

## Binding Symbols and Sounds: Evidence from Event-Related Oscillatory Gamma-Band Activity

Andreas Widmann<sup>1</sup>, Thomas Gruber<sup>1</sup>, Teija Kujala<sup>2,3,4</sup>, Mari Tervaniemi<sup>3,4</sup> and Erich Schröger<sup>1</sup>

<sup>1</sup>*Institute of Psychology I, University of Leipzig, Germany*

<sup>2</sup>*Helsinki Collegium for Advanced Studies, University of Helsinki, Finland*

<sup>3</sup>*Cognitive Brain Research Unit, Department of Psychology, University of Helsinki, Finland*

<sup>4</sup>*Helsinki Brain Research Center, Helsinki, Finland*

### Abstract

The present study intended to examine the neural basis of audiovisual integration, hypothetically achieved by synchronized gamma-band oscillations (30–80 Hz) that have been suggested to integrate stimulus features and top-down information. To that end, we studied the impact of visual symbolic information on early auditory sensory processing of upcoming sounds. In particular, we used a symbol-to-sound-matching paradigm in which simple score-like patterns predict corresponding sound patterns. Occasionally, a single sound is incongruent with the corresponding element of the visual pattern. In response to expected sounds congruent with the corresponding visual symbol, a power increase of phase-locked (evoked) activity in the 40-Hz band was observed peaking 42-ms poststimulus onset. Thus, for the first time, we demonstrated that the comparison process between a neural model, the expectation, and the current sensory input is implemented at very early levels of auditory processing. Subsequently, expected congruent sounds elicited a broadband power increase of non-phase-locked (induced) activity peaking 152-ms poststimulus onset, which might reflect the formation of a unitary event representation including both visual and auditory aspects of the stimulation. Gamma-band responses were not present for unexpected incongruent sounds. A model explaining the anticipatory activation of cortical auditory representations and the match of experience against expectation is presented.

### Keywords

audiovisual integration, audition, event-related brain potentials, gamma-band oscillatory activity, neural model, vision

### Introduction

The detection of a wrong note in a musical piece requires some prior knowledge, which might be derived from the preceding parts of the music or from memory representations of the piece. In musicians, they can also be formed by comparing a visually presented score with the auditory input. The brain processes associated with confirming, disconfirming, and integrating expectations of forthcoming sounds with the current sounds have been widely studied within the auditory modality by means of electroencephalogram (EEG) and event-related potentials (ERPs; see e.g., [Näätänen 1992](#); [Winkler et al. 1996](#); [Schröger 1998](#)). In a recent study ([Widmann et al. 2004](#)), we demonstrated that expected sounds are processed differently as compared with unexpected sounds in a symbol-to-sound-matching paradigm. Simple “score-like” visual patterns (cf., [Fig. 1](#)) predicted subsequently presented “melody-like” auditory patterns. Tones incongruent with the visually formed expectations elicited an ERP complex differing from that elicited by congruent tones, consisting of the so-called incongruency response (IR), followed by N2b-, P3a-, and P3b-like components.

Recently, it was suggested that oscillatory brain activity in the gamma range (> 30 Hz) might provide additional insights into brain processes complementary to ERPs ([Gruber et al. 2004](#); [Gruber and Müller 2005](#)). The early auditory-evoked gamma-band response (GBR) has an onset of approximately 10–15 ms after stimulus onset, lasts for about 100 ms, and reveals a peak amplitude of about 50–60 ms poststimulus

([Basar et al. 1987](#); [Pantev et al. 1991](#)). It has been shown to be modulated by attention ([Yordanova et al. 1997](#)), sensory gating ([Müller et al. 2001](#)), and stimulation rates ([Makeig 1990](#); [Pantev et al. 1993](#)). [Debener and colleagues \(2003\)](#) demonstrated top-down influences on the auditory-evoked GBR at around 60 ms using a novelty-oddball design. Enhanced GBRs were found in response to target but not to novel stimuli. In contrast to the early evoked GBR, the later induced GBR is supposed to reflect the activation of “cortical object representations,” which are composed of many different features and include visual ([Tallon-Baudry and Bertrand 1999](#)), auditory ([Kaiser and Lutzenberger 2005](#)), and semantic information ([Gruber and Müller 2005](#)).

Most of the researches on cortical object representations are confined to a single modality, and so far little is known on the neural basis of processes matching information across modalities. However, our perception is largely based on the integrated information from different modalities: for instance, speech perception is based on heard sounds and seen lip movements ([Kaiser et al. 2005](#)). Here, we report event-related gamma-band oscillations elicited in the symbol-to-sound-matching task developed by [Widmann et al. \(2004\)](#). In a reanalysis of EEG data from which ERPs were previously reported by [Widmann et al. \(2004\)](#), we tested whether early auditory-evoked GBRs that were reported for auditory target detection ([Debener et al. 2003](#)) do also occur when a sound has to be evaluated for congruency with a visual symbol. Moreover, we determined whether later induced GBR effects usually associated with binding and the establishment of a unitary object representation within a modality ([Keil et al. 2001](#)) are also elicited when features have to be compared across modalities. Such an effect has recently been shown for highly trained audiovisual speech stimuli ([Kaiser et al. 2005](#)). However, for arbitrarily defined relations between vision and audition, induced GBRs have not yet been reported.

