

# The Sounds of Sentences: Differentiating the Influence of Physical Sound, Sound Imagery, and Linguistically Implied Sounds on Physical Sound Processing

Carolin Dudschig<sup>1</sup> · Ian Grant Mackenzie<sup>1</sup> · Jessica Strozyk<sup>1</sup> · Barbara Kaup<sup>1</sup> · Hartmut Leuthold<sup>1</sup>

Published online: 29 July 2016  
© Psychonomic Society, Inc. 2016

**Abstract** Both the imagery literature and grounded models of language comprehension emphasize the tight coupling of high-level cognitive processes, such as forming a mental image of something or language understanding, and low-level sensorimotor processes in the brain. In an electrophysiological study, imagery and language processes were directly compared and the sensory associations of processing linguistically implied sounds or imagined sounds were investigated. Participants read sentences describing auditory events (e.g., “The dog barks”), heard a physical (environmental) sound, or had to imagine such a sound. We examined the influence of the 3 sound conditions (linguistic, physical, imagery) on subsequent physical sound processing. Event-related potential (ERP) difference waveforms indicated that in all 3 conditions, prime compatibility influenced physical sound processing. The earliest compatibility effect was observed in the physical condition, starting in the 80–110 ms time interval with a negative maximum over occipital electrode sites. In contrast, the linguistic and the imagery condition elicited compatibility effects starting in the 180–220 ms time window with a maximum over central electrode sites. In line with the ERPs, the analysis of the oscillatory activity showed that compatibility influenced early theta and alpha band power changes in the physical, but not in the linguistic and imagery, condition. These dissociations were further confirmed by dipole localization results showing a clear separation between the source of the compatibility effect in the physical sound condition (superior temporal area) and the source of the compatibility effect triggered by the linguistically implied sounds or the imagined sounds (inferior temporal area). Implications for grounded models of language understanding are discussed.

**Keywords** Embodied cognition · ERP · Modal · Amodal · Imagery · Language · Environmental sound processing

Mental imagery has always been of central interest for psychological and philosophical models of internal representations within the human mind (Kosslyn & Pomerantz, 1977; Kosslyn, Thompson, & Ganis, 2006). References to imagery can be found in the first approaches to understanding human thought in ancient philosophical works (Thomas, 2016). In modern cognitive science, an important question concerns the format of mental representations on which various cognitive processes—such as thinking, memory, or language—are operating. Despite the majority of imagery research referring to the visual domain, it is now acknowledged that imagery—a quasiperceptual experience—is a basic feature with regard to all senses: olfactory imagery, haptic imagery, motor imagery, and, most important for the current study—auditory imagery (Djordjevic, Zatorre, Petrides, Boyle, & Jones-Gotman, 2005; Djordjevic, Zatorre, Petrides, & Jones-Gotman, 2004; Halpern & Zatorre, 1999; Klatzky, Lederman, & Matula, 1991; Kraemer, Macrae, Green, & Kelley, 2005; Segal & Fusella, 1971; Stevenson & Case, 2005).

Imagery was first referred to as an internal subjective experience. For example, auditory imagery was described as “the introspective persistence of an auditory experience, including one constructed from components drawn from long-term memory, in the absence of direct sensory instigation of that experience” (Intons-Peterson, 1992, p. 46). However, in recent years, the term *imagery* has often been used to describe a type of representation format; specifically, imagery is regarded as perceptual-like and is typically opposed to mere propositional, abstract representations (e.g., Hubbard, 2010; Kosslyn & Pomerantz, 1977; but see Pylyshyn, 1973). Core research questions concern the similarity between imagery and perceptual processes. With the development of psychophysiological methods, these questions have been addressed by investigating the involvement of primary sensory cortices during imagery with the following logic: If imagery involves

✉ Carolin Dudschig  
carolin.dudschig@uni-tuebingen.de

<sup>1</sup> Universität Tübingen, Schleichstr. 4, 72076 Tübingen, Germany

(primary) sensory cortex activity, this would strongly suggest that perception and imagery are of the same quality and might only vary with regard to intensity rather than being associated with two qualitatively different types of representations. Penfield and Perot (1963) first reported that electrically activating the superior temporal gyrus in epilepsy patients resulted in auditory hallucinations. Further studies extended this line of investigation in both the visual and auditory domain. In the visual domain, there are mixed findings, whereby some show primary cortex activation and other studies show activation only in secondary sensory cortices (Ishai, Ungerleider, & Haxby, 2000; Kosslyn, Ganis, & Thompson, 2001; Kosslyn, Thompson, Kim, & Alpert, 1995; O'Craven & Kanwisher, 2000; Wheeler, Petersen, & Buckner, 2000). In the auditory domain, mainly secondary auditory cortex activation is reported during imagery (e.g., Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005; Wheeler et al., 2000; Zatorre, Halpern, Perry, Meyer, & Evans, 1996). Still, some studies also report activation of primary sensory regions when mentally reconstructing previously heard tones (Yoo, Lee, & Choi, 2001) or when imagery required the reconstruction of “high-level” perceptual features (Kraemer et al., 2005). However, with imaging techniques moving beyond the pure mapping of specific processes in the brain toward a more comprehensive analysis of the neural activity patterns, a recent study suggests that during auditory imagery of environmental sounds, the information encoded in the neural patterns is more abstract than during perception or memory maintenance of these sounds (Linke & Cusack, 2015).

In the domain of language comprehension, a similar debate has become very influential. Specifically, embodied or simulation models of language comprehension (e.g., Barsalou, 1999; Glenberg & Kaschak, 2002) propose a close connection between language processes and sensorimotor processes in the brain. These interconnections between motor, sensory, and language processing have been demonstrated for various linguistically described events. For example, Hauk, Johnsrude, and Pulvermüller (2004) showed that when reading words referring to specific motor actions (e.g., *kick*), motor regions in the brain become active similar to those that are activated during performing the according action—also including the primary motor area. For sensory events, similar interconnections have been found. For example, when reading words referring to specific smells (e.g., *cinnamon*), olfactory brain regions have been demonstrated to become activated (González et al., 2006). Kiefer, Sim, Herrnberger, Grothe, and Hoenig (2008) first reported a direct link for auditory and lexical conceptual brain systems. In line with other studies, they compared the brain regions involved during physical sound processing and processing sound-related words (e.g., *telephone*) in an fMRI study. The results showed that sound-related words specifically activate areas in the auditory cortex that are also involved during physical sound processing. The authors concluded that these results support modality-specific representations of the linguistically described concepts in line with the embodied language-processing model (e.g., Barsalou, 2008).

These debates regarding the processes underlying imagery and the processes underlying language comprehension have taken place largely independently. In our study, we aim at directly comparing the effects of imagery and sentence comprehension on subsequent auditory perception processes. If both—active imagery and passive language comprehension—work on similar representation formats, one would expect them both to influence subsequent perceptual processes in a similar manner. The first research question addressed in this article concerns the similarity of active imagery and passive sentence comprehension. The second goal of this study is to shed light on the nature of representations that become active during imagery and during sentence comprehension. Specifically, we will investigate whether imagery and sentence processing activate early sensory processes, resembling activation during perceiving the described or imagined events, as claimed in strong embodiment models of language comprehension (see Binder & Desai, 2011; Meteyard & Vigliocco, 2008). To address these questions, we will implement a paradigm that allows comparing the influence of sound imagery, of physical sound processing, and of comprehending sentences describing auditory events on subsequent physical sound processing.

It is well known that linguistic stimuli can directly influence subsequent sound processing (e.g., Ballas, 1993). In the grounded language comprehension literature, behavioral studies typically show that words or sentences influence subsequent picture, auditory, or motor processing (e.g., Brunyé, Ditman, Mahoney, Walters, & Taylor, 2010; Dudschig, de la Vega, & Kaup, 2015; Wolter, Dudschig, de la Vega, & Kaup, 2015; Zwaan, Stanfield, & Yaxley, 2002). Brunyé et al. (2010) demonstrated that sound categorization was faster if preceded by a compatible sentence. For example, participants were faster to indicate that a “waves on beach” sound is a natural and not a computer-generated sound if the sound was preceded by the sentence “The waves crashed against the beach as the tide began to rise” than by a sentence implying a different type of sound. The authors concluded that language comprehension is closely related to sensorimotor activations and that comprehension triggers the simulation of multimodal representations, even in situations where no active imagery is demanded. However, from behavioral data alone, it is often difficult to conclude that actual sensory processes are involved during such tasks (Hubbard, 2010).

In addition to the behavioral studies, there are also several electrophysiological studies showing cross-modal associations between language comprehension and sound processing. Van Petten and Rieffers (1995) first investigated the influence of spoken words on processing environmental sounds of 2,500 ms duration. They found that in both orders—sounds followed by words and words followed by sounds—the N400 measured on the second stimulus was larger in the case of a mismatch between the two stimuli. Orgs, Lange, Dombrowski, and Heil (2006, 2007, 2008) used a similar experimental design and also found N400-like effects in a 200–500 ms time interval poststimulus,

despite using a sound duration of only 300 ms and using different tasks with various processing demands regarding the word–sound association. Schön, Ystad, Kronland-Martinet, and Besson (2010) also investigated conceptual priming using both words and sounds as targets. Compared to previous studies, they implemented acousmatic sounds—nonmeaningful sounds with indefinite sources—with an average duration of 820 ms. Such sounds were implemented to reduce the possibility that the compatibility effects reflect linguistically moderated labelling effects (see also Cummings et al., 2006; Koelsch et al., 2004). Nevertheless, a N400-like relatedness effect developed in both the sound–word and the word–sound condition within the 250–600 ms time window. This finding is in line with the idea that labeling presumably does not improve the recognition process and is usually not used by participants because it adds additional processing costs during sound perception (Dick, Bussiere, & Saygin, 2002). Various other studies show that sounds influence processing of nonauditory events, and that nonauditory events influence subsequent sound processing (e.g., Daltrozzo & Schön, 2009; Koelsch et al., 2004; Schirmer, Soh, Penney, & Wyse, 2011). Critically, both the behavioral and the electrophysiological studies often do not allow strong conclusions to be drawn regarding the involvement of sensory processes in causing the behavioral or electrophysiological effects. In both cases, the effects are sometimes attributed to sensory-perceptual associations and sometimes to conceptual associations on a higher-level multimodal integration stage (e.g., Hubbard, 2010; Schirmer et al., 2011).

In this study, we aim at directly comparing the influence of sentence processing on subsequent physical sound processing with the effects of imagery on sound processing, and additionally differentiate sensory from nonsensory contributions to these effects. To this end, we implemented three experimental conditions (see Fig. 1). In the first condition (linguistic), participants read sentences (e.g., “The dog barks.”) and subsequently perceived an auditory sound that either matched or mismatched the described event. In the second condition (imagery), participants were instructed to actively imagine a specific sound (e.g., *barking*) and subsequently perceived a matching or mismatching sound (see Intons-Peterson, 1980). In the third condition (physical), a physical sound (e.g., *barking*) was followed by another matching (identical) or mismatching sound. This purely auditory condition served as a control condition; specifically, it allows to assess the time course as well as the perceptual processes underlying the match–mismatch effects. To compare the nature of linguistic, physical, or imagery events on subsequent sound processing regarding the underlying sensory or higher-level cognitive processes, distinct markers of brain activity were assessed. These included in the event-related potential (ERP) signal the time-course, the distribution, and the neural generators of specific ERP components as well as in the non-averaged EEG data modulations of oscillatory brain activity. We predict the following: If imagery and sentence comprehension work on

similar representations, we expect both conditions to similarly influence subsequent sound processing. Specifically, the time-course, distributions, and neural generators of the event-related effects should be identical. Such a finding would be in line with the hypothesis that during reading, imagery processes become automatically activated. Additionally, at least according to strong versions of embodiment accounts (e.g., Glenberg & Gallese, 2012), these influences should be based on purely perceptual processes. In this case, we predict very early perceptual effects to be present—for example, in P1/N1 attenuation effects (e.g., Näätänen & Picton, 1987)—and a strong similarity in the neural generators involved in the purely physical condition and the other two conditions. Additionally, in line with Fuentemilla, Marco-Pallarés, and Grau’s (2006) findings in a tone-repetition paradigm, we would predict that priming effects that are based on a sensory preactivation process should be reflected in theta and alpha band power change.

## Method

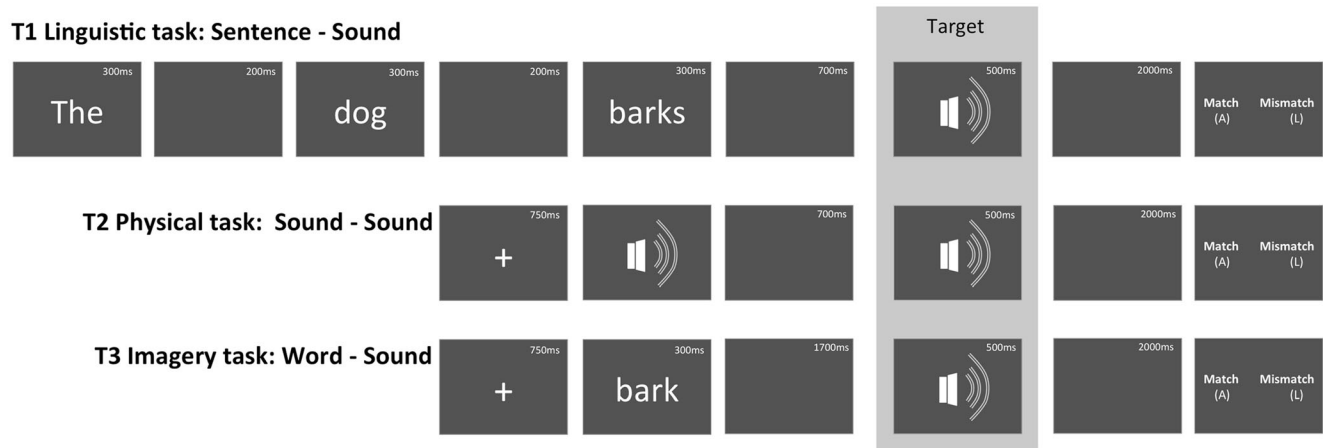
### Participants

Thirty participants were tested; two were excluded because they were tested in the wrong experimental version. The remaining sample consisted of 28 German native speakers ( $M_{\text{age}} = 24.86$  years,  $SD_{\text{age}} = 3.86$  years, seven male, one left-handed). Participants were naïve regarding the experimental hypothesis and received course credits or were paid 16 € for participation.

