenuation was not a "pure" N1-effect, but it included an ERP of different origin as well. Because of this, the topography of the tone-related N1 and P2 was compared to the corresponding ERP effect or effects: that is, the modulatory effect of condition on the level of attenuation (i.e. the No Tap-minus-Tap differences of the attenuation effects), the main ERP attenuation effect (i.e. the mean of the attenuation effects across conditions), or the attenuation in the No Tap condition. For these analyses the vector-normalized amplitudes (as described by McCarthy & Wood, 1985) were submitted to Signal (tone-related ERP vs. the appropriate ERP-effect) × Electrode (Fz, Cz, Pz, Oz) ANOVAs for N1 and P2 separately. In these analyses, a significant interaction would mean that the ERP-effect was not a "pure" modulation of the underlying tone-related ERP, rather, that it was (in part) due to a different ERP component. That is, for example, an interaction in this analysis could signal that the topography of the modulation effect was different from that of the N1. In these analyses Greenhouse-Geisser correction was applied; Greenhouse-Geisser epsilon, unadjusted degrees of freedom, and adjusted p-levels are reported. The alpha-level was set to 0.05. Generalized eta-squared effect sizes (Olejnik and Algina, 2003; Bakeman, 2005) are reported. Amplitude values are reported in the form of amplitude ± standard deviation. Statistical calculations were performed in R (version 3.0.0, R Core Team, 2013).

Experiment 1- Results

Participants complied with the instruction. The mean between-action interval was 3968 ± 194 ms in the Tap, and 4004 ± 137 ms in the No Tap condition (no significant difference: t[18] = 0.755, p = 0.460; paired, two-tailed t-test). In the Tap, 5.7 ± 1.1 %; in the No Tap condition 5.7 ± 0.9 % of the sounds were coincidences (no significant difference: t[18] = 0.278, p = 0.784). The ratio of intervals excluded from the analyses due to potentially undetected actions did not differ between conditions (t[18] = 1.439, p = 0.167); this affected 0.6 ± 1.1 % of all between-action intervals.

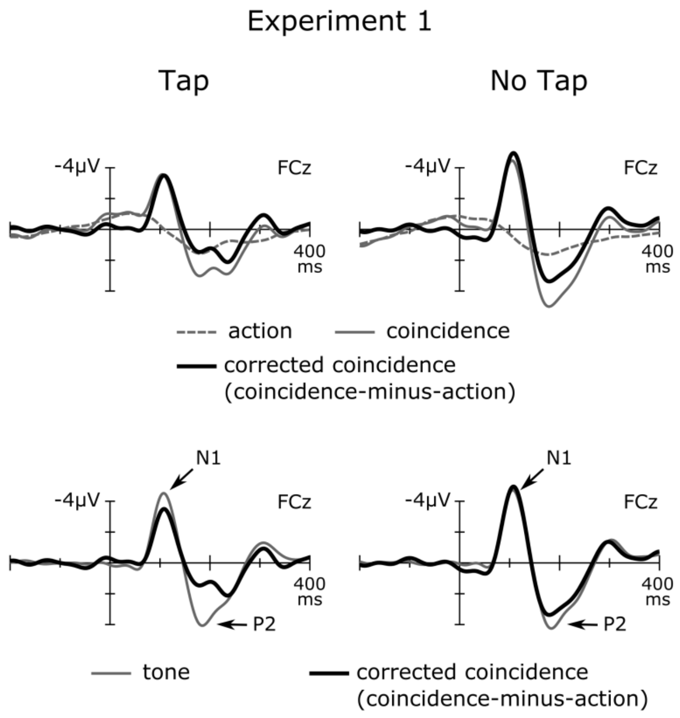

Figure 5.2. Group-average (N=19) ERP waveforms recorded in Experiment 1 at FCz in the Tap (left) and No Tap (right) conditions elicited by the respective actions, action-tone coincidences and the corresponding corrected coincidence waveforms (coincidence-minus-action difference waveforms; top row). Tone-related ERPs and corrected coincidence waveforms are contrasted in the bottom row.

The number of epochs contributing to coincidence ERPs was 23 ± 6 in both conditions. The number of epochs contributing to the tone ERPs was 231 ± 35 in the Tap and 223 ± 33 in the No Tap condition. Tones elicited the N1 and P2 waveforms (Figure 5.2) with peaks occurring at the expected latencies and electrode positions documented in the literature (Näätänen and Picton, 1987). For tones (at least 1200 ms after an action), N1 peaked at 107 ms at FCz in both conditions (Figure 5.2, the amplitudes - average signal in the 97-117 ms range at FCz - were -4.343 ± 2.331 , and -4.521 ± 2.452 μV in the Tap and No Tap conditions, respectively). The ANOVA of the amplitudes showed a significant Condition main effect: F(1,18) = 13.450, $\eta_G^2 = 0.028$, p = 0.002; and a Condition × Event interaction: F(1,18) = 6.934, $\eta_G^2 = 0.017$, p = 0.017, indicating stronger attenuation in the Tap, than in the No Tap condition. The modulation of the attenuation-effect (i.e. the mean between-condition difference of the attenuations) was 1.229 μV, yielding d = 0.604, and a (post-hoc) power of 0.702. The corrected coincidence and tone ERP amplitudes significantly differed in the Tap (t[18] = 3.229, p = 0.005, 1.002 μV attenuation), but not in the No Tap condition (t[18] = 0.697, p = 0.494, -0.226 μV "attenuation", i.e. the amplitude was numerically more negative for the coincidence).

P2 peaked at FCz at 185 ms and 184 ms in the Tap and No Tap conditions, respectively (the amplitudes - average signal in the 174-194 ms range at FCz – were 3.971 ± 2.797, and 4.157 ± 2.716 μV in the Tap and No Tap conditions, respectively).

The ANOVA of the amplitudes showed a significant Condition main effect: F(1,18) = 12.034, $\eta_G^2 = 0.039$, p = 0.003; a significant Event main effect: F(1,18) = 21.966, $\eta_G^2 = 0.099$, $p < 0.001$; and a Condition × Event interaction: F(1,18) = 12.489, $\eta_G^2 = 0.027$, p = 0.002, indicating stronger attenuation in the Tap, than in the No Tap condition. The modulation of the attenuation-effect was 1.708 µV, yielding d = 0.811, and a (post-hoc) power of 0.916. The corrected coincidence and tone ERPs amplitudes differed significantly in the No Tap condition: (t[18] = 2.125, p = 0.048, 0.856 µV attenuation), showing that despite the modulatory effect of the Condition factor, a significant attenuation was still present.

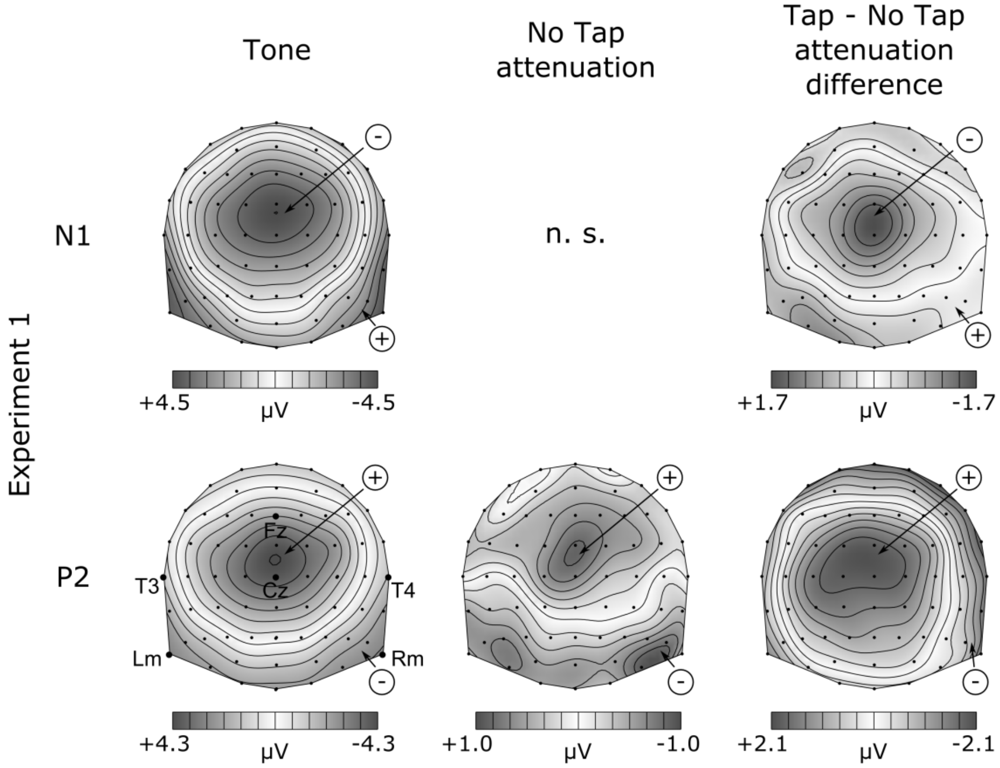

Figure 5.3. Group-average (N=19) topographical distributions of the N1 (top row) and P2 (bottom row) in Experiment 1. The left column shows topographies of ERPs elicited by tones (average signals in 20 ms long windows centered at the group-average peak latencies, averaged across conditions). The right column shows the corresponding attenuation-modulatory effect of the Condition factor, that is, the Tap-minus-No Tap attenuation differences. The middle column shows the attenuation effect (if significant) in the No Tap condition. Signal ranges differ between the individual topographical distributions in order to emphasize similarities or differences in shape. Positive and negative polarity areas are indicated by arrows.

The topographical distributions of the effects were visually similar to that of the respective tone-related ERPs (Figure 5.3). Comparing the topographies of the mean tone ERP and the attenuation-modulatory effect of the Condition factor (No Tap-minus-Tap attenuation differences) showed no significant Signal × Electrode interactions (N1: F[3,54] = 0.554, ε = 0.535, p = 0.543; P2: F[3,54] = 0.493, ε = 0.567, p = 0.585) providing no evidence for a different ERP component causing the modulatory effects of the Condition factor. Comparing the mean tone ERP and the attenuation effect in the No

Tap condition in the P2 time range showed no significant interaction ($F[3,54] = 0.229$, $\varepsilon = 0.608$, $p = 0.777$).

Experiment 1- Interim Discussion

The results of Experiment 1 are on-a-par with previous studies utilizing button-press actions: the estimated auditory N1 and P2 amplitudes for tones coinciding with an action in the Tap condition were reduced in comparison to the N1 and P2 elicited by tones separated from the preceding action by at least 1200 ms. Conforming to the hypothesis, when finger-movements did not result in mechanical contact with an external object (No Tap condition), the magnitude of attenuation was reduced. This suggests that mechanical impact played a substantial role in the observed ERP attenuation pattern. Because tapping caused faint, but probably audible transient sounds as evidenced by the measurements presented in the Methods section, interpreting these effects is not trivial, as discussed in the General Discussion section. The results, however, also indicate that mechanical impact was not the only factor behind the observed action-related ERP attenuation: a significant P2 attenuation was still present in the No Tap condition.

Because action-related auditory attenuation is often measured in a contingent stimulation arrangement, Experiment 2 was conducted with such an arrangement to investigate whether the effects observed in Experiment 1 were present in such a paradigm as well.

Experiment 2 - Materials & Methods

Participants

19 young adults (fourteen women, aged 20-26 years, mean 22 years, all right-handed) participated in Experiment 2 for monetary compensation. They reported normal hearing and no history of neurological disorders, and gave written informed consent after the experimental procedures were explained to them.