## Materials and Methods

### Participants

Twenty-four right-handed subjects (5 males, mean age 24 years, range 19–35 years) participated in the experiment. All of them reported normal or corrected-to-normal vision and normal hearing. None of the participants had a history of a neurological disease or injury. Subjects were paid for their participation in the experiment and gave their informed consent after the details of the procedure had been explained to them.

### Stimuli and Apparatus

Participants were comfortably seated in a dimly lit, sound-attenuated, and electrically shielded chamber. They held a response pad with buttons under their left index, right index, and right middle fingers. An LCD computer screen was placed about 130 cm in front of the subjects' eyes so that the stimuli appeared slightly below the horizontal line of sight. Visual stimuli, presented simultaneously on a black background, consisted of 4–6 light gray rectangles aligned above and below the horizontal midline of the screen. The rectangles subtended a visual angle of  $0.32^\circ \times 0.16^\circ$ . The empty spaces between the rectangles were of the same width as the rectangles. The upper corners of the rectangles were placed  $0.36^\circ$  above or  $0.04^\circ$  below the horizontal midline of the screen (see [Fig. 1](#) for a sample screenshot). Sounds of the auditory stimuli were triangle waves (containing only odd harmonics with an amplitude ratio proportional to the inverse square of the harmonic number) with a frequency of 352 (F<sup>4</sup>) or 422.4 Hz (G<sup>#4</sup>) and a duration of 300 ms including a 5-ms rise and a 5-ms fall time (Hann window). Sounds were presented via headphones (Sennheiser HD 25) at an intensity of 65 dB SPL. Each trial started with the visual presentation of a pattern of 4–6 rectangles on the screen. Visual stimuli stayed on the screen until both behavioral responses (see below) were given by the subject. The sound presentation started 1000 ms after the onset of the visual display. The number of auditory stimuli matched the number of rectangles. The tones were presented with an interstimulus interval of 300 ms. Congruent stimuli were defined as either high tones matching rectangles above the midline or low tones matching rectangles below the horizontal midline at the corresponding serial positions in the visual and auditory patterns. Incongruent stimuli were defined as either high tones matching rectangles below the midline or low tones matching rectangles above the midline at the corresponding serial positions in the visual and auditory patterns. The intertrial interval was 1200 ms.

### Design and Procedure

The study was preceded by a training session on a separate day consisting of 4 blocks of 53 trials each.

The experiment consisted of 12 blocks of 53 trials, each including a visual and an auditory pattern (note that a “trial” is defined as a complete sequence of 4–6 sounds related to one visual pattern consisting of 4–6 elements and not as a single sound), with a short break after the sixth block. The visual pattern in each trial was selected pseudorandomly so that the same visual pattern was never presented twice in consecutive trials. In 50% of the trials, the visual and the auditory patterns were incongruent in one pseudorandomly selected element. In the visual and the auditory patterns, at least one element had to be distinct from the other elements. That is, no visual pattern had rectangles only above or only below the horizontal midline, and none of the auditory patterns had only high or only low sounds. Altogether, 84 four-element, 180 five-element, and 372 six-element patterns (comprising 3468 auditory stimuli in 636 trials) were presented during the experiment. In 318 trials, one sound was incongruent with the corresponding visual element. Participants were instructed to press a button with their left index finger as rapidly as possible after the onset of the last tone of the auditory pattern. Subsequently, participants had to indicate as accurately as possible whether the visual and auditory patterns were congruent or not by pressing a button with their right index or middle finger, respectively. Participants received visual feedback if their first response was carried out before or more than 600 ms after the onset of the last tone by presenting an error message for 2400 ms at the center of the display. In this case, no second response was required. The first response was required from the participants in order to get them follow the complete auditory pattern in a similar manner as when reading text, even when an incongruent sound was presented. For the second response, there was no time limit. The subsequent trial started after both responses were given.

## Data Recording

The EEG was recorded with Ag–AgCl electrodes from the 19 standard positions of the extended 10–20-system and from the left and right mastoids (M1, M2). All electrodes were referenced to the tip of the nose. The vertical electrooculogram (EOG) was recorded between supra- and infraorbitally placed electrodes and the horizontal EOG between the outer canthi of the 2 eyes. Impedances of eye and mastoidal electrodes were kept below 10 k $\Omega$ , and all other electrode impedances were kept below 5 k $\Omega$ . EEG and EOG were filtered online with a bandpass of 0.1–100 Hz and sampled with a digitization rate of 500 Hz (Synamps EEG amplifier, NeuroScan, TX). Reaction times were recorded for each trial.