### Stimuli

Auditory stimuli were natural sounds sampled from several online databases (salamisound.de, mediamanual.at/mediamanual/workshop/audiovisuell/ton.php, freesound.org, theaterverlag.eu/28. Sounds\_Geraeusche.html). Sounds were sampled at a 44.1 kHz rate and stored as 16-bit mono .wav files. Further processing involved normalization to a maximum amplitude of 1.0 dB, DC offset removal and a 5 ms fading in and out interval.

In a rating study, 18 participants ( $M_{\text{age}} = 23.72$  years,  $SD_{\text{age}} = 3.58$  years, three male, all right-handed), who did not take part in the actual experiment, rated all sound stimuli according to their familiarity on a 7-point scale ranging from 1 (*unfamiliar*) to 7 (*very familiar*), because sound familiarity has been shown to affect environmental sound processing (Kirmse, Jacobsen, & Schröger, 2009; Kirmse, Schröger & Jacobsen, 2012). First, the sentence was displayed on the screen. After a 700 ms delay, the sound was presented for 500 ms via headphones. The presentation procedure was identical to the linguistic condition of the actual experiment. On average, the tones were rated as rather familiar ( $M = 4.92$ ,  $SD$



**Fig. 1** Trial procedure for the three tasks T1<sub>linguistic</sub> (sentence–sound), T2<sub>physical</sub> (sound–sound) and T3<sub>imagery</sub> (imagery–sound)

= 0.97), suggesting that participants perceived the sounds as natural. In a second rating study, 18 participants ( $M_{\text{age}} = 24.11$  years,  $SD_{\text{age}} = 3.74$  years, three male, all right-handed) rated the stimuli in a setup resembling the imagery condition. Here, participants were presented with the verb (e.g., *to bark*) and subsequently with the according sound. They also rated sound familiarity on a 7-point scale ranging from 1 (*unfamiliar*) to 7 (*very familiar*). Again, participants rated the sounds as rather familiar ( $M = 4.97$ ,  $SD = 0.73$ ). A comparison between the two rating studies showed no difference in familiarity ratings,  $t(79) = 0.34$ ,  $p = .73$ , suggesting that sentences and verbs were equally constraining with regard to sound familiarity. A final norming study addressed whether our sound stimuli can be verbally labeled. Previous studies suggested that sound labeling typically does not take place during sound recognition (Dick et al., 2002), although the contribution of labeling to sound processing cannot be completely ruled out (Schön et al., 2010). Therefore, a new set of 18 participants ( $M_{\text{age}} = 24.89$  years,  $SD_{\text{age}} = 4.50$  years, three male, all right-handed) first heard the 500 ms sound file and subsequently was asked to label the sound with a verb. Participants on average succeeded in 34.72 % of these trials in naming the correct verb. This leads to the conclusion that, in general, our sounds were rather hard to label for the participants.

### Procedure and Design

The experimental procedure was controlled using MATLAB, Psychtoolbox 3.0.10 (Kleiner et al., 2007). The visual stimuli were displayed in white on a black background on a 21 in. CRT monitor screen (resolution 1,280 × 960, 60 Hz). The sounds were played via headphones (Sennheiser PX-100-ii).

The experiment consisted of three different tasks. In task T1<sub>linguistic</sub> (linguistic task), participants had to indicate whether a sound matched with a previously displayed sentence. In task T2<sub>physical</sub> (physical task), participants had to indicate

whether two subsequently presented sounds matched or mismatched. In task T3<sub>imagery</sub> (imagery task), participants had to indicate whether a sound matched to their imagined sound, according to a previously displayed word.

Each trial began with a centrally presented fixation cross for 750 ms. In T1<sub>linguistic</sub>, the fixation cross was subsequently followed by three centrally presented words (e.g., *The dog barks*) for 300 ms, each separated by a 200 ms blank screen; 700 ms after the offset of the last word, a matching or mismatching sound was presented for 500 ms; 2,000 ms after tone offset, the response screen was displayed that indicated which key participants had to press for indicating that the sound matched or mismatched the sentence (*x* or *n* key on a standard German keyboard). The key assignments for match and mismatch responses varied randomly from trial to trial to prevent response planning before the onset of the response screen. Task T2<sub>physical</sub> was identical to task T1<sub>linguistic</sub>, with the only difference being that a sound was presented for 500 ms instead of the word-by-word sentence. Task T3<sub>imagery</sub> was identical to task T1<sub>linguistic</sub> with the difference being that only one word (e.g., *bark*, *ring*) was displayed, and participants were instructed to actively imagine that sound. To allow sufficient time to imagine the according sound, the interval between word offset and tone onset was increased to 2,000 ms.

The sequence of the three tasks was varied between participants, with half of the participants starting with the sentence task, followed by the sound task, and finally the imagery task. The other half of the participants started with the imagery task, followed by the sound task, and finally the sentence task. We choose these two versions of task order with T2<sub>physical</sub> always being the middle task, as we wanted to separate T1<sub>linguistic</sub> and T3<sub>imagery</sub> as far as possible within each participant. Each task consisted of 160 trials separated into four blocks each. Each participant was presented with each sound twice in each task (sentence, sound, imagery) once in a match and once in a mismatch condition; it was ensured that a sound was not repeated within a block or in consecutive blocks. Whether a



sound was presented first in a match or mismatch condition was randomly selected for half of the participants (14 random lists), with the other half of the participants being assigned to the mirrored experimental versions. The sounds that were assigned to the mismatch conditions for one participant were the matching sounds for the other participant. This procedure ensured that each sound and sentence occurred equally often in a match and mismatch condition. It also ensured that the sounds were identical across the three task conditions ( $T1_{\text{linguistic}}$ ,  $T2_{\text{physical}}$ , and  $T3_{\text{imagery}}$ ) and that potentially observed differences between the conditions (task and compatibility) cannot be attributed to low-level physical differences in the sounds used.

## EEG Recording and Analysis

**Electrophysiological measures** The electroencephalogram (EEG) and the electrooculogram (EOG) recordings were sampled at 512 Hz. The activity was recorded from the following midline electrodes: Fpz, AFz, Fz, FCz, Cz, CPz, Pz, POz, Oz, and Iz; the following left hemisphere electrodes: IO1, Fp1, AF3, AF7, F1, F3, F5, F7, F9, FC1, FC3, FC5, FT7, C1, C3, C5, M1, T7, CP1, CP3, CP5, TP7, P1, P3, P5, P7, PO3, PO7, O1, two additional positions PO9' located at 33 % and O9', located at 66 % of the M1–Iz distance, and corresponding right hemisphere electrodes. All ERP analyses were performed using EEGLAB (Delorme & Makeig, 2004), FieldTrip (Oostenveld et al., 2011) and custom MATLAB scripts. One analysis epoch was chosen for all analyses starting 1,500 ms before the onset of the critical tone and lasting 3 seconds. Off-line, all EEG channels were recalculated to an average reference and high-pass filtered (0.1 Hz, 30 dB/oct). Next, using a procedure similar to that by Nolan, Whelan, and Reilly (2010), (ocular) artifacts were removed and EEG data were corrected. The exact steps involved in this procedure are described next. A predefined  $z$ -score threshold of  $\pm 3$  was used for outlier identification (relating to channels, epochs, independent components, and single channels in single epochs). First, epochs containing extreme values in single electrodes (values larger than  $\pm 1000 \mu\text{V}$ ) were removed, as were trials containing values exceeding  $\pm 75 \mu\text{V}$  in multiple electrodes if independent of the EOG activity. Second,  $z$ -scored variance measures were calculated for all electrodes, and noisy EEG electrodes ( $z$  score  $> \pm 3$ ) were removed if their activity was uncorrelated to EOG activity. Third, this EEG data set was subjected to a spatial independent components analysis (ICA) (see infomax algorithm: Bell & Sejnowski, 1995). ICA components representing ocular activity were automatically identified using  $z$ -scored measures of the absolute correlation between the ICA component and the recorded hEOG and vEOG activity and confirmed by visual inspection before removal. Fourth, previously removed noisy

channels were interpolated using the average EEG activity of adjacent channels within a specified distance (4 cm,  $\sim 3$ –4 neighbors). For the event-related potential (ERP) analysis, the signal at each electrode site was averaged separately for each experimental condition, low-pass filtered (40 Hz, 36 dB/oct), and aligned to a 200 ms baseline before the onset of the critical tone. Visual inspection of the grand averaged waveforms identified six time windows of interest. The time windows are as follows: Window 1 (80–110 ms), Window 2 (180–220 ms), Window 3 (250–350 ms), Window 4 (375–425 ms), Window 5 (450–750 ms), and Window 6 (1,000–1,500 ms).

Statistical analyses were performed by means of analyses of variance (ANOVA), separately for ERP amplitudes at midline electrode and lateral electrode sites. For the midline analysis, three regions of interest (ROIs) were defined (anterior: FPz, AFz, Fz; central: FCz, Cz, CPz; posterior: Pz, POz, Oz). The lateral analysis followed procedures established in previous studies (e.g., Filik, Leuthold, Wallington, & Page, 2014). Lateral electrode sites were pooled to form eight ROIs, as recommended for the analysis of high-density electrode arrangements (e.g., Dien & Santuzzi, 2005) by dividing them along a left–right, an anterior–posterior, and a dorsal–ventral dimension. The four ROIs over the left hemisphere were defined as follows: left-anterior-ventral (AF7, F7, FT7, F5, FC5, C5, T7), left-anterior-dorsal (AF3, F3, FC3, F1, FC1, C1, C3), left-posterior-ventral (TP7, CP5, PO9', O9', P5, P7, PO7), and left-posterior-dorsal (CP3, CP1, P3, PO3, P1, O1). Four analogous ROIs were defined for homologue electrodes located over the right hemisphere. For all statistical analyses, the significance level was set to  $\alpha = .05$ .

**Dipole source analysis** We applied spatio-temporal source modeling to examine the brain sources producing the auditory N1 component in ERPs as well as the sources producing the match–mismatch effect—as visible in the difference waveforms—during target tone perception (see Fig. 4). To estimate the neural generators of N1 and mismatch-related ERP activity and the corresponding source waveforms, equivalent dipole sources were determined using BESA (Version 5.18; Scherg & Berg, 1991) with the four-shell spherical head model (cf. Berg & Scherg, 1994). Source locations are reported in approximate Talairach coordinates. The four-shell spherical head model specified brain, skull, cerebrospinal fluid (CSF), and scalp (cf. Berg & Scherg, 1994). The head radius was set to 92.5 mm; conductivity ratios and thickness (in parentheses) were as follows: brain = 0.3300, skull = 0.0042 (7 mm), scalp = 0.3300 (6 mm), and CSF = 1 (1 mm). The regularization constant was set to 2 % and the minimum distance criterion to 8.0 mm.

To maximize the signal-to-noise ratio, the dipole-source model was derived for band-pass filtered (0.1–40 Hz) grand average (across participants) N1 waveforms ( $1/2[\text{ERP}(t)\text{Mismatch plus}$

ERP(t)Match]) as well as grand average difference waveforms that resulted from the following subtraction: ERP(t)Mismatch minus ERP(t)Match. Principal component analysis (PCA) was used to estimate the minimal number of dipoles to be included in the model. Separate decomposition of an initial time interval around the pure tone (S1) in T2<sub>physical</sub> (-1,200 to -1,000 ms) and a later time interval (100–400 ms) around the target tone in the three tasks indicated that one PC, respectively, accounted for a large amount of the data variance in the two time intervals (> 85.4 %). A dipole pair was fitted for the early prime-time interval, and separately for each task, also for the target tone time interval at the time point where the difference between the match and the mismatch condition emerged (T1<sub>linguistic</sub> & T3<sub>imagery</sub>: 200–400 ms; T2<sub>physical</sub>: 100–300 ms).

## Results

Only trials where the participant responded correctly (i.e., responded “match” when the item was coded as a match and responded “mismatch” when the item was coded as a mismatch) were included in the ERP analysis. Overall, accuracy across all three tasks was ~94 % (T1<sub>linguistic</sub>: 94 %, T2<sub>physical</sub>: 98 %, T3<sub>imagery</sub>: 88 %). Analysis of accuracy rates was conducted with the factors task (T1<sub>linguistic</sub>, T2<sub>physical</sub>, T3<sub>imagery</sub>) and condition (match vs. mismatch). The ANOVA showed a main effect of task,  $F(2, 54) = 114.38, p < .001$ . Post hoc test indicated that accuracy differed among tasks, T1–T2:  $t(27) = 8.82, p < .001$ ; T2–T3:  $t(27) = 13.05, p < .001$ ; T1–T3:  $t(27) = 8.15, p < .001$ . There was also a main effect of condition,  $F(1, 27) = 89.78, p < .001$ , due to more errors in the match (90 % correct) than in the mismatch condition (97 % correct). The interaction between task and condition was also significant,  $F(2, 54) = 51.09, p < .001$ . The interaction was due to the match effect being significant in T1<sub>linguistic</sub>,  $t(27) = 8.68, p < .001$ , and T3<sub>imagery</sub>,  $t(27) = 8.53, p < .001$ , but only showing a trend in T2<sub>physical</sub>,  $t(27) = 1.96, p = .06$ .

Error trials and also the artefact-rejected trials were excluded from the averaging of condition-specific ERPs, leaving on average 69.22 trials per condition (out of 80; range = 38–80, median = 71). The ERP results are displayed in Figs. 2, 3, and 4. Across all conditions the sounds elicited the standard ERP component complex associated with physical sound processing, the initial central-maximal negativity (N1) was followed by a central-maximal positivity (P2) (e.g., Van Petten & Rheinfelder, 1995).