Stimuli and procedures

The stimulus and the action-related measurement arrangements were largely similar to that in Experiment 1 (see also Figure 5.1, left and middle columns). Participants were explicitly instructed to move only their fingers and avoid forcefully hitting the plate. Furthermore, a curtain was hanged above the arm, between the participant's head and hand, which occluded the sight of the hand and the frame. The instructions and procedures closely followed that described in the Study III (Horváth & Burgyán, 2013). As usual in contingent paradigms, there were three types of *stimulation arrangements*, which were administered in separate experimental blocks: In Motor-Auditory blocks, participants' finger-movements triggered the presentation of a tone (with 9 ms constant delay, which was taken into account during EEG processing). Participants were instructed to keep a steady, one-movement-every-4 s pace, and more importantly, count the elicited tones, because “at the end of the block the finger-movement will no longer elicit the tone”, and they will be required to report how many tones were elicited. At the end of the block, the experimenter provided feedback on accuracy, and also on movement timing. The number of tones was randomly sampled from a uniform distribution in the 61-70 range. In Motor blocks, participants were instructed to maintain a steady, one finger-movement every 4 s pace till the experimenter instructed them to stop. In Auditory blocks, participants listened to a sequence of tones and were instructed to count the tones, and report their number at the end of the block. As for Motor-Auditory blocks, the experimenter provided feedback on accuracy. Unknown to the participants, the sequence in each block was a replay of the first 51-60 tones (the sequence length determined randomly) from the tone-sequence produced in the preceding Motor-Auditory block. In contrast with paradigms in which no explicit instruction is given in these blocks, the instruction to count the tones and providing feedback on the accuracy allowed monitoring and motivating participants to attend the tones. The three types of stimulation arrangements were presented in *block triplets:* the first block was a Motor-Auditory, the second a Motor, and the third one was an Auditory block. Within each triplet, the condition (Tap or No Tap) was the same. For each participant two Tap and two No Tap triplets were administered. The order of the conditions was “TNNT” for ten, and “NTTN” for nine participants (where “T” stands for a Tap condition- and “N” for a No Tap condition triplet).

EEG-recording and analysis

Experimental blocks with improper hand positioning leading to unregistered movements were re-run immediately. The auditory ERPs elicited by action-tone events in the Motor-Auditory arrangement were estimated by subtracting the ERPs elicited by actions in the Motor arrangement from the ERPs elicited by the action-tone events. To assess whether the auditory ERP was attenuated, this estimate (*corrected coincidence – Motor-Auditory ERP*) was compared to the *tone – Auditory - ERP*, that is, the ERP elicited by tones in the Auditory arrangement. Only the first fifty trials of each experimental block were used in these calculations (corresponding to the lowest number of trials-minus-one in the Auditory blocks; this was chosen to maximize between-arrangement similarity). Peak latencies and topographical extrema of the N1, and P2 auditory ERPs were assessed in the group average ERPs to such tones. ERP amplitudes were then calculated as average signals in 20 ms long windows centered at these time points at the electrode with the maximal signal for N1 and P2. Amplitudes were analyzed in Condition (Tap vs. No Tap) × Event (corrected coincidence vs. tone) repeated-measures ANOVAs. Potential topographical differences for N1- or P2 attenuation-modulatory effects (between-condition attenuation differences) were assessed as in Experiment 1. To assess whether the experiment had sufficient power to detect differences in N1 and P2 attenuations equaling the mean between-condition attenuation differences measured in Experiment 1, the respective attenuation-difference measured in Experiment 1 was divided by the respective standard deviation of the attenuation-difference measured in Experiment 2. The resulting effect-size for each component was then used to calculate power. Furthermore, attenuation difference magnitudes detectable with 95% power were also calculated.

Experiment 2 - Results

Participants complied with the instructions. The between-action interval in the Tap condition was 3745 ± 333 ms in the Motor-Auditory and 4130 ± 347 ms in the Motor arrangements. In the No Tap condition it was 3849 ± 325 ms in the Motor-Auditory and 4268 ± 388 ms in the Motor arrangements. The Condition (Tap vs. No Tap) × Stimulation Arrangement (Motor-Auditory vs. Motor) ANOVA showed a significant Condition main effect: $F(1,18) = 4.864$, $\eta_G^2 = 0.030$, $p = 0.041$; and a Stimulation Arrangement main effect: $F(1,18) = 20.071$, $\eta_G^2 = 0.259$, $p < 0.001$. That is, participants kept a slower pace in No Tap than in the Tap conditions, and in the Motor than in the Motor-Auditory arrangements.

Most tone counting errors were off-by-one errors, with a couple of instances of larger deviations. The Condition (Tap vs. No Tap) × Stimulation Arrangement (Motor-Auditory vs. Auditory) ANOVA of the number of erroneous counts showed only a tendency for more errors in the Motor-Auditory than in the Auditory arrangement: $F(1,18) = 3.729$, $\eta_G^2 = 0.081$, $p = 0.069$ (overall, 29 counting errors occurred in the Motor-Auditory, and 14 in the Auditory arrangements out of a total of 76 experimental blocks).

The number of epochs contributing to the average ERPs were 85 ± 15, 86 ± 15, and 93 ± 7, in the Motor-Auditory, Motor, and Auditory arrangements, respectively, in the Tap; and 86 ± 19, 90 ± 12, and 93 ± 7, in the Motor-Auditory, Motor, and Auditory arrangements, respectively, in the No Tap condition.

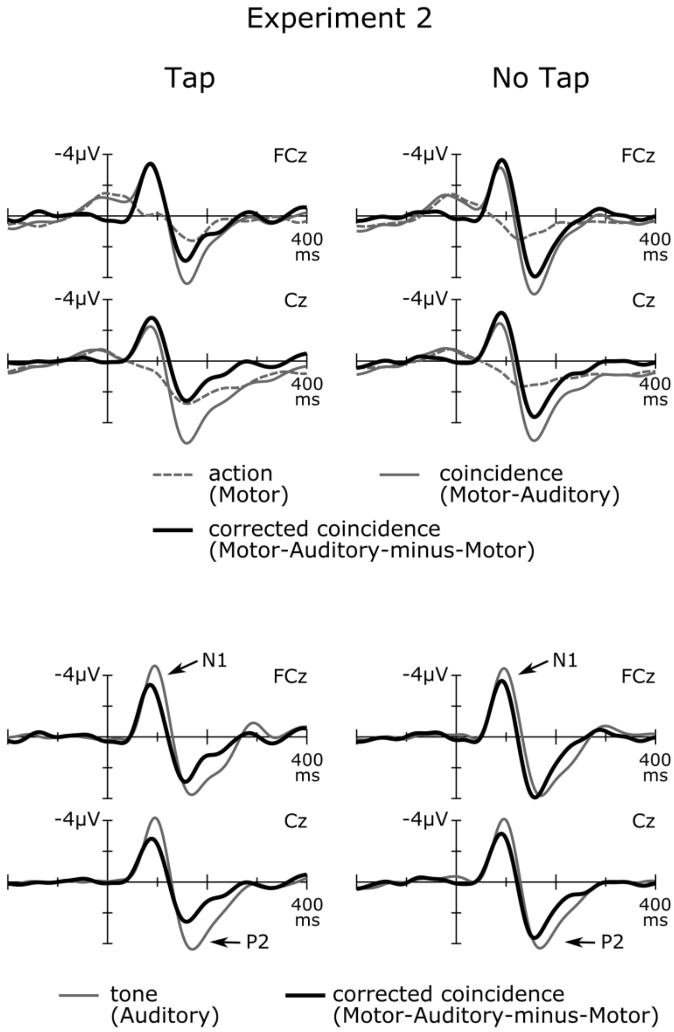

Figure 5.4. Group-average (N=19) ERP waveforms in Experiment 2 in the Tap (left) and No Tap (right) conditions elicited by the respective actions (obtained in the Motor arrangement), action-tone coincidences (obtained in the Motor-Auditory arrangement) and the corresponding corrected coincidence waveforms (coincidence-minus-action, that is, Motor-Auditory-minus-Motor difference waveforms; top panel). Tone-related ERPs (obtained in the Auditory arrangement) and corrected coincidence waveforms are contrasted in the bottom panel. Because N1 elicited by tones (in the Auditory arrangement) peaked at FCz, and P2 peaked at Cz, both signals are shown in separate rows.

Tones elicited the N1 and P2 waveforms (Figure 5.4) with peaks occurring at the expected latencies and electrode positions documented in the literature (Näätänen and Picton, 1987). For tones in the Auditory arrangement, N1 peaked at 95 ms at FCz in the Tap and at 96 ms in the No Tap condition (the amplitudes - average signal in the 86-106 ms range at FCz – were -4.435 ± 1.677, and 4.296 ± 1.979 µV in the Tap and No Tap conditions, respectively). The Condition (Tap vs. No Tap) × Event (corrected coincidence vs. tone) ANOVA of the N1 amplitudes showed a significant Event main effect: $F(1,18) = 34.417$, $\eta_G^2 = 0.114$, $p < 0.001$, indicating attenuated N1 response (by 1.273 µV) to the self-induced tones. (There was also a tendency for a Condition × Event interaction: $F(1,18) = 3.930$, $\eta_G^2 = 0.009$, $p = 0.063$, hinting at the possibility of stronger attenuation in the Tap, than in the No Tap condition). The effect size corresponding to the between-condition attenuation-difference in Experiment 1 (1.229 µV) was d = 0.853, and the power to detect such a difference was 0.940. The mean attenuation difference detectable with 95% power was 1.261 µV in this arrangement.

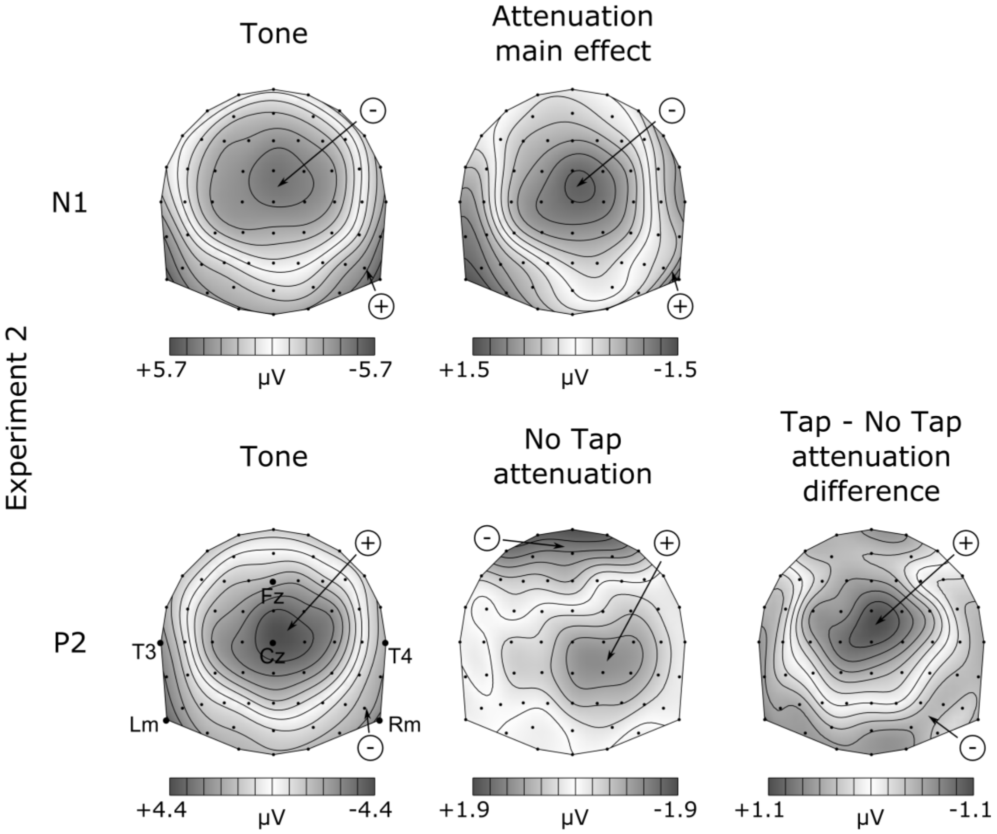

Figure 5.5. Group-average (N=19) topographical distributions of the N1 (top row) and P2 (bottom row) in Experiment 2. The left column shows topographies of ERPs elicited by tones (average signals in 20 ms long windows centered at the group-average peak latencies, averaged across conditions). In the top row, the middle column shows that attenuation main effect (tone-minus-corrected coincidence waveform averaged across conditions). In the bottom row the right column shows the attenuation-modulatory effect of the Condition factor (i.e., the Tap-minus-No Tap attenuation difference); the middle column shows the attenuation effect in the No Tap condition. Signal ranges differ between the individual topographical distributions in order to emphasize similarities or differences in shape. Positive and negative polarity areas are indicated by arrows.