## Data Analysis

Raw data processing apart from the wavelet transform was performed with the EEGLAB toolbox by [DeLorme and Makeig \(2004\)](#). EEG and EOG were divided into epochs of 2048 ms time locked to the onset of each auditory stimulus including an 800-ms prestimulus baseline. Due to contamination with muscular artifacts in some of the subjects at outermost electrode sites, analysis was restricted to electrodes sites F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4. Epochs with signal changes exceeding 150  $\mu$ V in a time window from –200 to 500 ms relative to stimulus onset at any of the remaining electrodes were discarded from the analysis. The first stimulus of each auditory pattern was excluded from the analysis because it required absolute judgment of pitch. The last stimulus of each auditory pattern was excluded from the analysis because it was contaminated by motor-related potentials. Furthermore, trials with an incorrect behavioral response were discarded from the EEG analysis. For each remaining incongruent stimulus, a congruent sibling stimulus from the nearest preceding or following trial with congruent visual and auditory patterns was selected. In particular, the congruent sibling stimulus was selected in such a way that it matched the incongruent stimulus in pitch, pattern length, and serial position within the pattern. This was done in order to guarantee an equal number of epochs—and, thus, an equal signal-to-noise ratio—for congruent and incongruent sounds. The congruent and the incongruent stimuli were averaged separately for each subject. On average, 211 stimuli (i.e., 66.4%, standard deviation [SD] = 26 stimuli) per condition per subject remained in the sample.

To compute the time-frequency representation of oscillatory gamma-band activity phase locked to stimulus onset (termed as evoked activity), the individual averages for congruent and incongruent stimuli were convolved with complex Morlet wavelets  $w(t, f_0) = A \exp(2\pi i t f_0) \exp(-t^2/2\sigma_t^2)$  normalized to unit energy  $A = (\sigma_t \sqrt{\pi})^{-1/2}$  with  $\sigma_t = 1/(2\pi \sigma_f)$  ([Tallon-Baudry et al. 1996](#)). The wavelet family used was defined by  $f_0/\sigma_f = 7$  ranging from 30 Hz (duration  $[2\sigma_t]$  74.2 ms, spectral bandwidth  $[2\sigma_f]$  8.6 Hz) to 80 Hz (duration 27.9 ms, bandwidth 22.9 Hz) in 1-Hz steps. The amplitude of the evoked activity was calculated by taking the moduli of the time-frequency representations. The time-frequency representation of the total oscillatory gamma-band activity (phase locked and non-phase locked) was calculated by averaging the moduli of the convolution of each epoch with the same family of complex Morlet wavelets separately for congruent and incongruent stimuli and each subject, respectively. The phase-locking factor was calculated

by the modulus of the average of the normalized complex time-varying energy of each epoch separately for congruent and incongruent stimuli and each subject (Jervis et al. 1983; Tallon-Baudry et al. 1996). For each frequency band, the mean of a –200- to –50-ms prestimulus baseline was subtracted from the time-frequency representation in order to eliminate uncorrelated noise and effects not related to sound onset. It should be noted that total activity is sometimes in the literature referred to as “induced” activity (Tallon-Baudry and Bertrand 1999; Herrmann and Mecklinger 2000). In the present study, we use the term induced activity for non-phase-locked activity calculated from the difference of total activity and evoked activity.

Based on the grand average time-frequency representation of congruent stimuli averaged across all electrodes, we defined time-frequency windows, which were used for subsequent repeated-measures analysis of variance (ANOVA) with the factors congruency (congruent vs. incongruent), anterior–posterior (frontal vs. central vs. parietal electrodes), and laterality (left hemisphere vs. midline vs. right hemisphere electrodes). For evoked activity, 2 time-frequency windows from 14 to 70 ms, 36 to 42 Hz and from 164 to 220 ms, 41 to 47 Hz were determined. For total activity, a time-frequency window from 116 to 188 ms, 30 to 80 Hz was determined. Window size was oriented on the duration and bandwidth of the respective wavelets, except for induced activity which was considered as broadband activity. To exclude baseline differences between conditions, the baseline uncorrected evoked and the induced frequency ranges were tested with the ANOVA model described above, in a time window from –200 to –50 ms relative to stimulus onset. All t-tests reported are two-sided. Greenhouse–Geisser adjustments of degrees of freedom were performed where applicable.

## Results

Figure 2 shows grand average time-frequency and phase-locking factor plots of evoked (*A, B*) and total activity (*D, E*) for congruent and incongruent sounds averaged across all the 9 channels included in the analysis. Furthermore, Figure 2 shows the grand average scalp topographies of the incongruent minus congruent difference amplitude for evoked (*C*; 14–70 ms, 36–42 Hz) and total activity (*F*; 116–188 ms, 30–80 Hz). Figure 3 shows time-amplitude plots for evoked gamma-band activity (36–42 Hz in *A* and 41–47 Hz in *B*) and total gamma-band activity (30–80 Hz in *C*). Furthermore, Figure 3 shows grand average ERPs for congruent and incongruent stimuli (*D*).