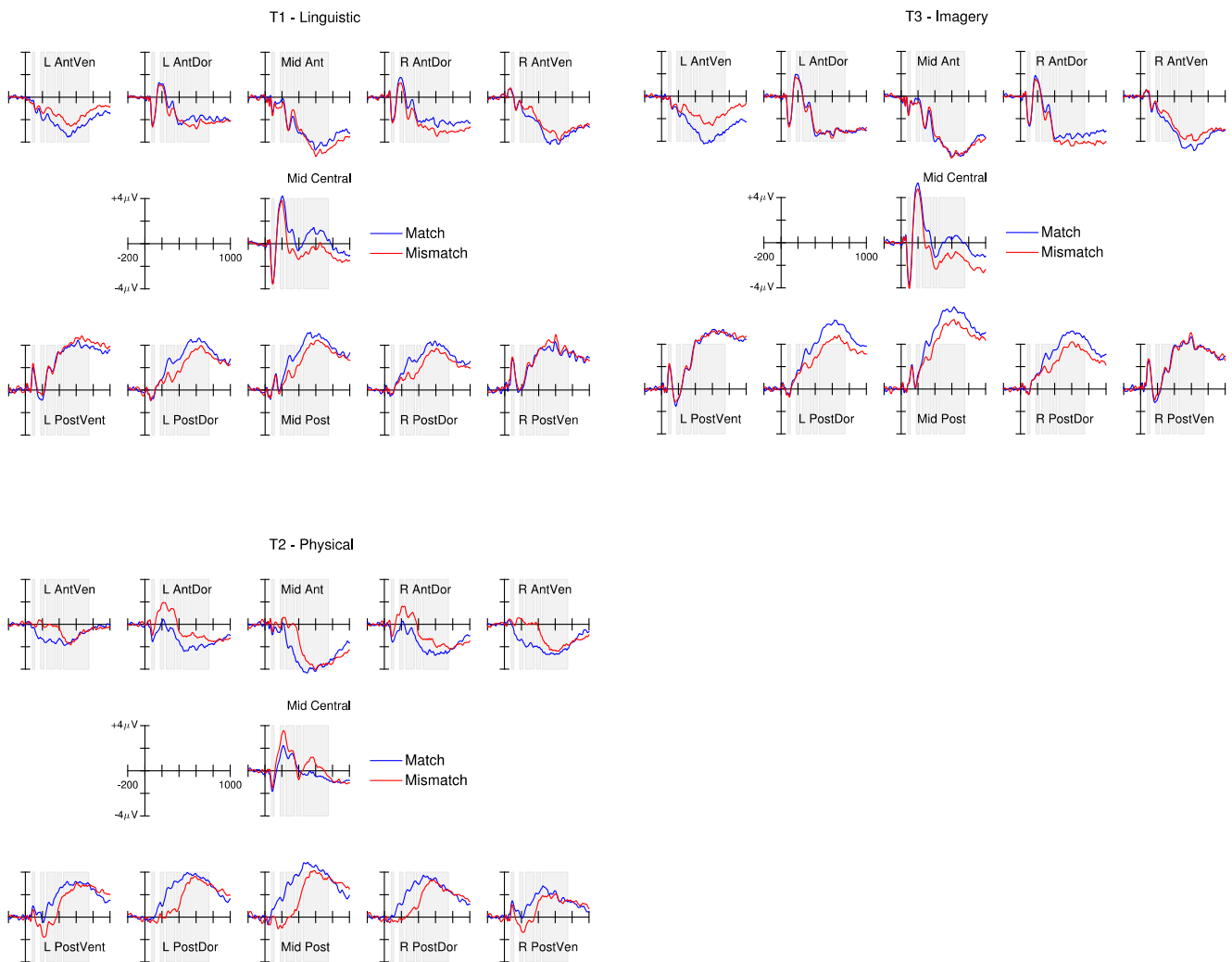
**Midline** For the midline analysis, a 2(compatibility: match vs. mismatch) \* 3(task: task T1<sub>linguistic</sub> vs. task T2<sub>physical</sub> vs. task T3<sub>imagery</sub>) \* 3(ROI: anterior vs. central vs. posterior) repeated-measures ANOVA was performed. The results are reported in the following paragraphs. Where appropriate, Huynh-Feldt corrected  $p$  values are reported. In order to check for influences of task order, we performed a further ANOVA with the

additional factor of the between-participant manipulation of task order (T1<sub>linguistic</sub> - T2<sub>physical</sub> - T3<sub>imagery</sub> vs. T3<sub>imagery</sub> - T2<sub>physical</sub> - T1<sub>linguistic</sub>). The factor task order did not affect any of the critical interactions between task and compatibility, suggesting that the compatibility effects within each task did not vary depending on whether the task was performed at the beginning or at the end of the experiment.

ERP amplitudes in the 80–110 ms time interval (N1) were largest at central electrodes, reflected in a main effect of ROI,  $F(2, 54) = 32.14, p < .001$ , with mean amplitudes over central sites (-2.73  $\mu$ V) being larger than those over anterior (-1.00  $\mu$ V) and posterior sites (0.48  $\mu$ V),  $F(1, 27) = 36.24, p < .001$  and  $F(1, 27) = 10.63, p < .01$ , respectively. The main effect of task was significant,  $F(2, 54) = 48.43, p < .001$ , due to smaller mean N1 amplitude for T2<sub>physical</sub> (-0.53  $\mu$ V) compared to both T1<sub>linguistic</sub> (-1.35  $\mu$ V),  $F(1, 27) = 67.32, p < .001$ , and T3<sub>imagery</sub> (-1.37  $\mu$ V),  $F(1, 27) = 67.78, p < .001$ , whereas T1<sub>linguistic</sub> and T3<sub>imagery</sub> amplitudes did not differ,  $F < 1$ , reflecting N1 component attenuation in T2<sub>physical</sub> (see Näätänen & Picton, 1987, for a review of the auditory N1). The significant two-way interaction between task and ROI,  $F(4, 108) = 12.24, p < .001$ , indicated the different topographic distributions for T1<sub>linguistic</sub> and T3<sub>imagery</sub> versus T2<sub>physical</sub> (see Fig. 3). Additionally, there was a trend for the three-way interaction between task, compatibility, and ROI,  $F(4, 108) = 2.68, p = .07$ , since in T2<sub>physical</sub> the compatibility effect in N1 amplitude was maximal at the occipital ROI (-0.65  $\mu$ V) and reversed at central (0.38  $\mu$ V) and frontal (0.32  $\mu$ V) electrode sides as indicated by the significant Compatibility  $\times$  ROI interaction (see Fig. 3),  $F(2, 54) = 6.64, p < .01$ , but this interaction was absent in T1<sub>linguistic</sub> ( $F < 1$ ) and T3<sub>imagery</sub> ( $F < 1$ ). This result suggests an earlier onset of the compatibility effect in T2<sub>physical</sub>.

Analysis of ERP amplitudes in the 180–220 ms time interval revealed the typical central P2 component distribution,  $F(2, 54) = 30.66, p < .001$ . Like for the N1 component, there was a main effect of task,  $F(2, 54) = 15.43, p < .001$ , due to component attenuation in T2<sub>physical</sub>. In contrast to N1, the main effect of compatibility was reliable for mean P2 amplitude,  $F(1, 27) = 16.11, p < .001$ , and this effect was maximal over central midline sites, as shown in an interaction with ROI,  $F(2, 54) = 11.40, p < .001$ . Additionally, the three-way interaction between compatibility, ROI, and task was significant,  $F(4, 108) = 20.40, p < .001$ . Post hoc tests showed a significant compatibility effect over midline electrode sites for both T1<sub>linguistic</sub>,  $F(1, 27) = 19.23, p < .001$ , and T3<sub>imagery</sub>,  $F(1, 27) = 12.79, p < .01$ , indicating its start during this time interval. Whereas for T2<sub>physical</sub> the negative-going mismatch effect over posterior sites was reversed in polarity over anterior sites, as indicated in an interaction between compatibility and ROI in T2<sub>physical</sub> only,  $F(2, 54) = 37.47, p < .001$ .

In the 250–350 ms time interval the compatibility effect in T1<sub>linguistic</sub> and T3<sub>imagery</sub> also interacted with ROI, T1<sub>linguistic</sub>:  $F(2, 54) = 13.90, p < .001$ ; T3<sub>imagery</sub>:  $F(2, 54) = 79.43, p < .001$ , due to an N400-like match–mismatch effect (see Fig. 3). Because the



**Fig. 2** ERP waveform for  $T1_{\text{linguistic}}$  (sentence–sound),  $T2_{\text{physical}}$  (sound–sound), and  $T3_{\text{imagery}}$  (imagery–sound) for the match (blue) and the mismatch (red) condition separate for the four lateral (anterior–

ventral, anterior–dorsal, posterior–ventral, posterior–dorsal) and the three central (anterior, central, posterior) ROIs. (Color figure online)

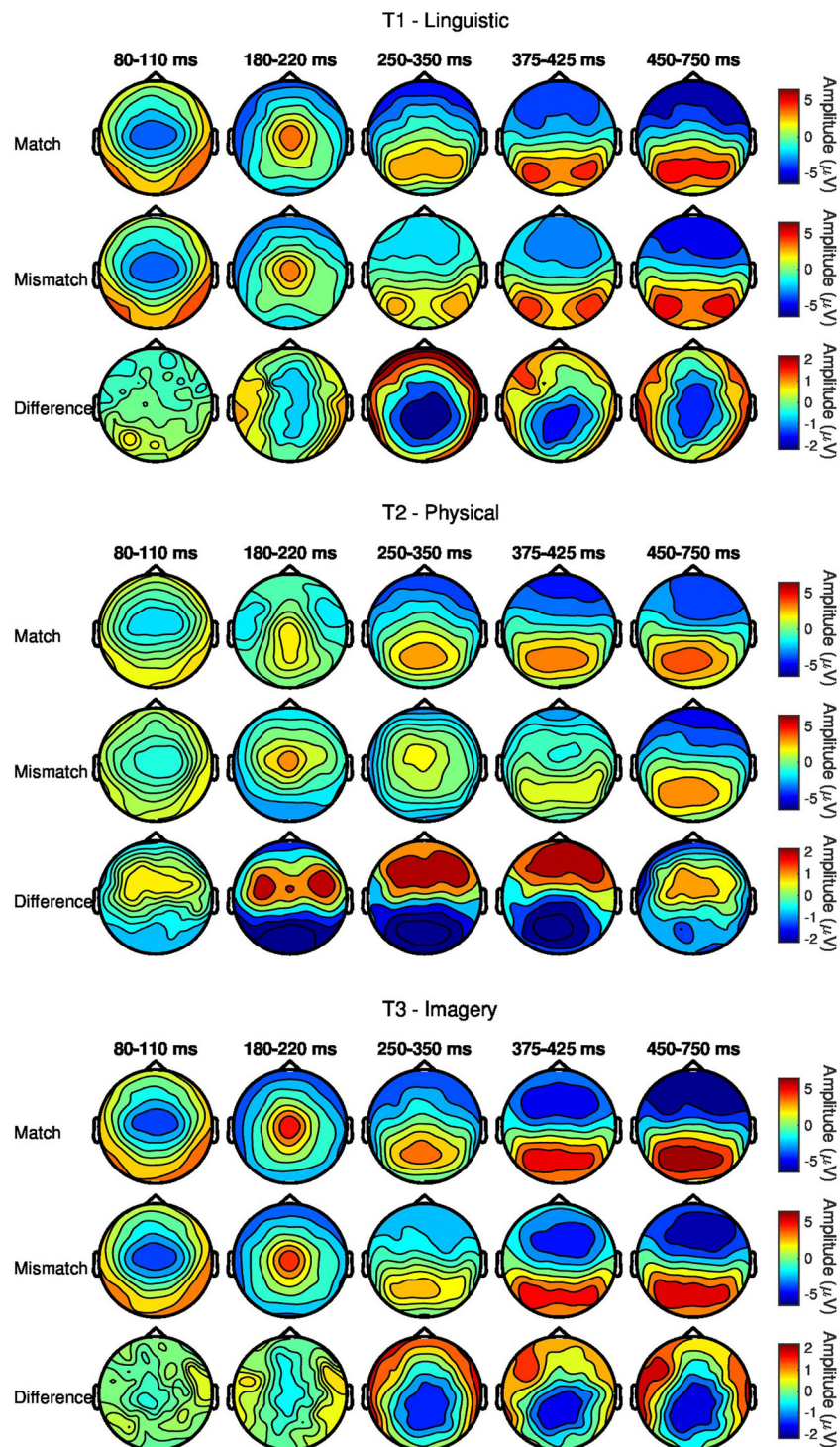
$T2_{\text{physical}}$  match–mismatch effect continued to be characterized by the anterior-positive and posterior-negative distribution, the distribution differences of the compatibility effect between the tasks persisted,  $F(4, 108) = 23.39$ ,  $p < .001$ , with post hoc tests showing that this was due to distribution differences between  $T2_{\text{physical}}$  versus  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$ . The main differences between  $T2_{\text{physical}}$  versus  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$  persisted throughout the subsequent time intervals (375–425 ms, 350–750 ms, 1000–1500 ms).

In summary, the midline analyses showed an earlier onset of the compatibility effect in  $T2_{\text{physical}}$  (80–110 ms) than in  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$  (180–220 ms). Additionally, the distribution of the compatibility effect indicated clear dissociations between  $T2_{\text{physical}}$  versus  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$ . Whereas  $T2_{\text{physical}}$  shows an anterior–posterior distribution of the match–mismatch effect, the  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$  conditions show an N400-like component with a central maximum (see Orgs et al., 2006, 2007, 2008; Van Petten & Rieffers, 1995). Additional

analysis of the difference waveforms (mismatch minus match) supported these findings (see Appendix C, Figure 12). The analysis of the  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$  difference waveforms did neither show a main effect of task at any time interval ( $F_s < 1.10$ ,  $p_s > .30$ ), nor an interaction of task and ROI ( $F_s < 1.64$ ,  $p_s > .21$ ). Only the main effect of ROI was significant starting in the 250–350-ms time interval,  $F(2, 54) = 11.63$ ,  $p < .001$ , due to larger central negativity as expected for N400-like components. The analysis of the difference waveforms of  $T1_{\text{linguistic}}$  and  $T2_{\text{physical}}$  showed a significant interaction between task and ROI starting in the 80–110 ms time interval,  $F(2, 54) = 4.85$ ,  $p < .05$ . The analysis of the difference waveforms for  $T2_{\text{physical}}$  and  $T3_{\text{imagery}}$  showed an interaction between task and ROI starting in the 180–220 ms time interval,  $F(2, 54) = 26.68$ ,  $p < .001$ .