P2 peaked at Cz at 170 ms in the Tap, and 168 ms in the No Tap condition (the amplitudes - average signal in the 159-179 ms range at Cz – were 4.294 ± 2.060, and 4.220 ± 1.699 µV in the Tap and No Tap conditions, respectively). The Condition × Event ANOVA of the amplitudes showed a significant Condition main effect: F(1,18) = 5.270, $\eta_G^2 = 0.015$, p = 0.034; a significant Event main effect: F(1,18) = 29.640, $\eta_G^2 = 0.162$, $p < 0.001$; and a Condition × Event interaction: F(1,18) = 4.560, $\eta_G^2 = 0.021$, p = 0.047, indicating stronger attenuation in the Tap, than in the No Tap condition (by 0.963 µV) . The effect size corresponding to the between-condition attenuation-difference in Experiment 1 (1.708 µV) was d = 0.850, and the power to detect such a difference was

0.938. The mean attenuation difference detectable with 95% power was 1.760 μV in this arrangement. The corrected coincidence and tone ERPs amplitudes significantly differed in the No Tap condition: t(18) = 2.921, p = 0.009, showing that despite the modulatory effect of the Condition factor, a significant 0.964 μV attenuation was still present.

The topographical comparison of the mean tone ERP and mean attenuation effect in the N1 range (Figure 5.5, top) showed no significant Signal × Electrode interaction: F[3,54] = 1.135, ε = 0.620, p = 0.330), providing no evidence for a different ERP component causing the ERP attenuation. In the P2 range (Figure 5.5, bottom), comparing the mean tone ERP with the difference of the attenuations showed no significant Signal × Electrode interaction: F[3,54] = 0.198, ε = 0.621, p = 0.806), providing no evidence for a different ERP component causing the modulatory effect of the Condition factor. Comparing the mean tone ERP and the attenuation effect in the No Tap condition in the P2 time interval, however, showed significant Signal × Electrode interaction: F[3,54] = 4.494, ε = 0.588, η_G^2 = 0.072, p = 0.023, indicating that the attenuation-effect involved ERP component(s) other than the tone-related P2.

Experiment 2 - Interim Discussion

As in previous studies utilizing button-press-contingent stimulation protocols, the estimated auditory N1 and P2 amplitudes were reduced for tones coinciding with (and also elicited by) the finger-movements in comparison to those elicited by tones only listened to. Conforming to the hypothesis, when finger-movements did not result in mechanical contact with an external object (No Tap condition), the magnitude of attenuation significantly decreased for P2, and there was a tendency for a similar N1 attenuation-decrease as well. Similarly to the results of Experiment 1, this suggests that mechanical impact played a substantial role in the observed ERP attenuation pattern. Similarly to Experiment 1, the observed action-related ERP attenuation was not exclusively related to the mechanical impact, because a significant N1 and P2 attenuation was still present in the No Tap condition. The topographical difference between the tone-related P2 and the attenuation effect in the P2 time range supports the notion that ERP attenuation effects measured in contingent paradigms reflect more, probably task dependent (Saupe et al., 2013) ERP effects, possibly superimposed on genuine N1 or P2 modulation.

Although participants were instructed to refrain from forcefully hitting the plate, tapping may have still caused faint, but audible transient sounds as evidenced by the intensity measurements presented in the Methods section of Experiment 1. That the N1 attenuation-decrease in the No Tap condition did not reach significance in contrast with Experiment 1, might have been caused by this instruction. Because of the potential presence of such transients, interpreting these effects is not trivial, as discussed in the General Discussion section.

The goal of the present study was to assess whether mechanical interaction resulting from the action played a role in action-related auditory (ERP) attenuation. The results of Experiment 1 and 2 suggest that mechanical impact does play a role, but due to the potentially confounding effect of transient sounds generated by tapping, this effect cannot be unequivocally attributed to attention. To eliminate this confound, in Experiment 3 participants maintained continuous contact with an object and applied force impulses from time-to-time (Press condition), or produced finger-movements without mechanical contact (No Tap condition identical to that in Experiment 1 and 2, see Figure 5.1). Because applying a force impulse did not result in displacement, no action-related sound was generated. It was hypothesized, as in Experiment 1 and 2, that the availability of tactile feedback (Press condition) would compel participants to allocate attention to the tactile stimulation as the action was performed, which would lead to a stronger action-related reduction of the auditory ERPs than in the No Tap condition, in which tactile feedback was not available.

Experiment 3 - Materials & Methods

Participants

21 young adults participated in Experiment 3, one of whom got tired during the experiment and could not perform the task, so the final sample included data from 20 participants (nine women, aged 18-28 years, mean 23 years, all right-handed) for monetary compensation. They reported normal hearing and no history of neurological disorders, and gave written informed consent after the experimental procedures were explained to them.

Stimuli and procedures

The experiment was very similar to Experiment 1. In the following, only the differing details are described. There was a No Tap condition identical to that in Experiments 1 and 2 (Figure 5.1, left column); and there was a Press condition, in which participants were instructed to put their index finger on a piezoelectric element and maintain contact throughout the experimental block (Figure 5.1, right column). The center of the piezoelectric element was marked with a small bump. They were instructed to apply short pressure impulses without moving their fingers once every 4 s. The pressure impulses did not produce sound. Each condition was administered in six blocks, each lasting 5 minutes. The blocks were presented in an interwoven ("PNNPPNN…", in which "P" denotes a block from the Press condition, and "N" denotes a block from the No Tap condition) order, with the starting condition counterbalanced between participants. As in Experiment 2, a curtain was hanged above the arm, between the participant's head and hand, which occluded the sight of the hand and the frame or piezoelectric element. The temporal adjustment (which was 300 ms in Experiment 1) was set to 500 ms. That is, tones scheduled to be delivered within 500 ms after an action were delivered immediately (an action-tone coincidence); tones scheduled to be delivered between 500 and 1000 ms were presented at 500 ms, tones scheduled to be delivered between 1000 and 1500 ms were presented at 1000 ms, and so on. This change in temporal adjustment would result in higher coincidence rates, and thus allow the recording of more coincidence epochs than in Experiment 1, which would result in enhanced signal-to-noise ratio, and more power to the detect attenuation- and attenuation-modulatory effects in the repeated-measures design. The ratio of coincidences would be lower than 10%, that is, coincidences would still be relatively rare.

EEG-recording and analysis

EEG recording and analysis was identical to that of Experiment 1, but in order to further enhance signal-to-noise ratio, eye-movement correction was applied on the basis of a two-minutes-long EEG recording (registered at the beginning of the session) featuring various eye-movements as described by Schlögl, Keinrath, Zimmermann, Scherer, Leeb, and Pfurtscheller (2007). To accommodate eye-movement correction into the processing chain, the EEG was off-line 1 Hz high pass filtered before eye-movement correction, and then 20 Hz low pass filtered. Also, because of the 500 ms adjustment window, tone ERPs were calculated as the ERP elicited by tones following an action by at least 1500 ms (and no action following within 400 ms).

Power calculations as described in the Methods section for Experiment 2 were performed to assess whether Experiment 3 had sufficient power to detect differences in N1 and P2 attenuation equaling the mean between-condition difference measured in Experiment 1. Attenuation-differences detectable with 95% power are also reported.

Because of its similarity to Experiment 1, further between-experiment comparisons were conducted separately for the N1 and P2 amplitudes measured in the Tap condition in Experiment 1 and the Press condition in Experiment 3, that is, amplitudes were analyzed in Experiment (Experiment 1 Tap vs. Experiment 3 Press, as between groups factor) × Event (corrected coincidence vs. tone, as repeated measures factor) ANOVAs. The main question in these analyses was whether an interaction would be found, which would suggest that although both tapping and applying more force involved a tactile change, differences in some other aspects of the two actions played a substantial role in the measured attenuations as well.

Because no significant N1 attenuation was found in the No Tap condition of Experiment 1, but a significant N1 attenuation was found in Experiment 3 (see below), between-experiment comparisons (Experiment × Event ANOVAs, as described above) were also conducted separately for the N1 and P2 amplitudes measured in the No Tap conditions of Experiment 1 and 3.

Experiment 3 - Results

Participants complied with the instruction. The mean between-action interval was 3808 ± 494 ms in the Press, and 3925 ± 417 ms in the No Tap condition (no significant difference: $t[19] = 1.140$, $p = 0.268$). The sound-coincidence rate was lower in the No Tap (8.7 ± 1.2 %) than in the Press (9.4 ± 1.1 %) condition: t[18] = 2.171, p = 0.043. The ratio of intervals excluded from the analyses due to potentially undetected actions did not differ between conditions (t[18] = 1.052, p = 0.306); this affected 3.1 ± 8.3 % of all between-action intervals.

The number of epochs contributing to coincidence ERPs was 35 ± 7 in the No Tap, and 39 ± 7 in the Press condition. The number of epochs contributing to the tone ERPs was 169 ± 31 in the No Tap and 176 ± 30 in the Press condition.

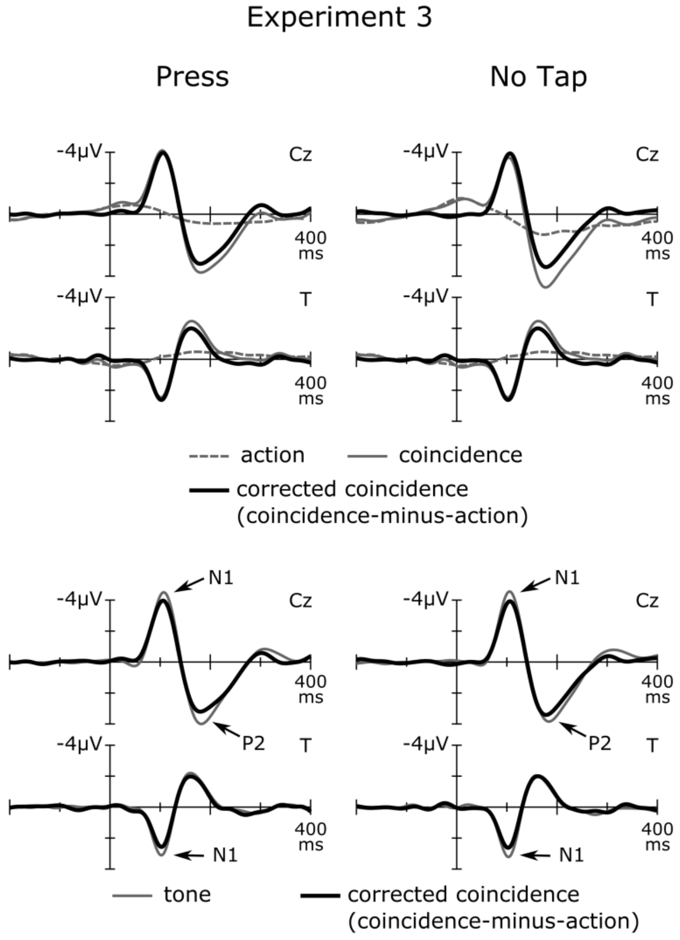

Figure 5.6. Group-average (N=20) ERP waveforms in Experiment 3 in the Press (left) and No Tap (right) conditions elicited by the respective actions, action-tone coincidences and the corresponding corrected coincidence waveforms (coincidence-minus-action difference waveforms; top panel). Tone-related ERPs and corrected coincidence waveforms are contrasted in the bottom panel. Because the N1 attenuation effect was also observable at the T3 and T4 electrodes (see Fig. 8) the average of these signals are shown in a separate row (marked by T) on both panels.