### Evoked Activity

Results showed increased phase-synchronized activity relative to baseline for congruent sounds in 2 time-frequency regions. First, there was an increase in the time and frequency range of the auditory middle latency response (MLR) peaking at 42 ms after sound onset at 39 Hz. The phase-locking factor in this time-frequency window was significantly different from baseline level ( $F_{1,23} = 6.72, P < 0.016$ ). The ANOVA in this time-frequency window (14–70 ms, 36–42 Hz) showed an interaction of the factors congruency and laterality ( $F_{2,46} = 4.95, P < 0.023, \epsilon = 0.699$ ), reflecting an amplitude difference between the responses elicited by congruent and incongruent sounds. It was pronounced at left hemisphere electrode locations ( $0.27 \mu\text{V}, t_{23} = 3.65, P < 0.001$ ) decreasing at midline ( $0.20 \mu\text{V}, t_{23} = 2.56, P < 0.018$ ) and even more at right hemisphere electrode locations ( $0.12 \mu\text{V}$ , not significant). Second, there was a later increase peaking at 192 ms after sound onset at 44 Hz. The phase-locking factor in this time-frequency window was not significantly different from baseline level. The ANOVA in this time-frequency window (164–220 ms, 39–49 Hz) showed a main effect of the factor congruency ( $F_{1,23} = 13.29, P < 0.001$ ).

### Total Activity

The results showed a broadband increase of total gamma-band activity relative to the baseline activity caused by the congruent sounds between about 100 and 200 ms after stimulus onset peaking at 152 ms and 61 Hz. This peak was not accompanied by phase-locked oscillations as derived from the frequency decomposition of the evoked response. The phase-locking factor in this time-frequency window was not significantly different from baseline level. Thus, this peak is referred to as induced gamma-band response. ANOVA in the 116- to 188-ms time window showed an interaction of the factors congruency and anterior–posterior electrode location ( $F_{2,46} = 8.77, P < 0.002, \epsilon = 0.783$ ). The amplitude difference between the responses to congruent and incongruent sounds increased from anterior ( $0.14 \mu\text{V}, t_{23} = 2.82, P < 0.01$ ) to central ( $0.23 \mu\text{V}, t_{23} = 3.71, P < 0.001$ ) to posterior electrodes ( $0.28 \mu\text{V}, t_{23} = 4.03, P < 0.001$ ). In addition, the ANOVA also revealed an interaction of the factors congruency and laterality ( $F_{2,46} = 3.8, P < 0.042, \epsilon = 0.769$ ). This interaction reflects the amplitude difference between the responses to congruent and incongruent sounds, which was maximal at the midline electrodes ( $0.24 \mu\text{V}, t_{23} = 4.13, P < 0.001$ ) being

slightly smaller at the left ( $0.23 \mu\text{V}$ ,  $t_{23} = 3.85$ ,  $P < 0.001$ ) and the right hemisphere electrodes ( $0.18 \mu\text{V}$ ,  $t_{23} = 3.01$ ,  $P < 0.006$ ).

Tests for baseline differences (from  $-200$  to  $-50$  ms prior to stimulus onset) between conditions revealed no significant effects for the evoked and the induced gamma-band response, respectively.

## Event-Related Potentials

ERP responses to incongruent sounds showed an increased negativity compared with congruent sounds in the N1 time range. This difference is referred to as IR. IR was followed by N2b- and P3a-like components. For a detailed description and statistical analysis of the ERP result, see [Widmann et al. \(2004\)](#).

## Discussion

Left frontally distributed early evoked and parietally distributed later induced GBRs were obtained when sounds were congruent with the corresponding visual symbol. We consider the early evoked (at 42 ms) and the later induced (at 152 ms) GBRs the most important findings and, thus, focus the discussion on these 2 peaks. Subsequently, we will relate results from the frequency domain to the ERP results reported by [Widmann et al. \(2004\)](#). Finally, we suggest a functional model on the mental processes during symbol-to-sound matching, accounting for both the time and the frequency domain.

Importantly, the measured early evoked GBR was not accompanied by an increase in total gamma power. Thus, it reflects phase synchronization relative to stimulus onset rather than additional cortical activation (cf., [Shah et al. 2004](#)). The early auditory-evoked GBR has been shown to be enhanced by the allocation of attention to the stimulation ([Tiitinen et al. 1993](#)) and sensory gating ([Müller et al. 2001](#)). Early auditory-evoked GBRs are not modulated by the probability of sounds in passive oddball situation ([Tiitinen et al. 1994](#)); however, when rare target sounds have to be attentively detected (active oddball situation), they elicit increased early evoked GBR compared with frequent nontarget sounds ([Debener et al. 2003](#)). We also found a modulation of oscillatory activity in the time and frequency range of the early evoked GBR for sounds being actively discriminated. An early evoked GBR was observed in response to congruent but not in response to incongruent sounds. In the present paradigm, visual symbols trigger expectations about forthcoming sounds that have to be compared with the current sensory input. The early auditory-evoked GBR seems to mirror this early matching mechanism. This mechanism might be controlled by top-down processes that activate the memory representation of the expected stimulus. These activated synaptic patterns might further enhance subthreshold oscillations in sensory cortical areas before stimulation (expressing the prediction about auditory input patterns; [Engel et al. 2001](#)). If bottom-up input and prediction match, this will result in “resonance phenomena” in the respective cortical network and, thus, in enhanced evoked gamma oscillations. This argument is in line with a match-and-utilization model proposed for the visual modality ([Herrmann et al. 2004](#)). These authors suggest that the visual early evoked GBR is a signature of a match of sensory information with stored representations in memory. It should be noted that transient early evoked gamma-band oscillations are supposed to contribute to the auditory MLR ([Basar et al. 1987](#); [Pantev et al. 1993](#); [Müller et al. 2001](#)) that originates in superior temporal areas and reflects the earliest activation of primary auditory cortex ([Ponton et al. 2002](#)).