**Lateral Analysis** For the lateral analysis, a 2 (compatibility: match vs. mismatch) \* 3 (task: Task 1 vs. Task 2 vs. Task 3) \* 2 (VentDor: ventral vs. dorsal) \* 2 (Hemi: left vs. right) \*





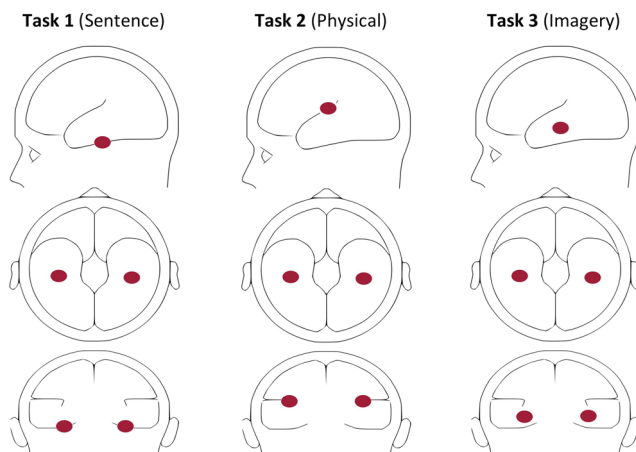
**Fig. 3** Topographic plots  $T1_{\text{linguistic}}$  (sentence–sound),  $T2_{\text{physical}}$  (sound–sound), and  $T3_{\text{imagery}}$  (imagery–sound) for the match, mismatch, and difference (mismatch minus match) waveforms separately for the five time intervals. (Color figure online)

2(AntPos: anterior vs. posterior) repeated-measures ANOVA was performed. Where appropriate, Huynh-Feldt corrected  $p$  values are reported. Again, the between-participant manipulation of task order ( $T1_{\text{linguistic}} - T2_{\text{physical}} - T3_{\text{imagery}}$  vs.  $T3_{\text{imagery}} - T2_{\text{physical}} - T1_{\text{linguistic}}$ ) did not affect any of the critical interactions between task and compatibility, suggesting that the compatibility effects

did not vary depending on whether a task was performed at the beginning or at the end of the experiment. Thus, the factor task order was excluded in the reported ANOVA.

The main results from the lateral analysis support the findings from the midline analysis. In the 80–110 ms time interval (N1), ERP amplitude was larger over left than right hemispheric ROIs





**Fig. 4** Dipoles fitted to the difference waveforms (mismatch minus match) at the interval where the difference started to become significant separately for  $T1_{\text{linguistic}}$ ,  $T2_{\text{physical}}$ , and  $T3_{\text{imagery}}$

( $-0.36$  vs.  $0.03$   $\mu\text{V}$ ),  $F(1, 27) = 10.11$ ,  $p < .01$ , over anterior than posterior ROIs ( $-1.00$  vs.  $0.61$   $\mu\text{V}$ ),  $F(1, 27) = 25.61$ ,  $p < .001$ , and dorsal than ventral ROIs ( $-1.13$  vs.  $0.75$   $\mu\text{V}$ ),  $F(1, 27) = 105.37$ ,  $p < .001$ . Also the interaction between hemisphere and the ventral–dorsal axis was significant,  $F(1, 27) = 16.24$ ,  $p < .001$ . More interestingly, there was a significant interaction between task and hemisphere,  $F(2, 54) = 4.74$ ,  $p < .05$ , due to an early lateralization in  $T1_{\text{linguistic}}$ ,  $F(1, 27) = 17.14$ ,  $p < .001$ , and a trend for this interaction in  $T3_{\text{imagery}}$ ,  $F(1, 27) = 3.41$ ,  $p = .08$ , but not in  $T2_{\text{physical}}$ ,  $F(1, 27) = 2.64$ ,  $p = .12$ . There were significant interactions between task and the anterior–posterior axis,  $F(2, 54) = 10.78$ ,  $p < .001$ , as well as between task and the ventral–dorsal axis,  $F(2, 54) = 49.72$ ,  $p < .001$ . Also the three-way interaction of task, anterior–posterior, and ventral–dorsal axis was significant,  $F(2, 54) = 11.82$ ,  $p < .001$ , due to larger task-related effects in terms of N1 amplitude attenuation in  $T2_{\text{physical}}$  compared to both  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$  over anterior–dorsal and posterior–ventral than anterior–ventral and posterior–dorsal ROIs. The three way interaction between task, compatibility, and the anterior–posterior axis,  $F(2, 54) = 3.88$ ,  $p < .05$ , was due to an early developing anterior–posterior distribution difference in matching versus mismatching trials in  $T2_{\text{physical}}$ ,  $F(1, 27) = 12.81$ ,  $p < .01$ , but not  $T1_{\text{linguistic}}$  ( $F = 1.65$ ) or  $T3_{\text{imagery}}$  ( $F < 1$ ).

In the 180–220 ms time interval (P2), mean ERP amplitudes showed a positivity over dorsal ROIs that turned into a negativity over ventral ROIs; and this effect was more pronounced over anterior sites as indicated by the significant interaction of the anterior–posterior axis with the ventral–dorsal axis,  $F(1, 27) = 18.59$ ,  $p < .001$ . The main effects of task,  $F(2, 54) = 22.89$ ,  $p < .001$  ( $T1_{\text{linguistic}}$ :  $0.23$   $\mu\text{V}$ ,  $T2_{\text{physical}}$ :  $0.05$   $\mu\text{V}$ ,  $T3_{\text{imagery}}$ :  $0.25$   $\mu\text{V}$ ), and of compatibility were significant,  $F(1, 27) = 8.98$ ,  $p < .01$ , as well as the interaction between task and compatibility,  $F(2, 54) = 8.51$ ,  $p < .001$ . Post hoc tests showed that the compatibility effect was significant for  $T2_{\text{physical}}$ ,  $F(1, 27) = 14.98$ ,  $p < .001$ , but not  $T1_{\text{linguistic}}$  ( $F < 1$ ) or  $T3_{\text{imagery}}$  ( $F < 1$ ). Task interacted with hemisphere,  $F(2, 54) = 5.55$ ,  $p < .01$ , due to larger P2 amplitudes

over right- than left-hemispheric ROIs in  $T1_{\text{linguistic}}$ ,  $F(1, 27) = 9.73$ ,  $p < .01$ , but not in  $T2_{\text{physical}}$  ( $F < 1$ ) or  $T3_{\text{imagery}}$  ( $F < 1$ ). The two-way interaction between compatibility and the anterior–posterior axis was significant,  $F(1, 27) = 23.34$ ,  $p < .001$ , as well as the two-way interaction between compatibility and the ventral–dorsal axis,  $F(1, 27) = 14.91$ ,  $p < .001$ , reflecting larger compatibility effects over anterior than posterior and ventral than dorsal ROIs. Moreover, there was a trend for the interaction between task, compatibility, and the ventral–dorsal axis,  $F(2, 54) = 2.53$ ,  $p = .09$ , and a significant four-way interaction between task, compatibility, the anterior–posterior axis, and the ventral–dorsal axis,  $F(2, 54) = 6.21$ ,  $p < .01$ . This result indicates that the size of the match–mismatch effect varied as a function of task across lateral ROIs. Further tests showed a reliable interaction between compatibility and anterior–posterior axis only for  $T2_{\text{physical}}$ ,  $F(2, 54) = 58.82$ ,  $p < .001$ , whereas the interaction between compatibility and the ventral–dorsal axis was significant for  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$ ,  $F(2, 54) = 15.11$ ,  $p < .001$  and  $F(2, 54) = 14.07$ ,  $p < .001$ , respectively. The key differences between  $T2_{\text{physical}}$  versus  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$  reported here persisted throughout the subsequent time intervals.

Analysis of the difference waveforms confirmed these results. When comparing  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$ , the earliest task difference showed in the 250–350-ms time interval,  $F(1, 27) = 17.06$ ,  $p < .001$ , due to a stronger compatibility effect in  $T1_{\text{linguistic}}$  than in  $T3_{\text{imagery}}$ . In the 450–750 ms time interval there was an interaction between task and the anterior–posterior factor,  $F(1, 27) = 4.43$ ,  $p < .05$ , and a significant interaction between task, hemisphere, and the anterior–posterior axis,  $F(1, 27) = 6.82$ ,  $p < .05$ . In contrast, when comparing  $T1_{\text{linguistic}}$  and  $T2_{\text{physical}}$ , the task difference already showed in the 80–110 ms time interval with the interaction between task and the anterior–posterior axis,  $F(1, 27) = 8.20$ ,  $p < .01$ . Similarly, when comparing  $T3_{\text{imagery}}$  and  $T2_{\text{physical}}$ , there was an interaction between task and the ventral–dorsal axis,  $F(1, 27) = 4.31$ ,  $p < .05$ , and a trend for the interaction between task and the anterior–posterior axis,  $F(1, 27) = 2.99$ ,  $p = .10$ , in the 80–110 ms time interval.

**Permutation Tests** The standard ERP analysis was extended by a permutation procedure. Cluster-based permutation tests were performed across 64 scalp electrode sites within the time interval starting from sound onset (0–1,000 ms poststimulus) with the constraint that at least two adjacent channels show a significant effect (see Maris & Oostenveld, 2007). These Montecarlo cluster-based permutation tests ( $N = 1,000$ ) are specifically designed to deal with multiple comparisons problems resulting from dense electrode layouts and multiple time-point testing. This procedure was applied to the difference waveforms comparing  $T1_{\text{linguistic}}$  with  $T2_{\text{physical}}$ ,  $T1_{\text{linguistic}}$  with  $T3_{\text{imagery}}$ , and  $T2_{\text{physical}}$  with  $T3_{\text{imagery}}$ . The results are displayed in Appendix A (Figure 6, Figure 7, and Figure 8). In summary, the results fully mirrored the ERP results;  $T2_{\text{physical}}$  shows early differences from

T1<sub>linguistic</sub> and T3<sub>imagery</sub>, whereas T1<sub>linguistic</sub> and T3<sub>imagery</sub> do not differentiate at any time point.

**Dipole Analysis** The fitted  $x$ ,  $y$ , and  $z$  coordinates were compared between the three tasks and the pure tone source location. The resulting dipole pair in T1<sub>linguistic</sub>, T2<sub>physical</sub>, and T3<sub>imagery</sub> was located in the temporal cortex (T1<sub>linguistic</sub>:  $x = 0.10$  mm,  $y = -0.03$  mm,  $z = -0.08$  mm; T2<sub>physical</sub>:  $x = 0.08$  mm,  $y = -0.00$  mm,  $z = 0.12$  mm, T3<sub>imagery</sub>:  $x = 0.02$  mm,  $y = -0.04$  mm,  $z = -0.07$  mm), with each task-dependent model accounting for a large proportion of the data variance, T1<sub>linguistic</sub>: RV (residual variance) = 6.15 %; T2<sub>physical</sub>: RV = 13.12 %; T3<sub>imagery</sub>: RV = 8.05 %. First, the source location of the pure tone source (measured in the time interval of component onset) and the source of the T2<sub>physical</sub>-difference (measured in the 40 ms interval after the onset of the match–mismatch effect) did not differ,  $x$ :  $t(27) = 0.93$ ,  $p = .36$ ;  $y$ :  $t(27) = 0.47$ ,  $p = .64$ ;  $z$ :  $t(27) = -1.62$ ,  $p = .12$ , suggesting that the match–mismatch effects measured on physical tone processing indeed are able to reflect sensory processes, probably originating in the primary and secondary auditory cortex (see Figure 4). Second, the source fitted to the T1<sub>linguistic</sub>-difference wave and the T3<sub>imagery</sub>-difference wave did not differ significantly,  $x$ :  $t(27) = 0.57$ ,  $p = .58$ ;  $y$ :  $t(27) = .13$ ,  $p = .90$ ;  $z$ :  $t(27) = -0.01$ ,  $p > .99$ . Finally, the sources of the T1<sub>linguistic</sub> and T2<sub>physical</sub> difference waves and of the T3<sub>imagery</sub> and T2<sub>physical</sub>-difference waves were compared. In both cases the source dipole was differently located on the  $z$  axis, T1<sub>linguistic</sub>-T2<sub>physical</sub>:  $x$ :  $t(27) = -0.15$ ,  $p = .88$ ;  $y$ :  $t(27) = .42$ ,  $p = .68$ ;  $z$ :  $t(27) = 4.39$ ,  $p < .001$ ; T3<sub>imagery</sub>-T2<sub>physical</sub>:  $x$ :  $t(27) = 0.38$ ,  $p = .70$ ;  $y$ :  $t(27) = 0.56$ ,  $p = .58$ ;  $z$ :  $t(27) = 3.23$ ,  $p < .01$ , suggesting the source of the match–mismatch effect in the sentence and the imagery condition to have underlying semantic-conceptual neural networks in the inferior temporal cortex.