For tones (at least 1500 ms after an action), N1 peaked at 105 ms in the No Tap and at 107 ms in the Press condition, (both) at Cz (Figure 5.6, the amplitudes - average signal in the 96-116 ms range at Cz – were 4.330 ± 1.938, and 4.372 ± 1.866 µV in the Press and No Tap conditions, respectively). The ANOVA of the amplitudes showed a significant Event main effect only: F(1,19) = 9.985, $\eta_G^2 = 0.022$, p = 0.002 (for all other effects F < 1), indicating 0.557 µV lower (less negative) ERP amplitude for corrected coincidence than for tones. The effect size corresponding to the between-condition difference in attenuation measured in Experiment 1 (1.229 μV) was d = 1.426, and the power to detect such a difference was >0.999. The mean attenuation difference detectable with 95% power was 0.733 μV in this arrangement. The corrected coincidence and tone ERPs amplitudes were compared in both conditions, and significant differences were found: t(19) = 2.859, p = 0.010, attenuation of 0.602 μV; and t(19) = 2.683, p = 0.015, attenuation of 0.511 μV, respectively in the No Tap and Press conditions.

For tones, P2 peaked at 185 ms at FCz in the No Tap, and at 182 ms at Cz in the Press condition (the amplitudes - average signal in the 174-194 ms range at Cz - were 3.925 ± 2.120, and 3.776 ± 2.156 μV in the Press and No Tap conditions, respectively). The ANOVA of the amplitudes showed a significant Event main effect only: F(1,19) = 6.504, $\eta_G^2 = 0.020$, p = 0.026 (for all other effects F < 1), indicating 0.618 μV lower (less positive) ERP amplitude for corrected coincidences than for tones. The effect size corresponding to the between-condition attenuation-difference in Experiment 1 (1.708 μV) was d = 1.594, and the power to detect such a difference was >0.999. The mean attenuation difference detectable with 95% power was 0.911 μV in this arrangement. The corrected coincidence and tone ERP amplitudes were compared in both conditions. A significant difference was only found in the Press condition (t[19] = 3.250, p = 0.004, attenuation of 0.777 μV; in the No Tap condition: t[19] = 1.540, p = 0.140, attenuation of 0.460 μV).

The topographical comparison of the mean tone ERPs and mean attenuation effects showed no significant Signal × Electrode interactions (N1: F[3,57] = 1.140, ε = 0.561, p = 0.324; P2: F[3,57] = 0.962, ε = 0.527, p = 0.375) providing no evidence for a different ERP component causing the attenuation effects of the Event factor. An unexpected finding observable at the topography of the attenuation effects (Figure 5.7) is the attenuation of the positive aspect of the N1 at the temporal (T3/T4) sites (the mean amplitudes were 2.866 ± 1.571; and 2.994 ± 1.353 μV, in the Press and No Tap conditions, respectively). The post-hoc Condition (No Tap vs. Press) × Event (corrected coincidence vs. tone) × Electrode Side (T3 or T4) repeated-measures ANOVA showed a significant main effect of Event only: F(1,19) = 13.998, $\eta_G^2 = 0.025$, p = 0.001 (for all other effects F < 1), indicating a 0.538 μV lower (less positive) ERP amplitude for corrected coincidences than for tones. The topographical comparison of the mean tone- and mean corrected-coincidence ERPs using the T3, T4, TP7, TP8, and the signals from the left and right mastoids (Lm and Rm respectively) showed significant Signal × Electrode interaction: F(5,95) =3.297, ε = 0.555, $\eta_G^2 = 0.037$, p = 0.023, indicating that the tone N1 and the attenuation-effect topographies differed in their shapes.

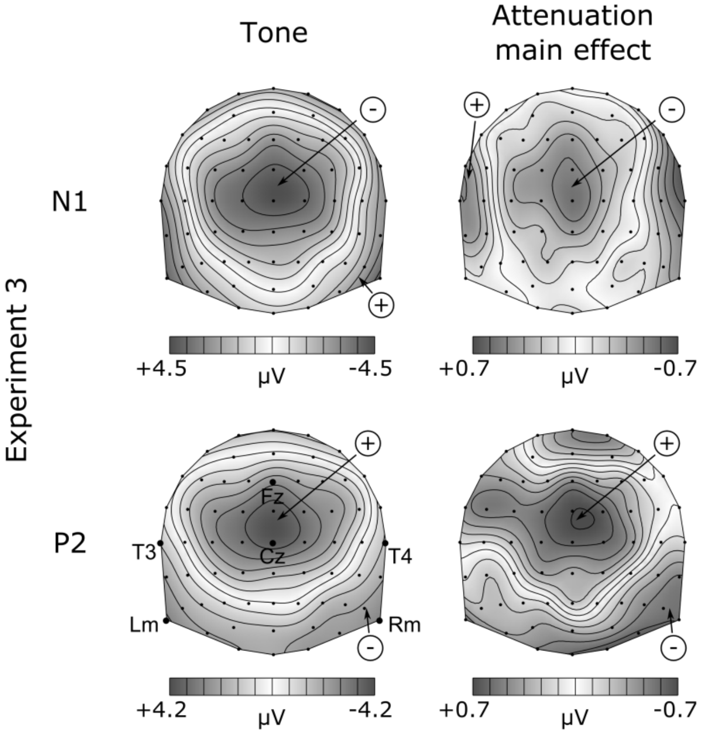

Figure 5.7. Group-average (N=20) topographical distributions of the N1 (top row) and P2 (bottom row) in Experiment 3 (average signals in 20 ms long windows centered at the tone group-average peak latencies). The left column shows topographies of ERPs elicited by tones (averaged across conditions). The right column shows the attenuation main effects (tone-minus-corrected coincidence waveform averaged across conditions). Signal ranges differ between the individual topographical distributions in order to emphasize similarities or differences in shape. Positive and negative polarity areas are indicated by arrows.

The Experiment (Experiment 1 Tap vs. Experiment 3 Press) × Event (corrected coincidence vs. tone) ANOVA for the N1 amplitudes showed a significant Event main effect only: $F(1,37) = 17.384$, $\eta_G^2 = 0.033$, $p < 0.001$, showing lower (less negative) ERP amplitude for corrected coincidences than for tones (Experiment main effect: $F<1$; Experiment × Event interaction: $F[1,37] = 1.860$, $\eta_G^2 = 0.004$, $p = 0.181$).

For the P2 amplitudes the same type of ANOVA yielded a significant Event main effect: $F(1,37) = 40.309$, $\eta_G^2 = 0.136$, $p < 0.001$, and an Experiment × Event interaction: $F(1,37) = 11.854$, $\eta_G^2 = 0.044$, $p = 0.001$, indicating a stronger P2 attenuation in the Tap condition of Experiment 1 than in the Press condition of Experiment 3 (the Experiment main effect was not significant: $F[1,37] = 1.799$, $\eta_G^2 = 0.040$, $p = 0.188$).

The Experiment (Experiment 1 No Tap vs. Experiment 3 No Tap) × Event (corrected coincidence vs. tone) ANOVA for the N1 amplitudes showed only a significant interaction: $F(1,37) = 4.682$, $\eta_G^2 = 0.009$, $p = 0.037$, showing that N1-attenuation was stronger in Experiment 3 than in Experiment 1 (Experiment main effect: $F<1$; Event main effect: $F[1,37] = 1.077$, $\eta_G^2 = 0.002$, $p = 0.306$).

For the P2 amplitudes the same type of ANOVA yielded a significant Event main effect only: $F(1,37) = 4.878$, $\eta_G^2 = 0.018$, $p = 0.012$, showing lower (less positive) P2 amplitude for corrected coincidences than for tones ($F<1$ for both the Experiment main effect, and the interaction).

Experiment 3 - Interim Discussion

The estimated auditory N1 and P2 amplitudes were reduced for tones coinciding with finger-movements or pressure impulses in comparison to tones following such actions by at least 1500 ms. No significant between-condition difference in the action-related attenuations were found, despite the increased power resulting from the introduction of the eye-movement correction and the stronger temporal adjustment. This result does not support the hypothesis that the action-related attenuation of auditory ERPs is brought about by allocating attention to the tactile feedback related to performing the action. The comparison of the N1 and P2 attenuations observed in the Tap condition of Experiment 1 and the Press condition of Experiment 3 showed that P2 attenuation was significantly stronger in the Tap condition of Experiment 1. The comparison of the N1 and P2 attenuations in the No Tap conditions of Experiment 1 and 3 showed that N1-attenuation was stronger in Experiment 3.

General Discussion

The goal of the present study was to explore the potential role of mechanical interaction in the action-related auditory ERP attenuation. It was hypothesized that having a reliable mechanical interaction with an external object as part of the given action compelled participants to allocate attention to the tactile feedback as the action was performed, which would draw processing resources away from the simultaneously presented auditory stimuli, and lead to attenuated N1 and P2 auditory ERPs. The contribution of the auditory ERP to the action-tone coincidence-related ERP was estimated by subtracting the action-related ERP registered when the action did not co-occur with the tone.

In Experiments 1 (coincidence paradigm) and 2 (contingent paradigm), mechanical contact was established by tapping on an external object at the endpoint of a finger-movement. Both the coincidence and the contingent arrangements replicated the results of previous studies: auditory processing activity was attenuated when the tones coincided with a finger-movement. Moreover, this attenuation as reflected by the auditory N1 and P2 ERPs was reduced when finger-movements did not result in mechanical contact with an external object in Experiment 1, and a reduced P2 attenuation (and a hint at a reduced N1 attenuation) was obtained in Experiment 2. These results clearly show that mechanical interaction played a substantial role in the finger-movement-related auditory ERP attenuation reported in the literature. Because mechanical impact inevitably causes faint, but audible transient sounds (see below for a detailed discussion), in Experiment 3 participants continuously maintained contact with the external object, and applied pressure impulses from time to time, which did not result in displacement and transient sounds. Although auditory ERPs were attenuated as in Experiment 1 and 2, applying pressure impulses did not result in stronger N1 or P2 attenuations in comparison to the condition in which no mechanical contact was made.

Whereas the pattern of results shows that mechanical impact, but not the mere presence of mechanical contact played an important role in finger-action-related auditory ERP attenuation, connecting this technical description to the hypothesis motivating the present study, that is, that auditory ERP attenuation was caused by shifting attention to the action-contingent tactile stimulation, is not trivial. Indeed, as discussed in the following, the manipulation of action-contingent mechanical interaction in the present paradigm allows for a number of interpretations regarding the cause of this pattern.

The first two interpretations are based on the fact that movements resulting in mechanical impact lead to the elicitation of potentially audible transient sounds. As evidenced by the measurements described in the Methods of Experiment 1, tapping may have produced faint, but probably audible sounds.

(1) The first interpretation suggests that the ERP attenuation effects are ERP artifacts stemming from this methodological shortcoming. One of the basic difficulties in measuring action-related auditory ERP attenuation is to estimate the auditory ERP contribution within the action-tone coincidence ERP waveform. The most widely used approach – which was used in the present study as well - is to subtract an “action-only” waveform from the coincidence waveform. Whereas this is a plausible approach, it may cause problems if the actions generate auditory impact-sound transients, because the “action only” waveform will actually be a superposition of the motor- and the impact-sound-ERP (for clarity, in the following *probe sound* refers to the tone presented by the experimenter, and *impact-sound* refers to the faint sound produced by the mechanical interaction with the device). When the action coincides with the probe sound, the contribution of the impact-sound to the ERP may be negligible due to masking, because the probe is much louder. In contrast, when the impact-sounds occur during periods of silence (far away from the probe sounds, or in a different experimental block), low-amplitude auditory ERPs may be elicited by the unmasked impact-sound (superimposed on the motor ERP). When estimating the auditory contribution to the coincidence ERP, the subtraction of this low-amplitude impact-sound ERP may artificially decrease the auditory ERP amplitude for the probe. Note that the presence or absence of the superimposed impact-sound related ERP to the motor ERP might not be visually salient in the ERP waveform at all. Obviously, this suggests that experiments in which actions produce faint sounds because of the nature of interaction with the response-device (e.g. tapping, or button clicks) overestimate the magnitude of auditory ERP attenuation. The finding that the estimated P2 attenuation was stronger in the Tap condition of Experiment 1 than in the Press condition of Experiment 3 fits this interpretation well (but note that no such difference was found for the N1). Note that whereas some studies (e.g. SanMiguel, Todd, & Schröger, 2013) used noise-attenuation techniques to counteract this type of confound, no study reported direct measurements whether these

were in fact effective or not (the present study being no exception: the effectiveness of the instruction to refrain from forcefully hitting the plate in Experiment 2 was not formally assessed).