The present finding poses an interesting question: Why are early evoked GBRs observed in response to rare target sounds in an oddball paradigm ([Debener et al. 2003](#)) but in response to frequent congruent sounds in the present paradigm? In oddball paradigms, targets are usually well defined by a constant feature and can be detected by matching each sound against a memory template of the target (as indicated by the increased early evoked GBRs in response to targets but not to novels). In the present paradigm, the rare incongruent sounds are not defined by a constant feature. Thus, subjects can solve the task either by matching each sound against the one not predicted by the visual pattern to detect the rare incongruent sounds or by matching each sound against the predicted one to detect all congruent sounds. The observed increased early evoked GBRs in response to congruent sounds suggest that the latter strategy was chosen by the participants. This was confirmed by the subjects’ verbal reports after the experiment (2 sample trials are provided at <http://www.uni-leipzig.de/~biocog/widmann/congruent.avi> and <http://www.uni-leipzig.de/~biocog/widmann/incongruent.avi>).

The increase of the early evoked GBR was found to be more pronounced over left hemispheric electrode sites, possibly indicating a stronger involvement of left hemisphere auditory areas. Recent imaging studies show that left hemispheric lateralization can also be observed during the processing of nonspeech sounds ([Devlin et al. 2003](#); [Yoo et al. 2005](#)). As the processing of temporal information is often associated with left hemispheric lateralization ([Tervaniemi and Hugdahl 2003](#); [Schönwiesner et al. 2005](#)) and “cor-

rect” sounds are defined in the frequency and the time domains, the left hemispheric preponderance of the early evoked GBR could indicate a relative dominance of temporal information (i.e., pitch changes) over spectral information (i.e., absolute pitch) in the present task.

The induced GBR subsequent to the early GBR may reflect some further processing when the sound was found to match the prediction. There are different attributions on the functional significance of later induced gamma oscillations mainly based on studies in the visual modality, which relate to memory and object formation. For example, [Tallon-Baudry and Bertrand \(1999\)](#); [Bertrand and Tallon-Baudry 2000](#)) suggest that they reflect mechanisms for the construction of object representations driven by sensory input or internal, top-down processes. In the match-and-utilization model ([Herrmann et al. 2004](#)), induced gamma is supposed to reflect processes of readout and utilization of earlier matches between incoming information and memory (see above). [Gruber et al. \(2004\)](#) associate them with memory encoding and retrieval. To our knowledge, only a limited number of EEG and magnetoencephalography studies report induced GBRs in the auditory modality with highly inconsistent results. Employing variations of the active and passive auditory oddball paradigms, the reported results range from decreased activity in response to targets ([Bertrand and Tallon-Baudry 2000](#)), no significant difference between targets and nontargets ([Debenner et al. 2003](#)), to increased activity in response to deviants (i.e., targets) ([Marshall et al. 1996](#); [Kaiser and Lutzenberger 2005](#)). Although [Bertrand and Tallon-Baudry \(2000\)](#) argue that memory representations of targets have to be kept active in standard trials (but not in target trials) in order to detect the targets, [Kaiser and Lutzenberger \(2005\)](#) relate induced gamma oscillations to auditory deviance processing. In the present study, induced GBR is only elicited when the visual and auditory information is congruent. It seems possible that these induced GBRs reflect the integration of the auditory expectation based on visual symbolic information and the auditory sensory information from the current stimulation into a unitary event representation underlying our perception of a correct sound. This integration can only succeed when visual and auditory information was found to match (early evoked GBR). This view is supported by recent results presented by [Ford et al. \(2005\)](#) showing that coherence between frontal and parietal areas and total gamma power increase when subjects are provided with veridical rather than distorted feedback of self-produced utterances. They interpret this augmentation as a reflection of “binding of expectation with experience.”

The results in the frequency domain reveal substantially new complementary insight about the mental processes involved in the symbol-to-sound-matching task as compared with the ERP results ([Widmann et al. 2004](#)). The IR as well as the N2b- and P3a-like ERP deflections were specific for the processing of incongruent stimuli. In contrast, evoked and induced GBRs are confined to congruent auditory stimuli. The IR in response to rare incongruent stimuli has been discussed ([Widmann et al. 2004](#)) as being functionally equivalent to the mismatch negativity (MMN) component elicited by rare deviant sounds in a series of regular standard sounds ([Näätänen 1992](#); [Schröger 1998](#)). The IR and MMN ERP components are supposed to reflect brain processes related to the detection of unexpected sounds, that is, a rare deviant sound in the oddball paradigm and, respectively, a sound incongruent to the predicted sound in the symbol-to-sound-matching paradigm. In contrast, the early evoked GBR is elicited when a current sound matches the expected sound.

In [Figure 4](#), we present a functional model of symbol-to-sound matching, utilizing data from evoked and induced EEG responses. This model is intended to serve as a framework for future empirical efforts further specifying the involved processes and their neuroanatomical basis. In particular, based on the time course and topography of the incongruency ERP response, it seems likely that the visual symbols activate auditory memory representations regarding the to-be-expected stimulus. Subsequently, this memory representation is compared with the currently presented auditory stimulus. The idea that visual symbols can, in fact, activate auditory memory representations is supported by the results from research on text reading showing that graphemes do activate their associated phonemes while reading (for review, see [Perfetti 1999](#); [Schroeder and Foxe 2005](#)).