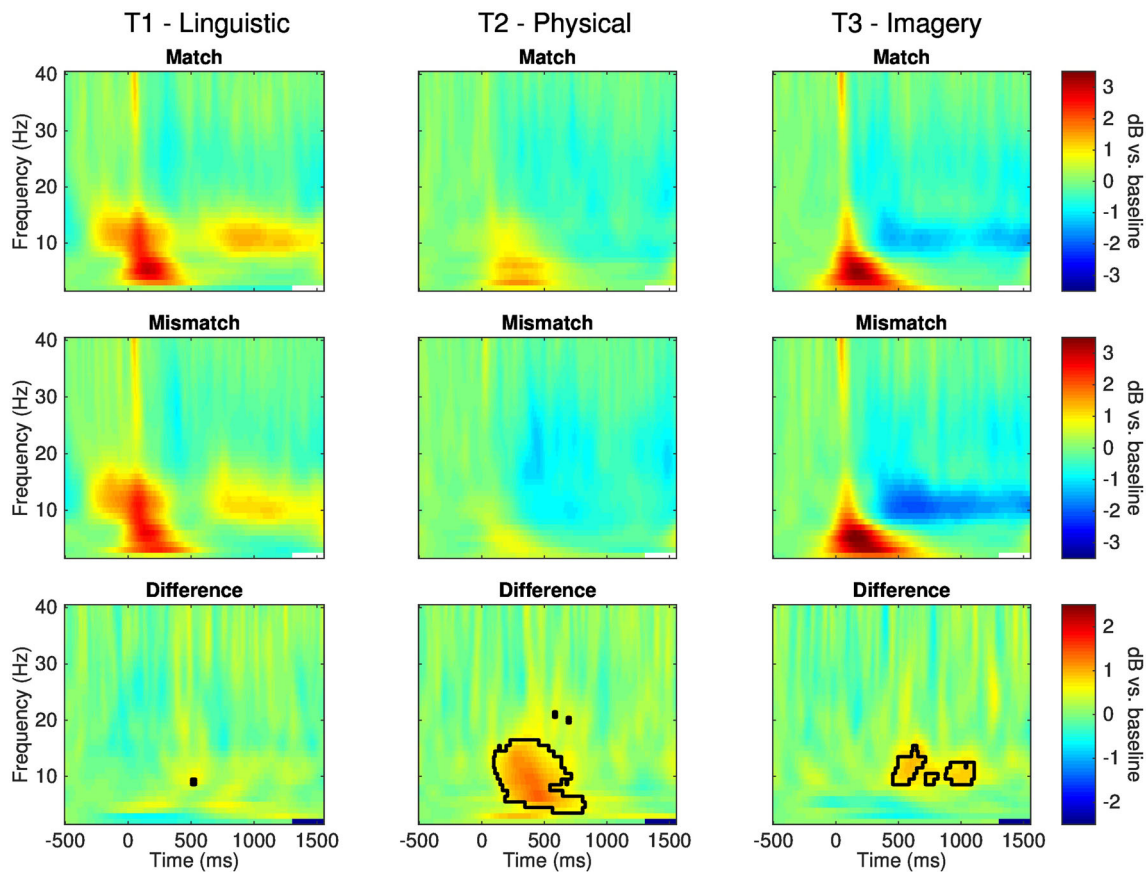
**Time Frequency Analysis** In all three tasks (linguistic, physical, imagery), oscillatory brain activity was analyzed during sound processing because non-phase-locked dynamic activity might be not observable in the average ERP but still be task modulated (see Fig. 5). Time-frequency representations were calculated using Morlet wavelets, as described previously (Tallon-Baudry, Bertrand, Delpuech, & Pernier, 1997), with linear spaced frequencies from 1 to 40 Hz in 1 Hz steps. The cycle width was set to 5, as this value has been shown to give a good tradeoff between time and frequency resolution (De Moortel, Munday, & Hood, 2004). The averaged power values were subsequently converted to a decibel (dB) scale relative to a prestimulus baseline from -500 to -300 ms. The difference values were obtained by subtraction between the match and the mismatch condition without further baseline correction. The time-frequency data were statistically compared using nonparametric cluster-based permutation  $t$  tests (Maris & Oostenveld, 2007) to control for Type I error rate in the context of multiple comparisons. Cluster of significant differences over electrode (72 electrodes), time (0–1.5 s),

and frequency (1–40 Hz) were identified. Cluster-level statistics were defined as the sum of  $t$  values within each cluster and were evaluated under the permutation distribution of the max/min cluster-level statistic. The distribution was obtained by  $N = 1,000$  random permutations of conditions within participants.

Across all three tasks, standard theta and alpha power change was observed during sound perception starting with sound presentation. This is in line with previous studies, suggesting that oscillatory and evoked activity contribute to the auditory N1 complex in nonattenuated conditions (see Fuentemilla et al., 2006). Only in T2<sub>physical</sub> the theta and alpha band power change was highly reduced in the match condition, also visible in the analysis of the difference time-frequency representation between the match and the mismatch condition. This finding is in line with previous studies on auditory attenuation processes (Fuentemilla et al., 2006). Subsequent between-task comparisons were performed by comparing the three difference representations in a permutation-based manner (Maris & Oostenveld, 2007). In line with the ERP results, this analysis showed no difference between T1<sub>linguistic</sub> and T3<sub>imagery</sub>. However, both T1<sub>linguistic</sub> and T3<sub>imagery</sub> differed significantly from T2<sub>physical</sub> (see Appendix B, Figures 9, 10, and 11 for visual displays of these analyses).

## Discussion

Research in the area of language comprehension has recently developed influential models suggesting that embodied—sensorimotor-like—representations are crucial during comprehension (Barsalou, 1999). The literature on imagery—a basic form of thought in the absence of the actual physical stimulus—has emphasized the role of sensory processes during mental imagery. These two research areas have developed rather independently, and they sometimes try to emphasize that the underlying processes differ (e.g., Brunyé et al., 2010). Especially in the area of embodied language comprehension, it is often stressed that sensorimotor representations during language comprehension are not the result of an active imagery process but occur rather automatically during language processing. Brunyé et al. (2010) investigated the influence of a sentence (e.g., “The waves crashed against the beach as the tide began to rise.”) on subsequent sound perception. They found that sound classification was faster if the sound was preceded by a matching sentence compared to a mismatching sentence. Critically, it remained open whether the effect of the sentence on subsequent physical sound classification is due to sensory preactivation or is triggered by other nonsensory processes. In our study, we directly compared the influence of active imagery, sentence processing, and sound perception on subsequent sound processing. In particular, we investigated whether the influence of language on subsequent sound processing reflects sensory, perceptual-like processes (see Brunyé et al., 2010), or alternatively whether during reading similar representations become available as during active imagery.



**Fig. 5** Time-frequency activity for  $T1_{\text{linguistic}}$ ,  $T2_{\text{physical}}$ , and  $T3_{\text{imagery}}$  at the Cz electrode. The top panels display the activity for the match conditions, followed by the activity in the mismatch conditions in the

middle panels and the difference (mismatch minus match) representation in the bottom panels. In the bottom panels statistically significant differences are highlighted. (Color figure online)

The results of this study are clear cut with regard to these open research issues: Sentence processing and active sound imagery both influence subsequent sound processing in a similar manner. Effects on sound processing were observed as early as 180–220 ms in both the linguistic ( $T1_{\text{linguistic}}$ ) (see Orgs et al., 2006, 2007, 2008; Van Petten & Rheinfelder, 1995) and the imagery ( $T3_{\text{imagery}}$ ) condition. However, these effects clearly differentiate from the effect observed in the physical control condition ( $T2_{\text{physical}}$ ) that served as a perceptual baseline condition. First, in the physical control condition, the influence of initial sound processing on subsequent sound processing was observed as early as 80–110 ms after stimulus onset. Second, an N1/P2 component attenuation—typically attributed to perceptual habituation processes (Näätänen & Picton, 1987)—was obtained only in this sound–sound condition. Dipole source localization of the match–mismatch effect shows clear activation in the sensory auditory cortex in the physical control condition, and the location of the dipole is actually identical to the one for the N1/P2 complex of the first tone presentation, suggesting that the match–mismatch effect relates to sensory processes of actual sound perception. In contrast, the effects observed in the linguistic and the imagery conditions are localized in inferior temporal cortex areas, probably because of semantic-conceptual representations being

involved in causing these effects (Fernandino et al., 2016; Lau, Philips, & Poeppel, 2008; Nobre, Allison, & McCarthy, 1994). Finally, analysis of the oscillatory activity showed clear similarities between the imagery ( $T3_{\text{imagery}}$ ) and the linguistic ( $T1_{\text{linguistic}}$ ) condition, which directly contrast with the physical control condition. In line with previous studies on sound processing and attenuation effects in the time-frequency analysis (e.g., Fuentemilla et al., 2006), there were clear theta and alpha power changes in the physical ( $T2_{\text{physical}}$ ) condition but not in the linguistic ( $T1_{\text{linguistic}}$ ) or the imagery ( $T3_{\text{imagery}}$ ) condition.

So why are there no sensory effects in our study, either in the linguistic or in the imagery condition? Our study investigates the influence of a sentence describing an auditory event and of a mental image of a sound on subsequent auditory stimulus processing. First, one could argue that the specific sensory representations (e.g., the neural activation when reading the sentence *The dog barks* that is similar to the activation when hearing a dog barking) might be present during imagery or during sentence comprehension, but then quickly decay until the presentation of the auditory stimulus. However, if there exist specific sensory representations similar to the ones present in the physical control condition, it is unlikely that the activation decays so quickly that it does not influence subsequent sensory



processes (Kiefer et al., 2008). In the physical control condition, clear sensory influences can be observed on target sound processing despite the sensory stimuli being separated by a 700 ms interval, identical to that used in the sentence condition. The influences measured in the imagery and the sentence conditions rather reflect N400-like conceptual-semantic influences—with regard to topography and timing—on sound processing (see Lau et al., 2008). As a second argument—with regard to the missing sensory effect in the imagery and linguistic conditions—one might suggest that our physical control condition ( $T2_{\text{physical}}$ ) was rather strong, because in the match condition we presented in succession two identical sounds rather than two physically different but conceptually related sounds (e.g., two different barking sounds). The choice of a physical sound repetition condition was motivated by the present goal to reveal a potential sensory match–mismatch effect in ERPs and its localization within the brain. It is obvious that localizing pure sound activation should result in sensory activation (e.g., Giard et al., 1994; Leaver & Rauschecker, 2010; Lewis et al., 2004), however, there are no a priori reasons why the physical match–mismatch effect would be generated alike at this sensory level. Strong versions of embodiment accounts would though suggest that language comprehension indeed activates such primary sensory representations (see Binder & Desai, 2011; Meteyard & Vigliocco, 2008). Hence, applying the dipole source localization method to physical match–mismatch ERP difference waveforms then allows testing whether this effect is generated within the same primary sensory brain area as those activated by single sounds. Thereby, it becomes possible to determine whether the physical match–mismatch effect indeed reflects a low-level perceptual effect. When using conceptually related sounds (e.g., two different barking sounds), this condition likely reflects rather nonsensory brain activity (see Aramaki, Marie, Kronland-Martinet, Ystad, & Besson, 2010), and thus would have not provided a strict control condition in order to analyze the purely sensory contributions to our match–mismatch effects. Of course, such a conceptual priming condition would come with the advantage of avoiding the stimulus repetition effect, although it would still include a modality repetition effect, which are both neither present in our linguistic nor in the imagery condition.

Of course, it would be interesting to know how conceptually related sounds influence sound processing. Fortunately, in a study currently conducted in our lab, we directly investigate this issue using a priming paradigm very similar to the one in this work. Preliminary results suggest that conceptually related physical sounds also result in a match–mismatch effect on the target sound. These effects occur rather late ( $> 300$  ms) and, as in the study of Aramaki et al. (2010), seem to reflect N400-like effects, as observed in the imagery and the sentence condition of the present study. This suggests that not only in conditions where a modality switch takes place (imagery–sound, language–sound) but also when staying within the same modality (sound–sound) conceptual N400-like effects

are observed. However, as the spatial resolution of source localization based on electrophysiological data is limited, future studies should be conducted to more precisely localize the brain sources contributing to the match–mismatch effect after imagery, sentence, or sound processing.

Another issue that needs to be discussed with regard to the missing sensory effects in  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$  is the fact that two specific task orders were used: Half of the participants first performed  $T1_{\text{linguistic}}$ , followed by  $T2_{\text{physical}}$  and finally  $T3_{\text{imagery}}$ . The other half of the participants performed the experiment in the reversed order starting with  $T3_{\text{imagery}}$ , followed by  $T2_{\text{physical}}$  and finally  $T1_{\text{linguistic}}$ . We choose these two versions as we aimed at separating  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$  as far as possible within each participant to avoid potential carryover effects. An additional analysis of task order also showed that the compatibility effects were not modified within each condition if presented at the start or at the end of the experiment. Still, one might ask whether the sensory effects in  $T2_{\text{physical}}$  might be due to the constant middle position of  $T2_{\text{physical}}$  in the experiment. Participants at that point might have familiarized with the listening task but not yet suffered from fatigue, whereas this does not hold true for  $T1_{\text{linguistic}}$  and  $T3_{\text{imagery}}$  conditions. This is because at the start of the experiment participants might be unfamiliar with the listening task, while at the end potentially suffering from fatigue, thereby differentially influencing these two task conditions. At the moment we cannot fully exclude this alternative explanation; however, given present ERP findings and those from conceptual sound priming experiments outlined above, we suggest that the cause of the sensory, early effects in  $T2_{\text{physical}}$  are more likely due to the physical nature of the task rather than to task order effects.

Of course, it should be noted that our study does not aim at fully excluding any sensory process involved during language comprehension. Kiefer et al. (2008) clearly demonstrated in an fMRI study activation in the primary and secondary sensory cortex during reading words that are highly associated with acoustic features (see also Fernandino et al., 2016). Also, when looking at ERPs, Kiefer et al. reported a dissociation between words associated with acoustic features and words that have no acoustic association. Specifically, acoustic-related words influenced ERP amplitudes in a time range of 150–200 ms after word onset at topographic regions that are typically associated with acoustic processing (Näätänen, 1992). Furthermore, source analyses identified standard auditory brain regions being involved in causing this effect. In this comparison, it is important to note that word-class effects were of central interest to the authors rather than the influence of language processes on subsequent nonlinguistic—in our case, perceptual auditory—processes.

Thus, our study clearly indicates that in studies investigating the influence of language on subsequent auditory processing, these influences seem not purely sensory in nature. Recent studies suggested that task demands might play an important role with regard to whether sensory activation is present during



language comprehension (Areshenkoff, Bub, & Masson, 2017; Dudschig & Kaup, *in press*; Lebois, Wilson-Mendenhall, & Barsalou, 2014). Because in our experiment participants had to process an auditory stimulus and directly compare it to the sentence or their mental image of a sound, we would argue that our task clearly demands perceptual activation. On the other side, one might argue that linguistic labeling might play a role in environmental sound recognition. The fact that our results resemble N400-like effects during sound perception might lead to the suggestion that the sounds are labeled and the outcome of the labeling triggers the match–mismatch effects. However, the fact that the effects in the linguistic and the imagery condition emerge as early as 180–220 ms following sound onset—and therefore earlier than typical N400 effects measured on words (e.g., Orgs et al., 2006)—makes it rather unlikely that additional, time-consuming labeling steps took place. Moreover, previous studies investigating sound recognition suggested that labeling adds additional processing costs and is usually not used by participants because it does not improve the sound recognition process (Dick et al., 2002).