Certain stimulation arrangements may be suitable for the elimination of the ERP artifact described above. For example, in contingent paradigms, introducing a delay between the action and the probe sound may allow one to temporally separate the probe-related ERP from the impact-related waveform. In such arrangements, however, one would have to control for refractoriness effects, which are known to affect the N1 (Näätänen & Picton, 1987). That is, the probe-related N1 might be elicited with lower amplitude in the Motor-Auditory arrangement than in the Auditory one because of the refractoriness caused by the impact sound.

(2) The second interpretation suggests that although the results do not reflect an ERP artifact as described above, they reflect prediction- or attention-based "filtering" related to the production of the impact-sound. This hypothesis suggests that action-related auditory attenuation (which may not only be reflected in ERP, but other measures as well) is "genuine" in the sense that it reflects a preparation for the *reliable* auditory consequence of the action, that is, the impact-sound. If there is a faint but audible sound produced by the mechanical interaction with the response device, then in the typical contingent or coincidence-based experimental designs, this is the most reliable auditory consequence of the action. In terms of the internal forward modeling framework the auditory attenuation would reflect an internal forward model predicting this action-contingent auditory stimulation, which would also lead to the attenuation of the probe tones. This would also readily explain why auditory attenuation occurs for probe-action coincidences in the absence of a contingent action-probe relationship (but see below). In terms of the attention-based framework, it could be speculated that participants adopt an attention set which is optimal for detection of the impact-sound (even against the background of other sounds), which leads to attenuated activity when a different sound – the probe – also occurs.

The most important methodological implication of these two lines of thought is that controlling the auditory consequences of the actions must be adopted as a standard procedure in research on action-related auditory attenuation. Whereas one may attempt to acoustically isolate the participant from the action-induced auditory transients, controlling whether the isolation of these extremely short, wide-spectrum sounds was successful might require extensive acoustical measurements specific to the given settings. It seems plausible that the best practice involves the use of devices designed for silent operation (like the infrared light beam- or piezoelectric element-based arrangements used in the present study).

The next two speculations suggest that despite the potential confounds, the observed ERP-modulations nonetheless reflect attention- or forward-model-based cognitive processes. That is, although one *could* cause impact-sounds with the given action (i.e. potentially hit the given response-device with a momentum sufficient to produce an audible sound), typically this does not happen, and the measured attenuation-differences reflect other, genuine cognitive effects.

(3) One may speculate (post-hoc) that although tapping (Experiment 1 and 2) and applying pressure (Experiment 3) on a surface both result in tactile changes related to the performance of the action, the information provided by these sensory changes is markedly different in the two contexts . Whereas tapping establishes a new mechanical contact and thereby allows one to confirm the presence of the object and its position in relation to the hand (which was essential for the performing the task in Experiment 1 and 2), changing the force applied to a continuously touched object does not provide more information on these task-relevant aspects of the situation. That is, tapping may compel participants to shift attention to the tactile feedback when the action is performed, because that is how the optimal interaction with the device can be monitored and maintained. In contrast, applying more force to the already touched rigid object does not allow for substantial information gain on the circumstances of the interaction. In different words, making contact with an object strongly reduces uncertainties regarding the (potential) interaction, whereas applying more pressure on an already touched rigid object does not. In similar vein, one may speculate that there might be a movement-, or movement-planning-related difference between moving the index finger in the air and tapping with it on a surface, which may result in the observed ERP attenuation patterns.

(4) A further speculative account suggests that although no audible transient sounds are generated on most trials, participants might still be aware of the possibility (or opportunity) to generate such sounds, and this “awareness” might be sufficient to produce a generalized expectation (see also Desantis, Hughes, & Waszak, 2012; Hughes, Desantis, & Waszak, 2013a). By attention- or forward model-based mechanisms, this expectation may activate sensory representations of the producible sounds, which results in stronger attenuation than in arrangements in which the action (pressure application, Press condition of Experiment 3) or the lack of object (No Tap conditions) does not warrant such an expectation.

Each of these interpretations is compatible with the pattern of results, still, these lines of thought have some plausible methodological implications, and offer some ideas for future research. Although the methodological implications of the results are severe, they do not invalidate previous research, because they also demonstrate that the ERP attenuation effects reported in the literature are not (entirely) due to the action-related mechanical impact-sound confound. Actions with silent response devices resulted in ERP attenuations: The contingent arrangement in Experiment 2 resulted in ERP attenuations in the N1 and P2 time range (although the different topography of the attenuation effect in the P2 range suggested that the P2 effect was not a pure P2-modulation). The coincidence paradigm in Experiment 3 showed significant N1 and P2-attenuations , and a significant P2-attenuation was found in Experiment 1. Interestingly, the between-experiment comparison showed that N1-attenuation was stronger in the No Tap condition of Experiment 3 than that of Experiment 1. Although this might be due to the better signal-to-noise ratio in Experiment 3 (brought about by the longer adjustment window and the eye-movement correction), or the difference in coincidence rate (5.7 vs. 8.7 %), one may also speculate that this difference may be related to the different contexts provided by the other conditions of the experiments. Although some evidence suggests that context may influence action-related ERP attenuation (i.e. Baess et al., 2011 found that N1-attenuation was stronger when the tones were self-induced among external tones), at this point it is not clear which aspect of the other conditions could cause such a modulation-difference.

Although the present study mainly focused at the fronto-central aspect of the N1 ERP, it has to be noted that the N1 waveform elicited by tones is not a unitary phenomenon, but the superposition of a number of sub-components (for an overview, see Näätanen & Picton, 1987). Unfortunately, only few studies speculated on the component-specificity of action-related N1-attenuation (Horváth et al., 2012, SanMiguel et al., 2013, Saupe et al., 2013), and a systematic investigation on sub-component attenuations has not been conducted yet. Studies utilizing the coincidence paradigm (e.g. Horváth et al., 2012) did not report significant N1 attenuation-effects at electrode sites where the N1 polarity is typically reversed for pure tones (that is, below the Sylvan fissure with a maximum at the mastoids when the EEG is recorded with nose-, or average reference, suggesting a supra-temporal source, Vaughan & Ritter, 1970). The lack of such a polarity reversal also observed in some studies using contingent stimulation arrangements was interpreted as a sign that action-related attenuation might mainly (but not exclusively) affect the non-specific N1 subcomponent (SanMiguel et al., 2013), which shows no such reversal. Interestingly and unexpectedly, in Experiment 3 of the present study, the N1-attenuation did show such a reversal. The effect, however, was not maximal at the mastoids, but at the T3 and T4 sites as evidenced by post-hoc analyses. This reversed attenuation effect therefore might not reflect the attenuation of the supra-temporal N1, but the attenuation of the positive Ta subcomponent of the so-called T-complex (Wolpaw & Penry, 1975). Whereas action-related attenuation of the Tb subcomponent (a negative peak observable typically at around 140 ms at temporal sites) have been observed (Horváth, 2013a; SanMiguel, et al., 2013; Saupe, et al., 2013), the results of Experiment 3 suggest that Ta might also be attenuated. Naturally, being a post-hoc finding, this interpretation should be handled with caution.

In summary, the present results demonstrate that using actions potentially resulting in mechanical impact sounds may bias the measurement of action-related auditory (ERP) attenuation. In the worst case scenario, the measured ERP attenuations may be artifacts resulting from the improper estimation of the auditory contribution to action-sound coincidence ERPs. Even in the absence of a direct ERP confound, measured action-related sensory attenuations (measured by ERPs or other methods) may reflect processing related to the impact sound instead of the probe sound presented by the experimenter. Importantly, however, the present study also demonstrated that when such confounds were eliminated by the use of silent response devices, tones coinciding with actions still elicited attenuated auditory ERPs .

Overview of the results

The most important result of the presented studies is that the action-related attenuation of the auditory N1 ERP was consistently present in coincidence arrangements: attenuations of the late auditory ERPs were found in all eight experiments administering this paradigm. N1 attenuation was consistently followed by the attenuation of the P2, but Study I suggested that the two components were nonetheless separable. These results demonstrate that auditory N1 and P2 attenuation occurs even in the absence of actual action-sound contingency. Although demonstrating the coincidence-effect is no direct evidence against the interpretations suggesting that action-related N1 attenuation is related to action-sound contingency representations, the coincidence paradigm provides a baseline condition for future studies on (additional) contingency-related attenuations.

The results of Study I allowed the formulation of a number of more economic, novel hypotheses regarding the cause of the attenuation effect. Testing these basic hypotheses provided a groundwork that was missing for the more elaborate hypotheses. The results of Study III are compatible with the notion that auditory ERP attenuation reflects central processes and not the co-activation of the middle-ear stapedius muscle with the tone-inducing movement. Although it is difficult to reject the hypothesis that action-related N1 attenuation reflects a dynamic, action-related change in the allocation of attention towards action-related activities, the results of Study II were compatible with the notion that well-known auditory selective attention effects did not substantially contribute to N1 attenuation. Timm et al., (2013), following up on this possibility in a contingent arrangement, reached similar conclusions. In a study aiming to uncover possible attention-related confounds in the context of the contingent paradigm, Saupe et al. (2013) showed that the although the contingent paradigm might be biased by between-condition differences in the allocation of attention, such differences can be dissociated by the affected auditory ERP subcomponents: whereas task-related attentional differences affected the vertex N1 waveform, Tb was only modulated by the coincidecence with an action.

The exploration of these hypotheses lead to a more detailed characterization of action-related auditory attenuation phenomena. The attenuation of the T-complex reported by others (Saupe et al., 2013; SanMiguel, Todd & Schröger, 2013) in contingent arrangements was also observed in some of the present studies in a coincidence arrangement. The pattern of attenuation was however, not consistent across the experiments: studies reporting T-complex attenuations reported mostly Tb-attenuation, but in Study V, a Ta-attenuation was observed, and no T-complex modulation was found (in retrospect) in the initial experiment of Study I. At this point, this pattern of results is to be interpreted with care, because this effect may also be brought about by the regularity of the action-sequence. Since the T-complex originates from the secondary auditory cortices (parabelt areas, Ponton et al., 2002), which are interconnected with the adjacent areas of temporal and parietal lobes, as well as the frontal lobe (Kaas & Hackett, 2000), a decrease in this component may fit hypotheses suggesting an interaction between auditory and other non-auditory subsystems.

Although the results of the MEG experiment in Study I suggest that the supratemporal N1 subcomponent is attenuated when tones coincide with a finger-movement based action, the attenuation-effect did not show a polarity reversal across the Sylvan-fissure (at the mastoids) when the EEG was recorded with nose reference. This suggests that the observed ERP attenuation effect might be dominated by the attenuation of the non-specific N1 subcomponent. It is important to note that in Studies I, II, and IV the coincidence-related N1 attenuation effect showed a posterior topography, which differs from the mostly fronto-central or central effects observed in contingent paradigms (for a typical attenuation effect in a contingent paradigm, see Study III). In Study II, the coincidence-related N1 attenuation was not substantially modulated by the multiplicity of concurrently presented tone frequencies. In contrast, as similar manipulation in the study by Baess et al. (2008) resulted in increased attenuation when the actions resulted in a single tone frequency in contrast to those resulting in random tone frequencies in a contingent arrangement. This, taken together with the topographical analyses suggesting that the N1-attenuation effect involves mainly the attenuation of the non-specific N1 subcomponent, hints at the possibility that the coincidence-effect is not stimulus feature specific.