The conclusion that the matching process takes place at early stages of auditory stimulus processing is supported by an enhancement of the early auditory-evoked GBR to congruent stimuli, which is interpreted as a sign of the successful matching process. The topography of the early evoked GBR with a frontal maximum is compatible with sources in the auditory cortices, and the latency of the response as early as the first activations of auditory cortex clearly supports this interpretation. The sources of the auditory MLR, presumably corresponding to the early auditory-evoked GBR in the time domain, are well known to reside in primary auditory areas in the superior temporal plane ([Ponton et al. 2002](#)).

Following the early comparison process in case of a mismatch between prediction and stimulation, the IR, N2b, P3a, and P3b are elicited ([Widmann et al. 2004](#)). N2b and P3a are associated with processes of target selection and orienting attention ([Escera et al. 2000](#); [Sokolov et al. 2002](#)). P3b is supposed to be asso-

ciated with the processes of anticipatory response selection here. In case of a match, the auditory stimulus and the visual symbolic information can be successfully bound together to an unequivocal event representation reflected by an induced GBR generated in widely dispersed parietooccipital, inferior temporal, and frontal structures (Gruber et al. 2006), possibly including also multisensory areas in the posterior auditory association cortex and the parietal association cortices (Schroeder and Foxe 2002; Ghazanfar and Schroeder 2006). Such integrated audiovisual representations could exist in, for instance, graphemes which also contain phonemic information. Taken together, according to this model, the early evoked GBR mirrors the comparison process per se, that is, the extraction of the information that is then used for further processing, which in turn results in the percept of a correct sound (reflected by the induced GBR). We suppose that this model can be generalized in the way that not only visual symbolic information but also information from any available source is used by the auditory cognitive system to continuously model the expected acoustic environment and to match the internal model with the experience.

## Conclusions

Our results show that during the audiovisual integration task, the early auditory-evoked GBR is sensitive to qualitative stimulus aspects. Expectations on the upcoming sounds triggered by asynchronously presented visual symbolic stimuli can be compared against the current auditory stimulation at early stages of auditory processing as indicated by the enhancement of the evoked GBR to correctly predicted sounds peaking as early as 42 ms after sound onset. Depending on whether the comparison results in a match or a mismatch, different sets of processes are elicited, with ERPs being more sensitive to processes following a mismatch and induced GBR being indicative of processes following a match of expectation and stimulation. The induced GBR obtained in the symbol-to-sound-matching paradigm is suggested to reflect the formation of a unitary event representation integrating auditory and symbolic visual aspects.

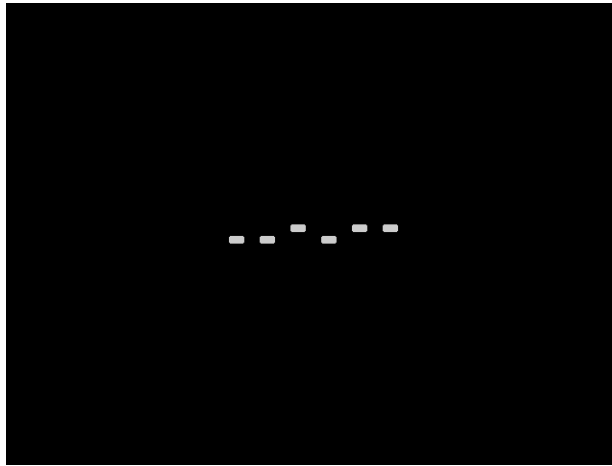
## Acknowledgements

The research was supported by grants from the Deutscher Akademischer Austauschdienst (DAAD; German Academic Exchange Service) and Marie Curie Individual fellowship by European Commission (QLK6-CT-2000-51227).