Another point that needs to be discussed is the role of the participants' task employed in this study. Participants had to decide after every sound presentation whether the sound was a matching or a mismatching sound. This task was used to ensure that participants had to process both the first and the second stimulus across all three tasks. With regard to the sentence condition, one might argue that our task enforced participants in performing imagery processes. Thus, the sentence condition might have resulted in a similar pattern as the imagery condition, whereas in other task setups these effects might not occur automatically during sentence reading. However, in previous studies by Orgs et al. (2007, 2008), it was shown that cross-modal match–mismatch effects occur under various task conditions, even when participants only had to indicate whether a sound was presented on the left or the right side. Thus, we would argue that it is not the specific task that resulted in the observed effects in the sentence task. Additionally, we used a rather short interval between sentence offset and subsequent sound presentation (700 ms) that is probably not long enough for active imagery to occur. Indeed, Intons-Peterson (1980) investigated the time required to build an internal auditory image of a named but unheard sound. On average participants took longer than 1 second to form a clear internal image of the described sounds. This timing was independent of the typical loudness of the according real sound. In the light of these findings it seems unlikely that our participants deliberately constructed an internal image of the described sound in the sentence condition, especially as there was no instruction to do so. In contrast, the use of the long interstimulus interval in the imagery condition ensured that participants had sufficient time to build an internal image of the described sound. There was also no effect of block order on the compatibility effects observed in either of the tasks. Even in the experimental version, where participants first performed the sentence

condition, the compatibility effects were not reliably different from those observed in the final imagery condition. Thus, we can exclude the possibility that our results are only due to a carryover effect in such a way that participants imagined the sound in the sentence condition because they were doing it in the preceding imagery condition. Of course, we have no definite check for participants' compliance in the imagery condition. However, the reasons outlined above—especially the increased time interval with no decay of the compatibility effect—let us suggest that participants complied with the task instruction. Also, the increased error rates in the imagery condition, especially in the match condition, might point toward the fact that participants constructed an internal image of the sound and subsequently compared the presented sound to this internal image (and rejected it as a mismatch if it did not fully match their internal representation).

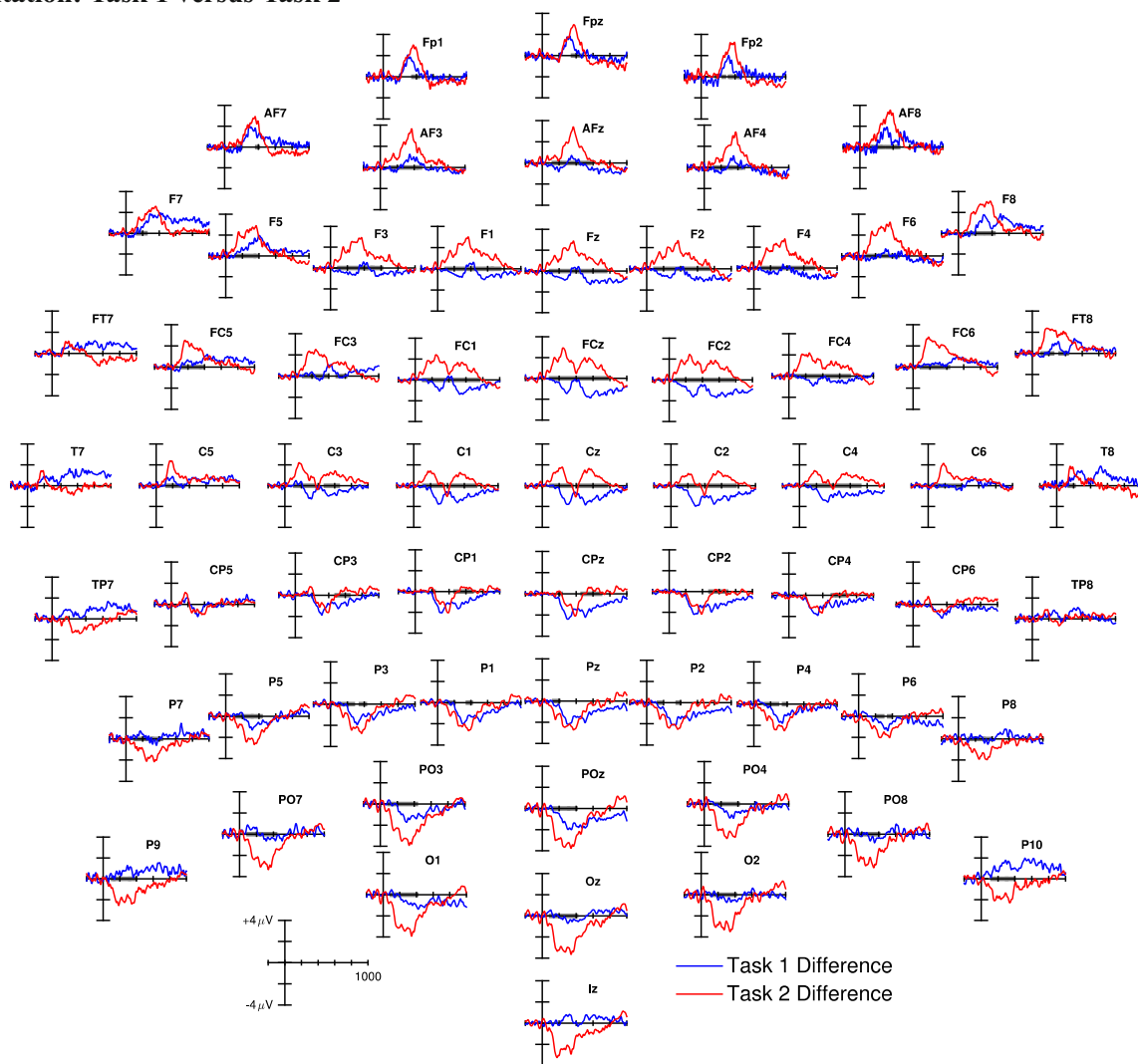
In summary, our study brings together research from imagery and language comprehension and suggests that during active imagery and language comprehension, similar representations and processes become involved. With regard to the first research question—addressing the issue of the similarity of the activations triggered by sound imagery and sentences describing auditory events—the results of this study are rather clear-cut and suggest that both types of events result in similar activations with comparable effects on subsequent physical sound processing. With regard to the second question, addressing the issue of the actual types of representation activated by sound imagery or sentence processing, this study points toward the involvement of higher semantic conceptual networks—beyond unimodal sensory processes. However, there are several possibilities how such higher level activations might look. First, sensory representations might play a role during the comprehension of language describing sound events; however, these representations might not be solely driven by primary sensory regions. Thus, in contrast to strong versions of the embodiment account—emphasizing the role of primary sensory activations—our findings suggest that both imagery and the processing of sentences describing auditory events result in the activation of higher level abstract representations. Indeed, a recent account of language comprehension by Fernandino et al. (2016) points toward the role of gradual abstraction during comprehension even for sensory representations (see also Kemmerer, 2015). Fernandino and colleagues suggested that the meaning representation of words associated with auditory events is less associated with primary, as compared to secondary, sensory regions in the brain. This is different for other sensory aspects of word meaning, such as visual aspects (e.g., color, shape), which are associated with primary sensory activation. Our findings support this assumption by indicating that both sound imagery and processing of sentences describing auditory events do not result in effects on subsequent sound processing that are clearly driven by primary sensory regions. In contrast

to such a potentially still sensory—although higher level sensory—representation activation, it also might be the case that our observed influence of sound imagery and of sentences describing sounds on subsequent physical sound processing might be driven by fully amodal representations (e.g., Machery, 2016; Pylyshyn, 1973) or even a hybrid version of meaning representation that allows both sensory modal and amodal aspects to be activated during comprehension (e.g., Binder & Desai, 2011; Dove, 2009, 2014; Shallice & Cooper, 2013). It might be difficult to disentangle these representational questions on the basis of electrophysiological data alone. Also, as mentioned above, our study was not designed to rule out the involvement of sensory processes during comprehension—and previous fMRI studies clearly show such sensory activations.

Finally, our study is important with regard to the literature of embodiment and language comprehension. In this literature, the influence of a linguistic stimulus on subsequent sensory processes is typically interpreted as core evidence in favor of the involvement of sensory processes during language comprehension (e.g., Brunyé et al., 2010; Zwaan et al., 2002; Wolter, Dudschig, de la Vega, & Kaup, 2015). Our study suggests that these associations between language and sensory processes might also be—at least partly—triggered by higher-level conceptual semantic networks beyond pure sensory processing networks. Thus, future studies are needed to get a clearer indication of which processes are underlying these influences of language on subsequent sensorimotor processes, instead of following the mere assumptions that such compatibility effects directly provide evidence for an embodied model of language comprehension.

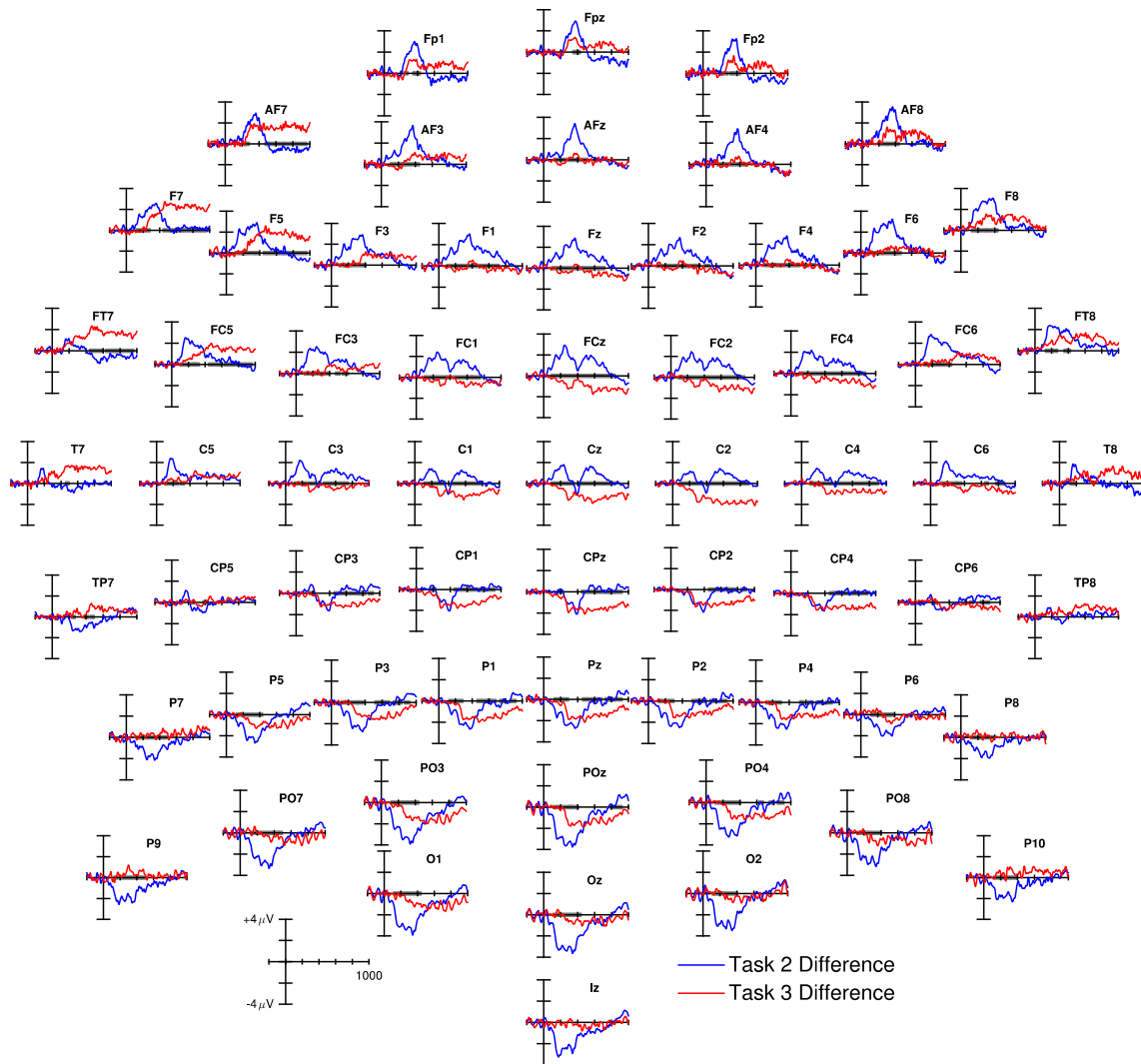
## Appendix 1

### Permutation: Task 1 versus Task 2



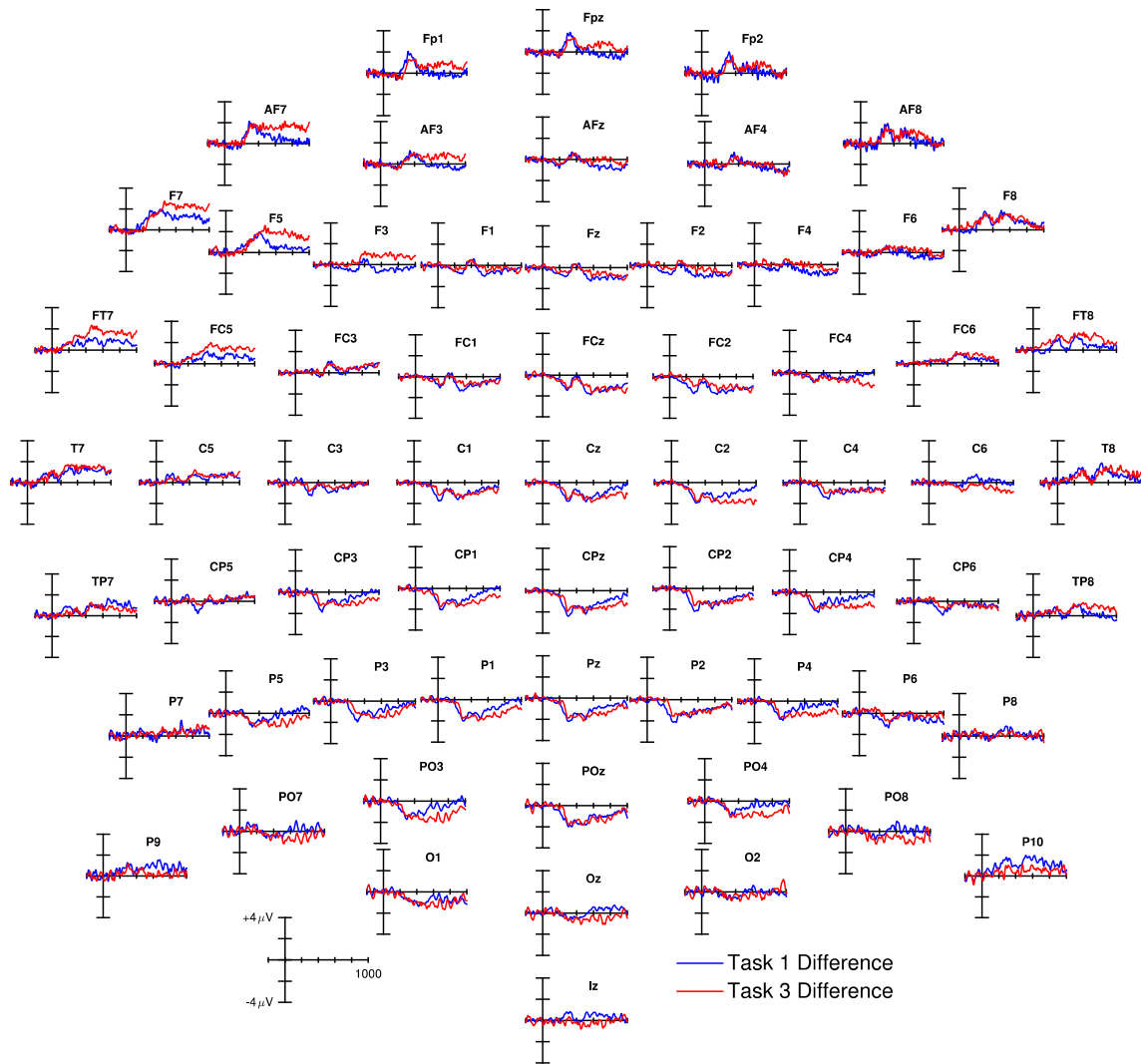
**Fig. 6**  $T1_{\text{linguistic}}$  and  $T2_{\text{physical}}$  difference waveforms. The gray bar indicates the time intervals where  $T1_{\text{linguistic}}$  and  $T2_{\text{physical}}$  significantly differ according to permutation tests (see Maris & Oostenveld, 2007)