A tentative framework for the interpretation of action-related auditory attenuation

Although the last five years have seen a burst of interest in action-related auditory attenuation, most current studies follow parallel lines of thought, and few studies go beyond the demonstration of action-related ERP attenuation in a given paradigm. Studies demonstrating successful modulations of the basic attenuation-effects are rare, as well as experiments directly comparing predictions based on the different hypotheses (with the exception of studies on the attention-based explanation of the attenuation-effect). A number of results in our own studies and others are null-effects, therefore, an integration of the available evidence is necessarily speculative, and a general framework integrating the currently available results can only be tentative. A number of assumptions used in the following rely on the overview of the parameters and manipulations used in the reported experiments, and assume that the choice of these manipulations and parameters are not random, but reflect those parameters and manipulations, which were successfully used in demonstrations of these effects. Naturally, these parameters and manipulations may be used simply because they are convenient (see Study III, in which the spectral properties of the sounds used in successful demonstrations of auditory attenuation were compatible with an alternative account for the effect, but in the end, this turned out to be simply a matter of convenience).

The coincidence-effect is difficult to reconcile with the fundamental assumption that N1 attenuation reported for non-speech actions (finger-movements and key-presses) reflects functions relying on action-sound contingency-representations, or specifically, internal forward models. In the following, I put forward a tentative hypothesis suggesting that the presently available evidence can be best explained by a two-component model. I hypothesize that the results obtained in the *coincidence paradigm* show the effects of “genuine” forward modeling, whereas experiments utilizing *contingent stimulation protocols* reflect additional preparatory attention effects, or preactivation (Roussel, et al., 2013) related to the cognitive representation of actions in the given task setting.

Both of these mechanisms rely on action-effect contingency representations, which differ in a number of features. The most important difference is their characteristic adaptation time, that is, the time needed to form an action-effect representation, or to accommodate an existing representation to a changed action-effect contingency. It is suggested that forming or changing and internal forward model takes relatively long time, whereas preactivation-related action-effect representations are formed and adapted rapidly.

The observation this line of thought relies on is the following: Despite the theoretical generalization of “action-related” auditory attenuation, the actions used to demonstrate these effects in a non-speech arrangement are rather similar: they are all finger-movements. We perform such finger-movements to press buttons, tap on objects in everyday life countless times to achieve our goals. I hypothesize that the actions performed in these experiments are already coupled to various, well-known sensory effects, which are represented by internal forward models predicting the occurrence of transient sensory (auditory) events when the finger movement is initiated. This prediction is rather general: it predicts the occurrence of the auditory stimulation without specifying its exact features.

The idea of separating the two mechanisms by their characteristic times originates from motor learning. Smith, Ghazizadeh, & Shadmehr (2006) demonstrated that motor adaptation patterns for reaching movements in externally disturbed force-fields can be described by the interaction of two processes: one, which responds rapidly to changes of the action-effect (force-displacement) contingency, but does not retain such information for long, and another, which, although does not follow changes as rapidly, but retains information on contingencies for longer periods of time. Interestingly, Joiner & Smith (2008) demonstrated that long-term (24 hour) retention of the learned contingency representation was determined by the level of adaptation reached by the slowly adapting process.

In the present context, it is assumed that the durations of our experimental sessions allow for different effects to emerge. It is assumed that finger-movement-sound *coincidence* paradigms actually reflect the workings of a “generalized” forward model which associates finger movements (button pressing) with a sound transient, which is activated whenever the participant presses a button. It is assumed that this forward model is not formed during the experiment, rather it is built up by countless button-presses and taps leading to various auditory transients in everyday life. Although button-presses do not produce sounds most of the time during a coincidence experiment, due to the slow adaptation of the forward model, the duration of the experiment is too short to result in substantial deactivation.

In comparison, paradigms administering contingent action-sound stimulation protocols are long enough to allow the buildup of a specific action-effect association underlying sensory preactivation in addition to the already present “generalized” key-press-sound-transient forward model. Because this specific contingency representation is built up rapidly, such paradigms may yield feature-specific effects as well. That is, under typical experiment durations, it is expectable that protocols featuring action-sound contingencies yield patterns of stimulus-specific attenuation, whereas paradigms investigating action-sound coincidences do not. Note that it is not suggested that only contingent paradigms can yield results reflecting stimulus-specific predictions, rather that key-presses are probably not the best candidates for making specific forward model based predictions visible. Long-term training with systematic action-effect associations may lead to the formation of robust, cognitively impenetrable and specific action-effect representations, presumably internal forward models. For example, Repp & Knoblich (2007) showed that performing patterns of finger-movements which would result in ascending or descending tone pairs on a piano induced a corresponding bias in the perception of an ambiguous pitch change for pianists but failed to induce a bias in non-pianists.

This hypothesis also suggests that speech-production-based experiments may reflect the effects of both “genuine” forward modeling and preactivation. It seems plausible that internal forward models for speech-production could be more easily adjusted than for other actions, because there is a strong coupling (close, one-to-one correspondence) between the action- and effect-parameters. Obviously, future studies should explore possibilities regarding novel actions and the effects of long-term training, and these may give strong empirical support to the notion that the sensory effects resulting from speech-production are similar to the effects resulting from other actions.

References

Alho, K., Paavilainen, P., Reinikainen, K., Sams, M., & Näätänen, R. (1986). Separability of different negative components of the event-related potential associated with auditory stimulus processing. *Psychophysiology, 23(6)*, 613–623. doi:10.1111/j.1469-8986.1986.tb00680.x

Alho, K., Rinne, T., Herron, T. J., & Woods, D. L. (2014). Stimulus-dependent activations and attention-related modulations in the auditory cortex: A meta-analysis of fMRI studies. *Hearing Research, 307,* 29–41. doi:10.1016/j.heares.2013.08.001

Aliu, S. O., Houde, J.F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience, 21*, 791-802. doi: 10.1162/jocn.2009.21055

Ariff, G., Donchin, O., Nanayakkara, T. & Shadmehr, R. (2002). A real-time state predictor in motor control: study of saccadic eye movements during unseen reaching movements. *Journal of Neuroscience, 22*, 7721–9

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. *International Journal of Psychophysiology, 70*, 137-143. doi: 10.1016/j.ijpsycho.2008.06.005

Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology, 48(9)*, 1276–1283. doi:10.1111/j.1469-8986.2011.01196.x

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. *European Journal of Neuroscience, 29*, 1514-1521. doi:10.1111/j.1460-9568.2009.06683.x

Bak, C. K., Lebech, J., & Saermark, K. (1985). Dependence of the auditory evoked magnetic field (100 msec signal) of the human brain on the intensity of the stimulus. *Electroencephalography and Clinical Neurophysiology*, *61*(2), 141–149.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37(3)*, 379–384. doi:10.3758/BF03192707

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology, 83(2)*, 120–131.

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

Blakemore, S. J., Wolpert, D.M., & Frith, C.D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience, 1*, 635-40.

Borg, E., & Zakrisson, J. E. (1974). Stapedius reflex and monaural masking. *Acta oto-laryngologica, 78(3-4)*, 155–161.

Brown, H., Friston, K., & Bestmann, S. (2011). Active Inference, Attention, and Motor Preparation. *Frontiers in Psychology, 2*. doi:10.3389/fpsyg.2011.00218

Butler, R. A. (1972). Frequency specificity of the auditory evoked response to simultaneously and successively presented stimuli. *Electroencephalography and Clinical Neurophysiology*, *33*(3), 277–282. doi:10.1016/0013-4694(72)90154-X

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* Hillsdale, N.J: L. Erlbaum Associates.

Colzato, L. S., de Bruijn, E. R. A., & Hommel, B. (2012). Up to "Me" or Up to "Us"? The Impact of Self-Construal Priming on Cognitive Self-Other Integration. *Frontiers in Psychology, 3,* 341. doi:10.3389/fpsyg.2012.00341

Colzato, L. S., Zech, H., Hommel, B., Verdonschot, R., Wildenberg, W. P. M., & Hsieh, S. (2012). Loving-kindness brings loving-kindness: The impact of Buddhism on cognitive self–other integration. *Psychonomic Bulletin & Review, 19(3)*, 541–545. doi:10.3758/s13423-012-0241-y

Crapse, T. B., & Sommer, M. A. (2008a). Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience, 9(8)*, 587–600. doi:10.1038/nrn2457

Crapse, T. B., & Sommer, M. A. (2008b). Corollary discharge circuits in the primate brain. *Current Opinion in Neurobiology, 18(6)*, 552–557. doi:10.1016/j.conb.2008.09.017

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clinical Neurophysiology, 115(4)*, 732–744. doi:10.1016/j.clinph.2003.11.021

Cullen, K.E. (2004). Sensory signals during active versus passive movement. *Current Opinion in Neurobiology, 14*, 698–706. doi: 10.1016/j.conb.2004.10.002

Cullen, K. E., Brooks, J. X., & Sadeghi, S. G. (2009). How actions alter sensory processing: reafference in the vestibular system. *Annals of the New York Academy of Sciences, 1164,* 29–36. doi:10.1111/j.1749-6632.2009.03866.x

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping, 9*, 183-91. doi: 10.1002/(SICI)1097-0193(200004)9:4<183::AID-HBM1>3.0.CO;2-Z

Davidson, P. R., & Wolpert, D. M. (2005). Widespread access to predictive models in the motor system: a short review. *Journal of Neural Engineering, 2(3)*, S313–S319. doi:10.1088/1741-2560/2/3/S11

Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N., Hobart G. (1939) Electrical reactions of the human brain to auditory stimulation during sleep. *Journal of Neurophysiology 2*, 500-514

Davis, H., Mast, T., Yoshie, N., & Zerlin, S. (1966). The slow response of the human cortex to auditory stimuli: recovery process. *Electroencephalography and Clinical Neurophysiology*, *21*(2), 105–113.

Davis, H., & Zerlin, S. (1966). Acoustic Relations of the Human Vertex Potential. *The Journal of the Acoustical Society of America*, *39*(1), 109. doi:10.1121/1.1909858

Davis, P. A. (1939). Effects of acoustic stimuli on the waking human brain. *Journal of Neurophysiology*, *2*, 494–499.

Desantis, A., Hughes, G., & Waszak, F. (2012). Intentional binding is driven by the mere presence of an action and not by motor prediction. *PLoS ONE, 7(1)*, e29557. doi:10.1371/journal.pone.0029557

Desantis, A., Weiss, C., Schütz-Bosbach, S., & Waszak, F. (2012). Believing and perceiving: Authorship belief modulates sensory attenuation. *PLoS ONE, 7(5)*, e37959. doi:10.1371/journal.pone.0037959

Dimitrijevic, A., Michalewski, H. J., Zeng, F.-G., Pratt, H., & Starr, A. (2008). Frequency changes in a continuous tone: Auditory cortical potentials. *Clinical Neurophysiology, 119(9)*, 2111–2124. doi:10.1016/j.clinph.2008.06.002

Dolk, T., Hommel, B., Prinz, W., & Liepelt, R. (2013). The (not so) social Simon effect: A referential coding account. *Journal of Experimental Psychology: Human Perception and Performance, 39(5)*, 1248–1260. doi:10.1037/a0031031

Duhamel, J.R., Colby, C.L., Goldberg, M.E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science, 255*, 90–92

Elberling, C., Bak, C., Kofoed, B., Lebech, J., & Særmark, K. (1980). Magnetic auditory responses from the human brain. A preliminary report. *Scandinavian Audiology, 9*(3), 185–190. doi:10.3109/01050398009076353

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance, 27(1)*, 229.

Elsner, B., & Hommel, B. (2004). Contiguity and contingency in action-effect learning. *Psychological Research, 68*,138-54.