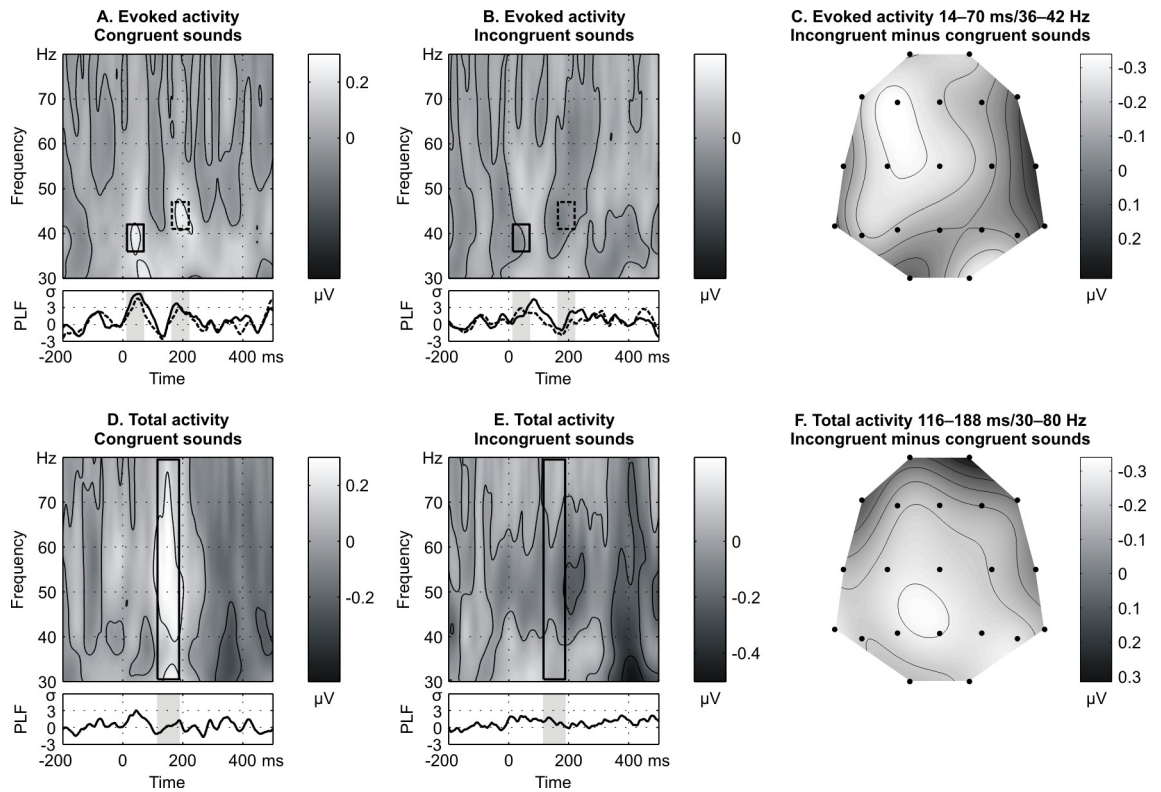
## References

- Basar E, Rosen B, Basar-Eroglu C, Greitschus F. 1987. The associations between 40 Hz-EEG and the middle latency response of the auditory evoked potential. *Int J Neurosci*. 33:103–117. [[PubMed: 3610489](#)]
- Bertrand O, Tallon-Baudry C. 2000. Oscillatory gamma activity in humans: a possible role for object representation. *Int J Psychophysiol*. 38:211–223. [[PubMed: 11102663](#)]
- Debener S, Herrmann CS, Kranczoch C, Gembris D, Engel AK. 2003. Top-down attentional processing enhances auditory evoked gamma band activity. *Neuroreport*. 14:683–686. [[PubMed: 12692463](#)]
- Delorme A, Makeig S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*. 134:9–21. [[PubMed: 15102499](#)]
- Devlin JT, Raley J, Tunbridge E, Lanary K, Floyer-Lea A, Narain C, Cohen I, Behrens T, Jezzard P, Matthews PM, et al. 2003. Functional asymmetry for auditory processing in human primary auditory cortex. *J Neurosci*. 23:11516–11522. [[PubMed: 14684855](#)]
- Engel AK, Fries P, Singer W. 2001. Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci*. 2: 704–716. [[PubMed: 11584308](#)]
- Escera C, Alho K, Schröger E, Winkler I. 2000. Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol Neurootol*. 5:151–166. [[PubMed: 10859410](#)]
- Ford JM, Gray M, Faustman WO, Heinks TH, Mathalon DH. 2005. Reduced gamma-band coherence to distorted feedback during speech when what you say is not what you hear. *Int J Psychophysiol* 57:143–150. [[PubMed: 15967529](#)]
- Ghazanfar AA, Schroeder CE. 2006. Is neocortex essentially multisensory? *Trends Cogn Sci*. 10:278–285. [[PubMed: 16713325](#)]
- Gruber T, Müller MM. 2005. Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. *Cereb Cortex*. 15:109–116. [[PubMed: 15238442](#)]
- Gruber T, Trujillo-Barreto NJ, Giabbiconi CM, Valdes-Sosa PA, Müller MM. 2006. Brain electrical tomography (BET) analysis of induced gamma band responses during a simple object recognition task. *Neuroimage*. 29:888–900. [[PubMed: 16242965](#)]
- Gruber T, Tsivilis D, Montaldi D, Müller MM. 2004. Induced gamma band responses: an early marker of memory encoding and retrieval. *Neuroreport*. 15:1837–1841. [[PubMed: 15257158](#)]