## Permutation: Task 2 versus Task 3



**Fig. 7**  $T2_{\text{physical}}$  and  $T3_{\text{imagery}}$  difference waveforms. The gray bar indicates the time intervals where  $T2_{\text{physical}}$  and  $T3_{\text{imagery}}$  significantly differ according to permutation tests (see Maris & Oostenveld, 2007)

Permutation: Task 1 versus Task 3

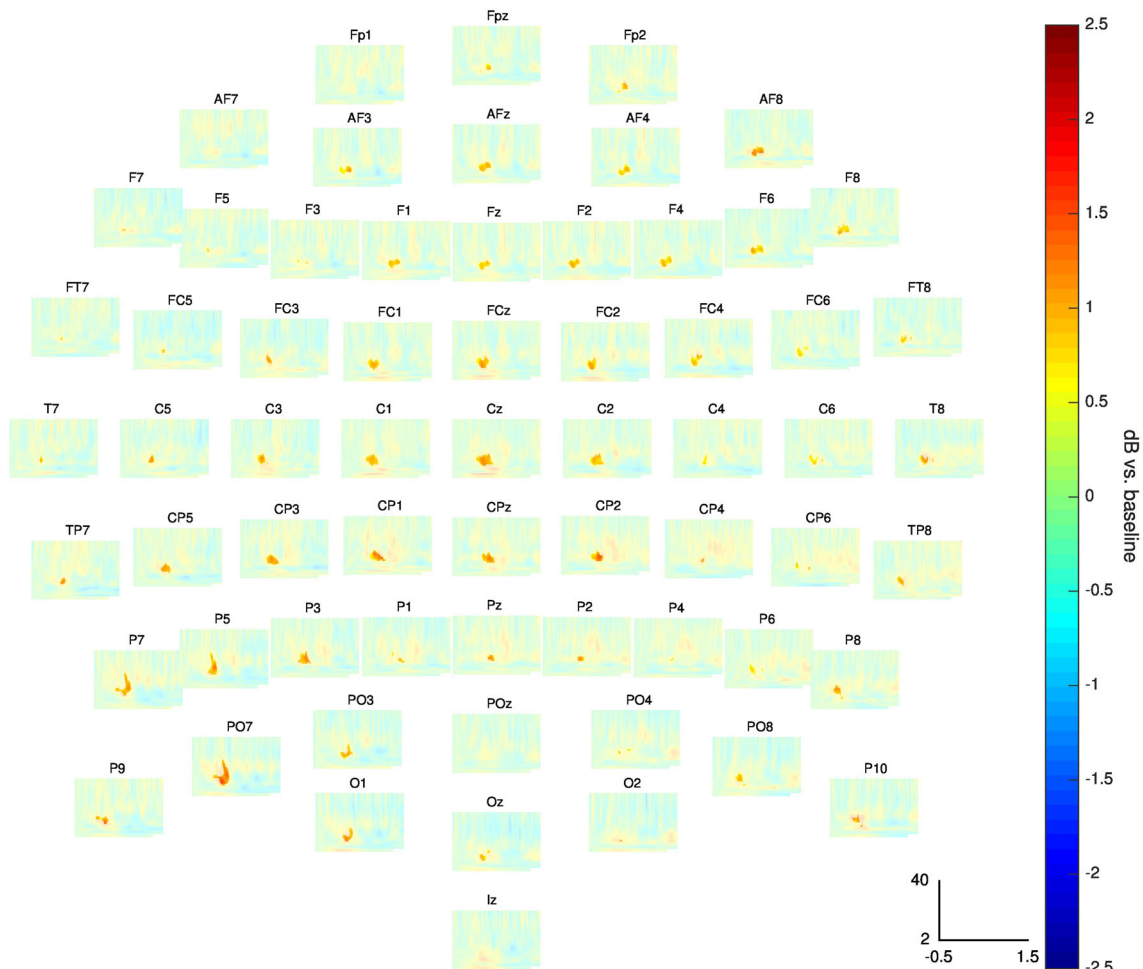


**Fig. 8** T1<sub>linguistic</sub> and T3<sub>imagery</sub> difference waveforms. The gray bar indicates the time intervals where T1<sub>linguistic</sub> and T3<sub>imagery</sub> significantly differ according to permutation tests (see Maris & Oostenveld, 2007)



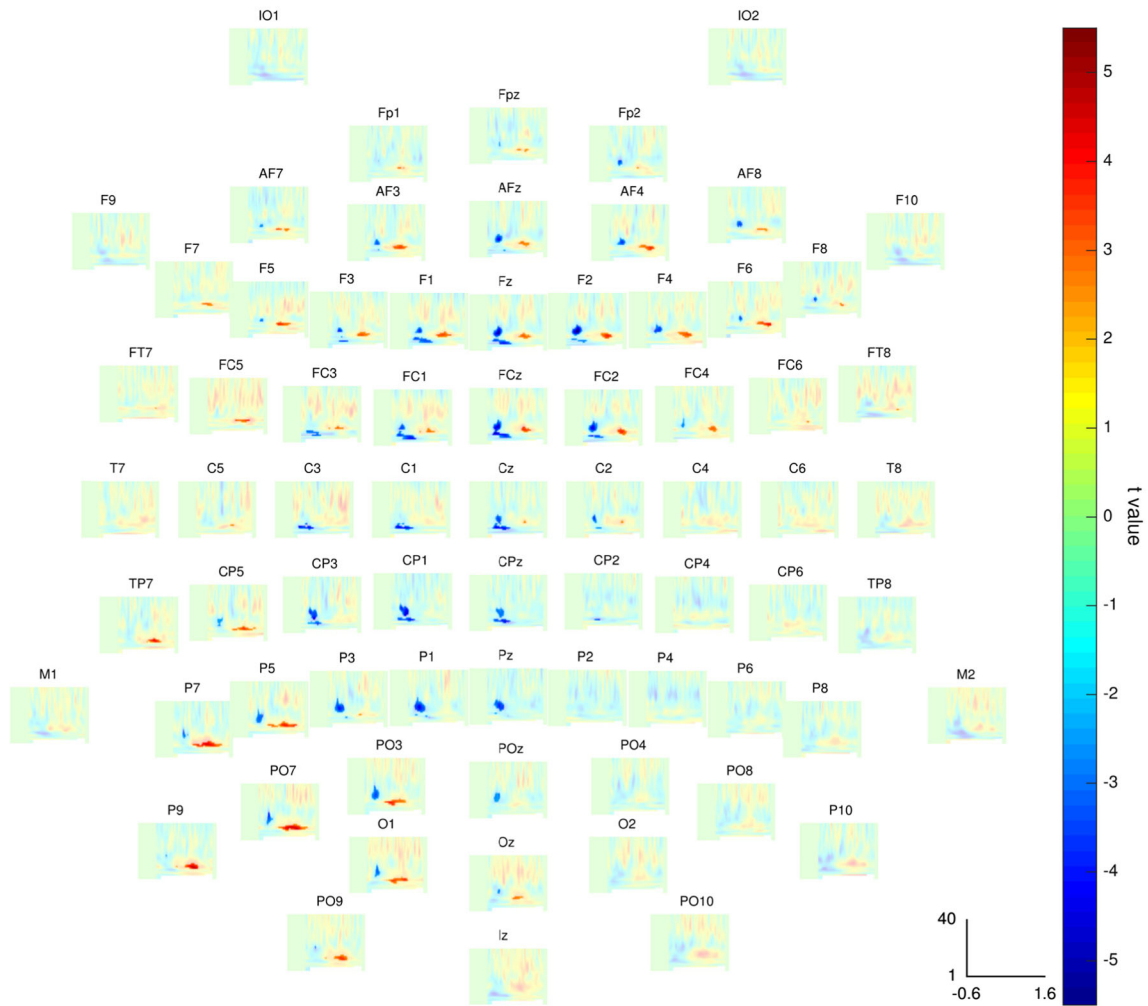
## Appendix 2

## Time-frequency: Task 1 versus Task 2



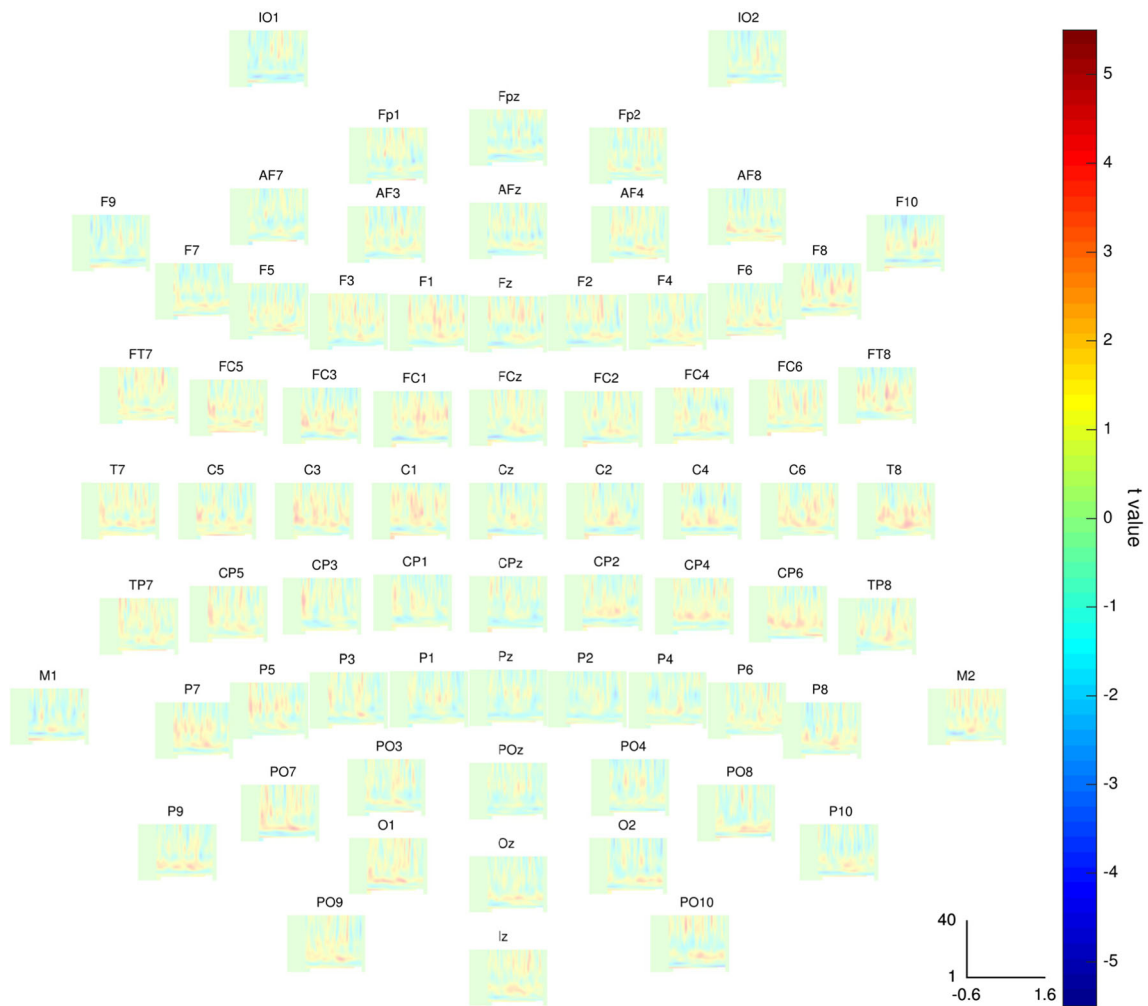
**Fig. 9** Permutation tests across frequency and time comparing the differences in  $T1_{\text{linguistic}}$  (mismatch minus match) with the differences  $T2_{\text{physical}}$  (mismatch minus match). Nonsignificant time points and frequencies are alpha masked (see Maris & Oostenveld, 2007)