Flinker, A., Chang, E. F., Kirsch, H. E., Barbaro, N. M., Crone, N. E., & Knight, R. T. (2010). Single-trial speech suppression of auditory cortex activity in humans. *Journal of Neuroscience, 30(49)*, 16643–16650. doi:10.1523/JNEUROSCI.1809-10.2010

Ford, J. M., Gray, M., Faustman, W. O., Roach, B. J., & Mathalon, D. H. (2007). Dissecting corollary discharge dysfunction in schizophrenia. *Psychophysiology, 44(4)*, 522–529. doi:10.1111/j.1469-8986.2007.00533.x

Ford, J.M., & Mathalon, D.H. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of Psychiatric Research, 38*, 37-46. doi: 10.1016/S0022-3956(03)00095-5

Ford, J. M., & Mathalon, D. H. (2012). Anticipating the future: Automatic prediction failures in schizophrenia. *International Journal of Psychophysiology, 83(2)*, 232–239. doi:10.1016/j.ijpsycho.2011.09.004

Ford, J. M., Mathalon, D. H., Kalba, S., Whitfield, S., Faustman, W.O., Roth, W.T. (2001). Cortical responsiveness during talking and listening in schizophrenia: an event-related brain potential study. *Biological Psychiatry, 50*, 540-549. doi: 10.1016/S0006-3223(01)01166-0

Ford, J. M., Palzes, V. A., Roach, B. J., & Mathalon, D. H. (in press). Did I do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone. *Schizophrenia Bulletin*. doi:10.1093/schbul/sbt072

Fruhstorfer, H. (1971). Habituation and dishabituation of the human vertex response. *Electroencephalography and Clinical Neurophysiology*, *30*(4), 306–312.

Fruhstorfer, H., Soveri, P., & Järvilehto, T. (1970). Short-term habituation of the auditory evoked response in man. *Electroencephalography and Clinical Neurophysiology*, *28*(2), 153–161.

Galazky, I., Schütze, H., Noesselt, T., Hopf, J.-M., Heinze, H.-J., & Schoenfeld, M. A. (2009). Attention to somatosensory events is directly linked to the preparation for action. *Journal of the Neurological Sciences, 279*(1-2), 93–98. doi:10.1016/j.jns.2008.12.006

Geisler, C. D., Frishkopf, L. S., & Rosenblith, W. A. (1958). Extracranial responses to acoustic clicks in man. *Science*, *128*(3333), 1210–1211. doi:10.1126/science.128.3333.1210

Gentili, R., Cahouet, V., Ballay, Y. & Papaxanthis, C. (2004). Inertial properties of the arm are accurately predicted during motor imagery. *Behavioral Brain Research, 155*, 231–9

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston, Houghton Mifflin.

Glasberg, B. R. & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47*, 103-138. doi: 10.1016/0378-5955(90)90170-T

Hansen JC, Hillyard SA (1980): Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology, 49,* 277–290. doi:10.1016/0013-4694(80)90222-9.

Hari, R., Aittoniemi, K., Järvinen, M.-L., Katila, T., & Varpula, T. (1980). Auditory evoked transient and sustained magnetic fields of the human brain localization of neural generators. *Experimental Brain Research, 40*, 237–240.

Hari, R., Kaila, K., Katila, T., Tuomisto, T., & Varpula, T. (1982). Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: implications for their neural generation. *Electroencephalography and Clinical Neurophysiology, 54(5)*, 561–569.

Hazemann, P., Audin, G., & Lille, F. (1975). Effect of voluntary self-paced movements upon auditory and somatosensory evoked potentials in man. *Electroencephalography and Clinical Neurophysiology, 39(3)*, 247–254. doi:10.1016/0013-4694(75)90146-7

Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V. (1993) Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics, 65*, 413–497.

Heinks-Maldonado, T.H., Mathalon, D.H., Gray, M., & Ford, J.M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology, 42*, 180-90. doi: 10.1111/j.1469-8986.2005.00272.x

Heinks-Maldonado, T.H., Nagarajan, S.S., & Houde, J.F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport, 17*, 1375-9. doi: 10.1097/01.wnr.0000233102.43526.e9

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13(2)*, 135–145.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182(4108)*, 177–180. doi:10.1126/science.182.4108.177

Holst, E., & Mittelstaedt, H. (1950). Das reafferenzprinzip. *Naturwissenschaften, 37(20)*, 464–476.

Hommel, B. (1993). Inverting the Simon effect by intention. *Psychological Research, 55(4)*, 270–279.

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica, 136(2)*, 189–202. doi:10.1016/j.actpsy.2010.04.011

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): a framework for perception and action planning. *Behavioral and Brain Sciences, 24(5)*, 849–878; discussion 878–937.

Horváth, J. (2013a). Action-sound coincidence-related attenuation of auditory ERPs is not modulated by affordance compatibility. *Biological Psychology, 93(1)*, 81–87. doi:10.1016/j.biopsycho.2012.12.008

Horváth, J. (2013b). Attenuation of auditory ERPs to action-sound coincidences is not explained by voluntary allocation of attention. *Psychophysiology, 50(3)*, 266–273. doi:10.1111/psyp.12009

Horváth, J. (2014a). Probing the sensory effects of involuntary attention change by ERPs to auditory transients: Probing the sensory impact of distraction. *Psychophysiology, 51(5)*, 489–497. doi:10.1111/psyp.12187

Horváth, J. (2014b). Sensory ERP effects in auditory distraction: did we miss the main event? *Psychological Research, 78(3)*, 339–348. doi:10.1007/s00426-013-0507-7

Horváth, J., & Burgyán, A. (2013). No evidence for peripheral mechanism attenuating auditory ERPs to self-induced tones. *Psychophysiology, 50(6)*, 563–569. doi:10.1111/psyp.12041

Horváth, J., Maess, B., Baess, P., & Tóth, A. (2012). Action–sound coincidences suppress evoked responses of the human auditory cortex in EEG and MEG. *Journal of Cognitive Neuroscience, 24(9)*, 1919–1931. doi:10.1162/jocn_a_00215

Horváth, J., & Winkler, I. (2010). Distraction in a continuous-stimulation detection task. *Biological Psychology, 83(3)*, 229–238. doi:10.1016/j.biopsycho.2010.01.004

Houde, J.F., Nagarajan, S.S., Sekihara, K., & Merzenich, M.M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience, 14,* 1125-1138. doi:10.1162/089892902760807140

Hughes, G., Desantis, A., & Waszak, F. (2013a). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin, 139(1)*, 133–151. doi:10.1037/a0028566

Hughes, G., Desantis, A., & Waszak, F. (2013b). Attenuation of auditory N1 results from identity-specific action-effect prediction. *European Journal of Neuroscience, 37(7)*, 1152–1158. doi:10.1111/ejn.12120

Hume, D. (1739/1896) *A treatise of human nature. Reprinted from the Original Edition in three volumes and edited, with an analytical index, by L.A. Selby-Bigge, M.A.* Oxford, Clarendon. Retrieved Feb 6, 2012 from http://oll.libertyfund.org/title/342.

Jasper, H.H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10*, 371-375

Joiner, W. M., & Smith, M. A. (2008). Long-term retention explained by a model of short-term learning in the adaptive control of reaching. *Journal of Neurophysiology*, *100*(5), 2948–2955. doi:10.1152/jn.90706.2008

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences, 97(22)*, 11793–11799.

Kauramäki, J., Jääskeläinen, I. P., Hänninen, J. L., Auranen, T., Nummenmaa, A., Lampinen, J., & Sams, M. (2012). Two-Stage processing of sounds explains behavioral performance variations due to changes in stimulus contrast and selective attention: An MEG study. *PLoS ONE, 7(10)*, e46872. doi:10.1371/journal.pone.0046872

Kauramäki, J., Jääskeläinen, I. P., & Sams, M. (2007). Selective attention increases both gain and feature selectivity of the guman auditory cortex. *PLoS ONE, 2(9)*, e909. doi:10.1371/journal.pone.0000909

Knight, R. T., Hillyard, S. A., Woods, D. L., & Neville, H. J. (1980). The effects of frontal and temporal-parietal lesions on the auditory evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, *50*(1-2), 112–124. doi:10.1016/0013-4694(80)90328-4

Knight, R. T., Hillyard, S. A., Woods, D. L., & Neville, H. J. (1981). The effects of frontal cortex lesions on event-related potentials during auditory selective attention. *Electroencephalography and Clinical Neurophysiology, 52(6)*, 571–582. doi:10.1016/0013-4694(81)91431-0

Knolle, F., Schröger, E., Baess, P., & Kotz, S. A. (2012). The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. *Journal of Cognitive Neuroscience, 24(3)*, 698–706. doi:10.1162/jocn_a_00167

Knolle, F., Schröger, E., & Kotz, S. A. (2013a). Prediction errors in self- and externally-generated deviants. *Biological Psychology, 92(2)*, 410–416. doi:10.1016/j.biopsycho.2012.11.017

Knolle, F., Schröger, E., & Kotz, S. A. (2013b). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex*, *49*(9), 2449–2461. doi:10.1016/j.cortex.2012.12.012

Lange, K. (2011). The reduced N1 to self-generated tones: An effect of temporal predictability? *Psychophysiology, 48(8)*, 1088–1095. doi:10.1111/j.1469-8986.2010.01174.x

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

Lange, K., Rösler, F., & Röder, B. (2003). Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: An event-related potential study. *Psychophysiology, 40(5)*, 806–817.

Liberman, M. C., & Guinan, J. J., Jr. (1998). Feedback control of the auditory periphery: anti-masking effects of middle ear muscles vs. olivocochlear efferents. *Journal of Communication Disorders, 31(6)*, 471–482

Lütkenhöner, B., & Klein, J.-S. (2007). Auditory evoked field at threshold. *Hearing Research*, *228*(1-2), 188–200. doi:10.1016/j.heares.2007.02.011

Makeig, S., Mueller, M. M., & Rockstroh, B. (1996). Effects of voluntary movements on early auditory brain responses. *Experimental Brain Research, 110(3)*, 487–492.

Martikainen, M. H., Kaneko, K., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex, 15(3),* 299–302. doi:10.1093/cercor/bhh131

Masson, M., & Loftus, G.R. (2003). Using confidence intervals for graphically based interpretation. *Canadian Journal of Experimental Psychology, 57*, 203–220.

McCarthy, G., & Donchin, E. (1976). The Effects of Temporal and Event Uncertainty in Determining the Waveforms of the Auditory Event Related Potential (ERP). Psychophysiology, 13(6), 581–590. doi:10.1111/j.1469-8986.1976.tb00885.x

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 62(3), 203–208. doi:10.1016/0168-5597(85)90015-2

McGrenere, J., Ho, W. (2000): Affordances: Clarifying and evolving a concept. In: *Proceedings of Graphics Interface 2000 May 15-17, 2000, Montreal, Quebec, Canada.* pp. 179-186.

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. Neural Networks, 9(8), 1265–1279.

Moore, B. C. J. (2012). *An introduction to the psychology of hearing*. Bingley: Emerald.

Mukerji, S., Windsor, A. M., & Lee, D. J. (2010). Auditory brainstem circuits that mediate the middle ear muscle reflex. *Trends in Amplification, 14(3)*, 170–191. doi:10.1177/1084713810381771

Näätänen, R. (1982). Processing negativity: An evoked-potential reflection of selective attention. *Psychological Bulletin, 92(3)*, 605-640.

Näätänen R. (1988) Implications of ERP data for psychological theories of attention. *Biological Psychology, 26,* 117-163

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences, 13(02)*, 201–233. doi:10.1017/S0140525X00078407

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42(4)*, 313–329. doi:10.1016/0001-6918(78)90006-9

Näätänen, R., & Michie, P. T. (1979). Early selective-attention effects on the evoked potential: a critical review and reinterpretation. *Biological Psychology, 8(2)*, 81–136.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24(4),* 375–425. doi:10.1111/j.1469-8986.1987.tb00311.x

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin, 125(6)*, 826–859.