**Time-frequency: Task 2 versus Task 3**



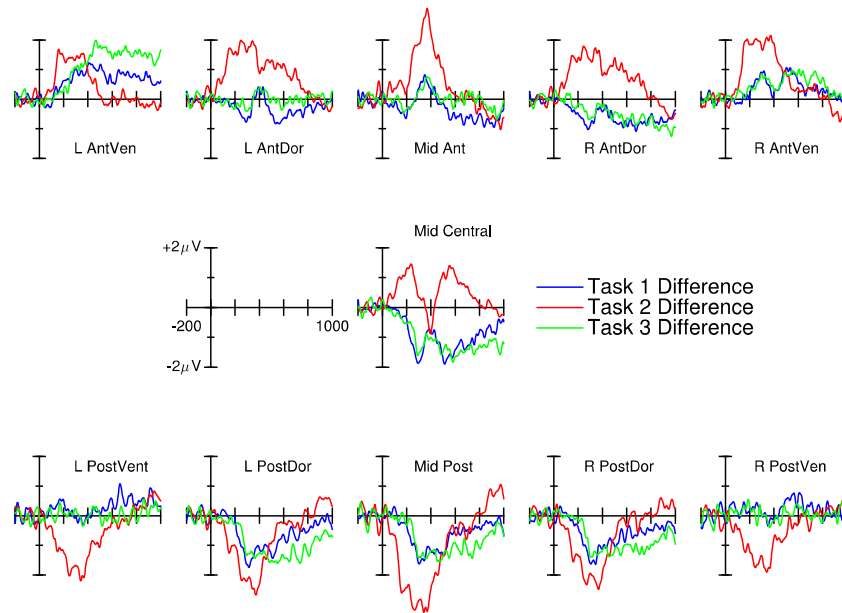
**Fig. 10** Permutation tests across frequency and time comparing the differences in  $T2_{\text{physical}}$  (mismatch minus match) with the differences  $T3_{\text{imagery}}$  (mismatch minus match). Nonsignificant time points and frequencies are alpha masked (see Maris & Oostenveld, 2007)

### Time-frequency: Task 1 versus Task 3



**Fig. 11** Permutation tests across frequency and time comparing the differences in  $T1_{\text{linguistic}}$  (mismatch minus match) with the differences  $T3_{\text{imagery}}$  (mismatch minus match). Nonsignificant time points and frequencies are alpha masked (see Maris & Oostenveld, 2007)

### Appendix 3



**Fig. 12** Difference waveforms for the three tasks T1<sub>linguistic</sub>, T2<sub>physical</sub>, and T3<sub>imagery</sub>

### Appendix 4

Material: Die Waschmaschine schleudert [The washing machine spins]. Das Schwein grunzt [The pig grunts]. Der Anrufbeantworter piept [The answering machine beeps]. Die Henne gackert [The hen clucks]. Die Kuh muht [The cow moos]. Der Motor brummt [The motor hums]. Das Wasser blubbert [The water is bubbling]. Das Geschirr klappert [The crockery rattles]. Die Nähmaschine ratter [The sewing machine chatters]. Der Wolf heult [The wolf howls]. Der Regen prasselt [The rain crepitates]. Der Alarmton dröhnt [The alarm is blaring]. Der Wecker tickt [The alarm clock is ticking]. Das Publikum applaudiert [The audience applauds]. Die Tür knallt [The door slams]. Der Korke knallt [The cork pops]. Die Limonadenflasche zischt [The bottle of lemonade fizzes]. Die Tastatur klappert [The keyboard clacks]. Der Schuss kracht [The shot crashes]. Das Motorrad dröhnt [The motorcycle roars]. Der Hahn kräht [The rooster crows]. Der Presslufthammer wummert [The jackhammer booms]. Der Wind heult [The wind howls]. Die Münzen klippeln [The coins are jingling]. Die Kanone donnert [The cannon thunders]. Die Vögel zwitschern [The birds are chirping]. Der Schmied hämmert [The blacksmith is hammering]. Das Mädchen kichert [The girl giggles]. Der Bach plätschert [The stream ripples]. Der Wasserhahn tropft [The faucet leaks]. Der Föhn blast [The hair dryer blows]. Die Möwe kreischt [The seagull screeches]. Der Hund knurrt [The dog growls]. Der Hund bellt [The dog barks]. Die Fahrradbremsen quietschen [The bicycle's brakes squeal]. Der Schiedsrichter pfeift [The referee whistles]. Die Mücke sirt [The mosquito

buzzes]. Der Kugelschreiber klickt [The pen clicks]. Der Frosch quakt [The frog croaks]. Der Specht hämmert [The woodpecker is hammering]. Das Mädchen schlürft [The girl slurps]. Der Schlüsselbund klimpert [The keychain strums]. Das Telefon klingelt [The phone rings]. Der Ballon platzt [The balloon bursts]. Der Hund winselt [The dog whines]. Die Taube gurr [The dove coos]. Der Holzboden knarrt [The wooden floor creaks]. Das Handy vibriert [The mobile phone vibrates]. Die Ziege meckert [The goat bleats]. Der Löwe brüllt [The lion roars]. Das Radio rauscht [The radio swooshes]. Die Katze miaut [The cat meows]. Das Baby schreit [The baby cries]. Das Schaf blökt [The sheep bleats]. Der Zug ratter [The train rattles]. Die Reifen quietschen [The tires screech]. Der Junge niest [The boy sneezes]. Die Grillen zirpen [The crickets chirp]. Die Geige quietscht [The violin squeaks]. Der Wecker klingelt [The alarm clock is ringing]. Der Kessel pfeift [The kettle whistles]. Die Blätter rascheln [The leaves rustle]. Die Sirene heult [The sirens are howling]. Die Stöckelschuhe klackern [The stiletto heels are clacking]. Das Papier raschelt [The paper is rustling]. Das Feuer knistert [The fire crackles]. Der Stoff reißt [The fabric tears]. Die Tür knarrt [The door creaks]. Die Säge kreischt [The saw screeches]. Die Kirchenglocke läutet [The church bells ring]. Die Katze faucht [The cat hisses]. Der Elefant trompetet [The elephant trumpets]. Das Auto hupt [The car honks]. Das Herz schlägt [The heart beats]. Die Peitsche knallt [The whip cracks]. Die Biene summt [The bee buzzes]. Die Frau hustet [The woman coughs]. Der Patient gurgelt [The patient gargles]. Der Mann trommelt [The man drums]. Das Pferd wiehert [The horse neighs].



## References

- Aramaki, M., Marie, C., Kronland-Martinet, R., Ystad, S., & Besson, M. (2010). Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. *Journal of Cognitive Neuroscience*, *22*(11), 2555–2569.
- Areshenkoff, C. N., Bub, D. N., & Masson, M. E. (2017). Task-dependent motor representations evoked by spatial words: Implications for embodied accounts of word meaning. *Journal of Memory and Language*, *92*, 158–169.
- Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(2), 250–267.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.
- Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.
- Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*(3), 229–241.
- Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536.
- Brunyé, T. T., Ditman, T., Mahoney, C. R., Walters, E. K., & Taylor, H. A. (2010). You heard it here first: Readers mentally simulate described sounds. *Acta Psychologica*, *135*(2), 209–215.
- Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H. J., & Jancke, L. (2005). Scanning silence: Mental imagery of complex sounds. *NeuroImage*, *26*(4), 1119–1127.
- Cummings, A., Čeponienė, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, *1115*(1), 92–107.
- Daltrozzo, J., & Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience*, *21*(10), 1882–1892.
- De Moortel, I., Munday, S. A., & Hood, A. W. (2004). Wavelet analysis: The effect of varying basic wavelet parameters. *Solar Physics*, *222*(2), 203–228.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.
- Dick, F., Bussiere, J., & Saygin, A. P. (2002). The effects of linguistic mediation on the identification of environmental sounds. *Center for Research in Language Newsletter*, *14*(3), 3–9.
- Dien, J., & Santuzzi, A. M. (2005). Application of repeated measures ANOVA to high-density ERP datasets: A review and tutorial. In T. C. Handy (Ed.), *Event-related potentials: A method handbook* (pp. 57–82). Cambridge, MA: MIT Press.
- Djordjevic, J., Zatorre, R. J., Petrides, M., Boyle, J. A., & Jones-Gotman, M. (2005). Functional neuroimaging of odor imagery. *NeuroImage*, *24*(3), 791–801.
- Djordjevic, J., Zatorre, R. J., Petrides, M., & Jones-Gotman, M. (2004). The mind's nose effects of odor and visual imagery on odor detection. *Psychological Science*, *15*(3), 143–148.
- Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, *110*, 412–431.
- Dove, G. (2014). Thinking in words: Language as an embodied medium of thought. *Topics in Cognitive Science*, *6*(3), 371–389.
- Dudschig, C., de la Vega, I., & Kaup, B. (2015). What's up? Emotion-specific activation of vertical space during language processing. *Acta Psychologica*, *156*, 143–155.
- Dudschig, C., & Kaup, B. (in press). Is it all task-specific? The role of binary responses, verbal mediation and saliency for eliciting language-space associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., . . . Seidenberg, M. S. (2016). Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cerebral Cortex*, *26*(5), 2018–2034.
- Filik, R., Leuthold, H., Wallington, K., & Page, J. (2014). Testing theories of irony processing using eye-tracking and ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 811–828.
- Fuentemilla, L., Marco-Pallarés, J., & Grau, C. (2006). Modulation of spectral power and of phase resetting of EEG contributes differentially to the generation of auditory event-related potentials. *NeuroImage*, *30*(3), 909–916.
- Giard, M. H., Perrin, F., Echallier, J. F., Thevenet, M., Froment, J. C., & Pernier, J. (1994). Dissociation of temporal and frontal components in the human auditory N1 wave: A scalp current density and dipole model analysis. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *92*(3), 238–252.
- Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905–922.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.
- González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading 'cinnamon' activates olfactory brain regions. *NeuroImage*, *32*(2), 906–912.
- Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, *9*(7), 697–704.
- Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307.
- Hubbard, T. L. (2010). Auditory imagery: Empirical findings. *Psychological Bulletin*, *136*(2), 302–329.
- Intons-Peterson, M. J. (1980). The role of loudness in auditory imagery. *Memory & Cognition*, *8*, 385–393.
- Intons-Peterson, M. J. (1992). Components of auditory imagery. In D. Reisberg (Ed.), *Auditory imagery* (pp. 45–72). New York, NY: Psychology Press.
- Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron*, *28*(3), 979–990.
- Kemmerer, D. (2015). Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychonomic Bulletin & Review*, *22*(4), 1068–1075.
- Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience*, *28*(47), 12224–12230.
- Kirmse, U., Jacobsen, T., & Schröger, E. (2009). Familiarity affects environmental sound processing outside the focus of attention: An event-related potential study. *Clinical Neurophysiology*, *120*, 887–896.
- Kirmse, U., Schröger, E., & Jacobsen, T. (2012). Familiarity of environmental sounds is used to establish auditory rules. *Neuroreport*, *23*, 320–324.
- Klatzky, R. L., Lederman, S. J., & Matula, D. E. (1991). Imagined haptic exploration in judgments of object properties. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 314.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1.

- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, 7(3), 302–307.
- Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, 2(9), 635–642.
- Kosslyn, S. M., & Pomerantz, J. R. (1977). Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, 9(1), 52–76.
- Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, 378(6556), 496–498.
- Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York: Oxford University Press.
- Kraemer, D. J., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: Sound of silence activates auditory cortex. *Nature*, 434(7030), 158–158.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920–933.
- Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *The Journal of Neuroscience*, 30(22), 7604–7612.
- Lebois, L. A., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2014). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, 39, 1–38.
- Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex*, 14(9), 1008–1021.
- Linke, A. C., & Cusack, R. (2015). Flexible information coding in human auditory cortex during perception, imagery, and STM of complex sounds. *Journal of Cognitive Neuroscience*, 27, 1322–1333.
- Machery, E. (2016). The amodal brain and the offloading hypothesis. *Psychonomic Bulletin & Review*. Advance online publication. doi:10.3758/s13423-015-0878-4.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190.
- Meteyard, L., & Vigliocco, G. (2008). The role of sensory and motor information in semantic representation: A review. In P. Calvo & A. Gomila (Eds.), *Handbook of cognitive science: An embodied approach* (pp. 293–312). London: Academic Press.
- Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.
- 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.
- Nobre, A. C., Allison, T., & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature*, 372(6503), 260–263.
- Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods*, 192(1), 152–162.
- O’Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, 12(6), 1013–1023.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 156869. doi:10.1155/2011/156869
- Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2006). Conceptual priming for environmental sounds and words: An ERP study. *Brain and Cognition*, 62(3), 267–272.
- Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2007). Is conceptual priming for environmental sounds obligatory? *International Journal of Psychophysiology*, 65(2), 162–166.
- Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2008). N400-effects to task-irrelevant environmental sounds: Further evidence for obligatory conceptual processing. *Neuroscience Letters*, 436(2), 133–137.
- Penfield, W., & Perot, P. (1963). The brain’s record of auditory and visual experience. *Brain*, 86(4), 595–696.
- Pylyshyn, Z. W. (1973). What the mind’s eye tells the mind’s brain: A critique of mental imagery. *Psychological Bulletin*, 80(1), 1–24.
- Scherg, M., & Berg, P. (1991). Use of prior knowledge in brain electromagnetic source analysis. *Brain Topography*, 4(2), 143–150.
- Schirmer, A., Soh, Y. H., Penney, T. B., & Wyse, L. (2011). Perceptual and conceptual priming of environmental sounds. *Journal of Cognitive Neuroscience*, 23(11), 3241–3253.
- Schön, D., Ystad, S., Kronland-Martinet, R., & Besson, M. (2010). The evocative power of sounds: Conceptual priming between words and nonverbal sounds. *Journal of Cognitive Neuroscience*, 22(5), 1026–1035.
- Segal, S. J., & Fusella, V. (1971). Effect of images in six sense modalities on detection of visual signal from noise. *Psychonomic Science*, 24(2), 55–56.
- Shallice, T., & Cooper, R. P. (2013). Is there a semantic system for abstract words? *Frontiers in Human Neuroscience*, 7(175), 1–10.
- Stevenson, R. J., & Case, T. I. (2005). Olfactory imagery: A review. *Psychonomic Bulletin & Review*, 12(2), 244–264.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1997). Oscillatory  $\gamma$ -band (30–70 Hz) activity induced by a visual search task in humans. *The Journal of Neuroscience*, 17(2), 722–734.
- Thomas, N. J. (2016). Mental imagery. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/sum2016/entries/mental-imagery/>.
- Van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, 33(4), 485–508.
- Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory’s echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, 97(20), 11125–11129.
- Wolter, S., Dudschig, C., de la Vega, I., & Kaup, B. (2015). Musical metaphors: Evidence for a spatial grounding of non-literal sentences describing auditory events. *Acta Psychologica*, 156, 126–135.
- Yoo, S. S., Lee, C. U., & Choi, B. G. (2001). Human brain mapping of auditory imagery: Event-related functional MRI study. *Neuroreport*, 12(14), 3045–3049.
- Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. (1996). Hearing in the mind’s ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience*, 8(1), 29–46.
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13(2), 168–171.