Neszmélyi, B (2014) *A cselekvés révén kialakuló hallási szuppresszió viselkedéses korrelátumainak vizsgálata* (Behavioral correlates of action-related auditory suppression; in Hungarian). (Unpublished M.A. thesis in psychology). Budapest University of Technology and Economics, Hungary

Norman, D. A. (1989). *The design of everyday things.* New York: Doubleday

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience*, *33*(41), 16110–16116. doi:10.1523/JNEUROSCI.2137-13.2013

Nuwer, M. R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guérit, J.-M., Hinrichs, H., … Rappelsburger, P. (1998). IFCN standards for digital recording of clinical EEG. *Electroencephalography and Clinical Neurophysiology, 106(3)*, 259–261. doi:10.1016/S0013-4694(97)00106-5

Okamoto, H., Stracke, H., Wolters, C.H., Schmael, F., Pantev, C. (2007). Attention Improves Population-Level Frequency Tuning in Human Auditory Cortex. *The Journal of Neuroscience 27*(39), 10383–10390. doi: 10.1523/JNEUROSCI.2963-07.2007

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8(4)*, 434–447. doi:10.1037/1082-989X.8.4.434

Ozaki, I., Jin, C. Y., Suzuki, Y., Baba, M., Matsunaga, M., & Hashimoto, I. (2004). Rapid change of tonotopic maps in the human auditory cortex during pitch discrimination. *Clinical Neurophysiology, 115(7)*, 1592–1604. doi:10.1016/j.clinph.2004.02.011

Pang, X. D., & Guinan, J. J., Jr. (1997). Effects of stapedius-muscle contractions on the masking of auditory-nerve responses. *The Journal of the Acoustical Society of America, 102(6)*, 3576–3586. doi:10.1121/1.420399

Pang, X. D., & Peake, W. T. (1986). How do contractions of the stapedius muscle alter the acoustic properties of the ear?. In: J.B. Allen, J.L. Hall, A. Hubbard, S.T. Neely, & A. Tubis (Eds.), *Peripheral auditory mechanisms: proceedings of a conference held at Boston University, Boston, MA, August 13-16* (pp. 36–43), Springer-Verlag

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology, 72(2)*, 184–187. doi:10.1016/0013-4694(89)90180-6

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1990). Corrigendum. *Electroencephalography and Clinical Neurophysiology, 76*, 565.

Picton, T.W., Woods D.L., Baribeau-Braun, J., Healey, T.M.G. (1977). Evoked potential audiometry. *Journal of Otolaryngology 6(2)*, 89-119.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Ponton, C., Eggermont, J. J., Khosla, D., Kwong, B., & Don, M. (2002). Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. *Clinical Neurophysiology, 113(3)*, 407–420. doi:10.1016/S1388-2457(01)00733-7

Pratt, H., Starr, A., Michalewski, H. J., Dimitrijevic, A., Bleich, N., & Mittelman, N. (2009). Auditory-evoked potentials to frequency increase and decrease of high- and low-frequency tones. *Clinical Neurophysiology, 120(2)*, 360–373. doi:10.1016/j.clinph.2008.10.158

Proske, U., & Gandevia, S. C. (2012). The proprioceptive senses: Their roles in signaling body shape, body position and movement, and muscle force. *Physiological Reviews*, *92*(4), 1651–1697. doi:10.1152/physrev.00048.2011

R Core Team. (2013). R: A language and environment for statistical computing. (Version 3.0.0). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rabinowitz, M.W. (1977) *Acoustic-reflex effects on the input admittance and transfer characteristics of the human middle-ear*. (Ph. D. Thesis). Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science. Retrieved from http://hdl.handle.net/1721.1/16406.

Rapin, I., Schimmel, H., Tourk, L. M., Krasnegor, N. A., & Pollak, C. (1966). Evoked responses to clicks and tones of varying intensity in waking adults. *Electroencephalography and Clinical Neurophysiology*, *21*(4), 335–344.

Repp, B. H., & Knoblich, G. (2007). Action can affect auditory perception. *Psychological Science, 18*, 6–7. doi: 10.1111/j.1467-9280.2007.01839.x

Rif, J., Hari, R., Hämäläinen, M.S., Sams, M. (1991). Auditory attention affects two different areas in the human supratemporal cortex. *Electroencephalography and Clinical Neurophysiology 79 (6)*, 464–472. doi:10.1016/0013-4694(91)90166-2.

Ritter, W., Simson, R., Vaughan, H. G. Jr., Friedman, D. (1979) A brain event related to the making of a sensory discrimination. *Science, 203*, 1358-1361. doi: 10.1126/science.424760

Ritter, W., Vaughan Jr, H. G., & Costa, L. D. (1968). Orienting and habituation to auditory stimuli: a study of short terms changes in average evoked responses. *Electroencephalography and Clinical Neurophysiology*, *25*(6), 550–556.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16(2)*, 225–237. doi:10.3758/PBR.16.2.225

Roth, W. T., Krainz, P. L., Ford, J. M., Tinklenberg, J. R., Rothbart, R. M., & Kopell, B. S. (1976). Parameters of temporal recovery of the human auditory evoked potential. *Electroencephalography and Clinical Neurophysiology*, *40*(6), 623–632.

Roussel, C., Hughes, G., & Waszak, F. (2013). A preactivation account of sensory attenuation. *Neuropsychologia, 51*, 922–929. doi:10.1016/j.neuropsychologia.2013.02.005

Roussel, C., Hughes, G., & Waszak, F. (2014). Action prediction modulates both neurophysiological and psychophysical indices of sensory attenuation. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00115

Roy, J.E., Cullen, K.E. (2004) Dissociating self-generated from passively applied head motion: neural mechanisms in the vestibular nuclei. *Journal of Neuroscience, 24*, 2102-2111.

Salomon, G., & Starr, A. (1963). Electromyography of middle ear muscles in man during motor activities. *Acta Neurologica Scandinavica, 39(2)*, 161–168. doi:10.1111/j.1600-0404.1963.tb05317.x

SanMiguel, I., Saupe, K., & Schröger, E. (2013). I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted ”what” but not ”when”. *Frontiers in Human Neuroscience, 7*. doi:10.3389/fnhum.2013.00407

SanMiguel, I., Todd, J., & Schröger, E. (2013). Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. *Psychophysiology, 50,* 334–343. doi:10.1111/psyp.12024

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schroger, E. (2013). Hearing silences: Human auditory processing relies on preactivation of sound-specific brain activity patterns. *Journal of Neuroscience, 33(20)*, 8633–8639. doi:10.1523/JNEUROSCI.5821-12.2013

Saupe, K., Widmann, A., Trujillo-Barreto, N. J., & Schröger, E. (2013). Sensorial suppression of self-generated sounds and its dependence on attention. *International Journal of Psychophysiology, 90*(3), 300–310. doi:10.1016/j.ijpsycho.2013.09.006

Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. *Consciousness and Cognition, 17(4)*, 1219–1227. doi:10.1016/j.concog.2008.01.003

Schafer, E. W. P., & Marcus, M. M. (1973). Self-stimulation alters human sensory brain responses. *Science, 181(4095)*, 175–177. doi:10.1126/science.181.4095.175

Scherg, M, & Von Cramon, D. (1985). Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 62(1)*, 32–44. doi:10.1016/0168-5597(85)90033-4

Scherg, M., & Von Cramon, D. (1986). Evoked dipole source potentials of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *65*(5), 344–360. doi:10.1016/0168-5597(86)90014-6

Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., & Pfurtscheller, G. (2007). A fully automated correction method of EOG artifacts in EEG recordings. Clinical Neurophysiology, 118(1), 98–104. doi:10.1016/j.clinph.2006.09.003

Schwent, V.L., Hillyard, S.A., Galambos, R. (1976). Selective attention and the auditory vertex potential. II. Effects of signal intensity andmasking noise. *Electroencephalography and Clinical Neurophysiology 40 (6)*, 615–622. doi:10.1016/0013-4694(76)90136-X.

Seppänen, M., Hämäläinen, J., Pesonen, A.-K., & Tervaniemi, M. (2012). Music Training Enhances Rapid Neural Plasticity of N1 and P2 Source Activation for Unattended Sounds. *Frontiers in Human Neuroscience*, *6*. doi:10.3389/fnhum.2012.00043

Shadmehr, R., Smith, M.A., & Krakauer, J.W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience, 33*, 89–108. doi: 10.1146/annurev-neuro-060909-153135

Simon, J. R., & Rudell, A. P. (1967). Auditory SR compatibility: the effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51(3)*, 300-304. doi: 10.1037/h0020586

Simmons, F. B. (1964). Perceptual theories of middle ear muscle function. *The Annals of Otology, Rhinology, and Laryngology, 73*, 724–739.

Sitek, K. R., Mathalon, D. H., Roach, B. J., Houde, J. F., Niziolek, C. A., & Ford, J. M. (2013). Auditory cortex processes variation in our own speech. *PLoS ONE*, *8*(12), e82925. doi:10.1371/journal.pone.0082925

Slegel, D. E., Benson, K. L., Zarcone, V. P., Jr, & Schubert, E. D. (1991). Middle-ear muscle activity (MEMA) and its association with motor activity in the extremities and head in sleep. *Sleep, 14(5)*, 454–459.

Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biology*, *4*(6), e179. doi:10.1371/journal.pbio.0040179

Sowman, P. F., Kuusik, A., & Johnson, B. W. (2012). Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. Experimental Brain Research, 222(1-2), 149–157. doi:10.1007/s00221-012-3204-7

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative & Physiological Psychology, 43(6)*, 482–489.

Sutter, C., Sülzenbrück, S., Rieger, M., & Müsseler, J. (2013). Limitations of distal effect anticipation when using tools. *New Ideas in Psychology, 31(3)*, 247–257. doi:10.1016/j.newideapsych.2012.12.001

Suzuki, Y., & Takeshima, H. (2004). Equal-loudness-level contours for pure tones. *The Journal of the Acoustical Society of America, 116(2)*, 918–933. doi:10.1121/1.1763601

Taulu, S., Kajola, M., Simola, J. (2004). Suppression of interference and artifacts by the signal space separation method. *Brain Topography, 16*, 269–275.

Thura, D., Hadj-Bouziane, F., Meunier, M., & Boussaoud, D. (2011) Hand modulation of visual, preparatory, and saccadic activity in the monkey frontal eye field. *Cerebral Cortex, 21*, 853-864.

Timm, J., SanMiguel, I., Saupe, K., & Schröger, E. (2013). The N1-suppression effect for self-initiated sounds is independent of attention. *BMC Neuroscience, 14(1)*, 2. doi:10.1186/1471-2202-14-2

Timm, J., SanMiguel, I., Keil, J., Schröger, E., & Schönwiesner, M. (2014). Motor intention determines sensory attenuation of brain responses to self-initiated sounds. *Journal of Cognitive Neuroscience, 26(7)*, 1481–1489. doi:10.1162/jocn_a_00552

Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., & Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Frontiers in Systems Neuroscience*, *8*. doi:10.3389/fnsys.2014.00028

Vaughan Jr, H. G., & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography and Clinical Neurophysiology, 28(4)*, 360–367.

Vaziri, S., Diedrichsen, J., Shadmehr, R. (2006). Why does the brain predict sensory consequences of oculomotor commands? Optimal integration of the predicted and the actual sensory feedback. *Journal of Neuroscience, 26*, 4188–97.

Ventura, M.I., Nagarajan, S.S., & Houde, J.F. (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience, 10*, 58. doi:10.1186/1471-2202-10-58

Weiss, C., Herwig, A., & Schütz-Bosbach, S. (2011). The self in action effects: Selective attenuation of self-generated sounds. Cognition, 121(2), 207–218. doi:10.1016/j.cognition.2011.06.011

Weitzman, E. D., & Kremen, H. (1965). Auditory evoked responses during different stages of sleep in man. *Electroencephalography and Clinical Neurophysiology*, *18*(1), 65–70.

Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology, 79*(3), 170–191. doi: 10.1016/0013-4694(91)90136-R

Wolpaw, J. R., & Penry, J. K. (1975). A temporal component of the auditory evoked response. Electroencephalography and clinical neurophysiology, 39(6), 609–620. doi:10.1016/0013-4694(75)90073-5

Wolpert, D.M., Ghahramani, Z., & Jordan, M.I. (1995). An internal model for sensorimotor integration. *Science, 269*, 1880-1882.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences, 2(9)*, 338–347.

Zakrisson, J. E., & Borg, E. (1974). Stapedius reflex and auditory fatigue. *Audiology, 13(3)*, 231–235.