- Herrmann CS, Mecklinger A. 2000. Magnetoencephalographic responses to illusory figures: early evoked gamma is affected by processing of stimulus features. *Int J Psychophysiol.* 38:265–281. [[PubMed: 11102667](#)]
- Herrmann CS, Munk MH, Engel AK. 2004. Cognitive functions of gamma-band activity: memory match and utilization. *Trends Cogn Sci.* 8:347–355. [[PubMed: 15335461](#)]
- Jervis BW, Nichols MJ, Johnson TE, Allen E, Hudson NR. 1983. A fundamental investigation of the composition of auditory evoked potentials. *IEEE Trans Biomed Eng.* 30:43–50. [[PubMed: 6826185](#)]
- Kaiser J, Hertrich I, Ackermann H, Mathiak K, Lutzenberger W. 2005. Hearing lips: gamma-band activity during audiovisual speech perception. *Cereb Cortex.* 15:646–653. [[PubMed: 15342432](#)]
- Kaiser J, Lutzenberger W. 2005. Human gamma-band activity: a window to cognitive processing. *Neuroreport.* 16:207–211. [[PubMed: 15706221](#)]
- Keil A, Gruber T, Müller MM. 2001. Functional correlates of macroscopic high-frequency brain activity in the human visual system. *Neurosci Biobehav Rev.* 25:527–534. [[PubMed: 11595272](#)]
- Makeig S. 1990. A dramatic increase in the auditory middle latency response at very low rates. In: Brunia C, Gaillard A, Kok A, editors. *Psychophysiological brain research*. Tilburg (The Netherlands): Tilburg University Press. p. 56–60.
- Marshall L, Mölle M, Bartsch P. 1996. Event-related gamma band activity during passive and active oddball tasks. *Neuroreport.* 7:1517–1520. [[PubMed: 8856711](#)]
- Müller MM, Keil A, Kissler J, Gruber T. 2001. Suppression of the auditory middle-latency response and evoked gamma-band response in a paired-click paradigm. *Exp Brain Res.* 136:474–479. [[PubMed: 11291728](#)]
- Näätänen R. 1992. *Attention and brain function*. Hillsdale (NJ): Erlbaum.
- Pantev C, Elbert T, Makeig S, Hampson S, Eulitz C, Hoke M. 1993. Relationship of transient and steady-state auditory evoked fields. *Electroencephalogr Clin Neurophysiol.* 88:389–396. [[PubMed: 7691563](#)]
- Pantev C, Makeig S, Hoke M, Galambos R, Hampson S, Gallen C. 1991. Human auditory evoked gamma-band magnetic fields. *Proc Natl Acad Sci USA.* 88:8996–9000. [[PubMed: 1924362](#)]
- Perfetti CA. 1999. Comprehending written language: a blueprint of the reader. In: Hagoort P, Brown C, editors. *Neurocognition of language processing*. Oxford (UK): Oxford University Press. p. 167–208.
- Ponton C, Eggermont JJ, Khosla D, Kwong B, Don M. 2002. Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. *Clin Neurophysiol.* 113: 407–420. [[PubMed: 11897541](#)]
- Schönwiesner M, Rübsem R, von Cramon DY. 2005. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur J Neurosci.* 22:1521–1528. [[PubMed: 16190905](#)]
- Schroeder CE, Foxe J. 2005. Multisensory contributions to low-level, ‘unisensory’ processing. *Curr Opin Neurobiol.* 15:454–458. [[PubMed: 16019202](#)]
- Schroeder CE, Foxe JJ. 2002. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res Cogn Brain Res.* 14:187–198. [[PubMed: 12063142](#)]
- Schröger E. 1998. Measurement and interpretation of the mismatch negativity (MMN). *Behav Res Methods Instrum Comput.* 30:131–145.
- Shah AS, Bressler SL, Knuth KH, Ding M, Mehta AD, Ulbert I, Schroeder CE. 2004. Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cereb Cortex.* 14:476–483. [[PubMed: 15054063](#)]
- Sokolov EN, Spinks JA, Lyytinen H, Näätänen R, editors. 2002. *The orienting response in information processing*. Mahwah (NJ): Erlbaum.
- Tallon-Baudry C, Bertrand O. 1999. Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci.* 3:151–162. [[PubMed: 10322469](#)]
- Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J. 1996. Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *J Neurosci.* 16:4240–4249. [[PubMed: 8753885](#)]
- Tervaniemi M, Hugdahl K. 2003. Lateralization of auditory-cortex functions. *Brain Res Rev.* 43:231–246. [[PubMed: 14629926](#)]
- Tiitinen H, Sinkkonen J, May P, Näätänen R. 1994. The auditory transient 40-Hz response is insensitive to changes in stimulus features. *Neuroreport.* 6:190–192. [[PubMed: 7703412](#)]
- Tiitinen H, Sinkkonen J, Reinikainen K, Alho K, Lavikainen J, Näätänen R. 1993. Selective attention enhances the auditory 40-Hz transient response in humans. *Nature.* 364:59–60. [[PubMed: 8316297](#)]
- Widmann A, Kujala T, Tervaniemi M, Kujala A, Schröger E. 2004. From symbols to sounds: visual symbolic information activates sound representations. *Psychophysiology.* 41:709–715. [[PubMed: 15318877](#)]
- Winkler I, Karmos G, Näätänen R. 1996. Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Res.* 742:239–252. [[PubMed: 9117400](#)]
- Yoo SS, O’Leary HM, Dickey CC, Wei XC, Guttmann CR, Park HW, Panych LP. 2005. Functional asymmetry in human primary auditory cortex: identified from longitudinal fMRI study. *Neurosci Lett.* 383:1–6. [[PubMed: 15936503](#)]
- Yordanova J, Kolev V, Demiralp T. 1997. The phase-locking of auditory gamma band responses in humans is sensitive to task processing. *Neuroreport.* 8:3999–4004. [[PubMed: 9462481](#)]

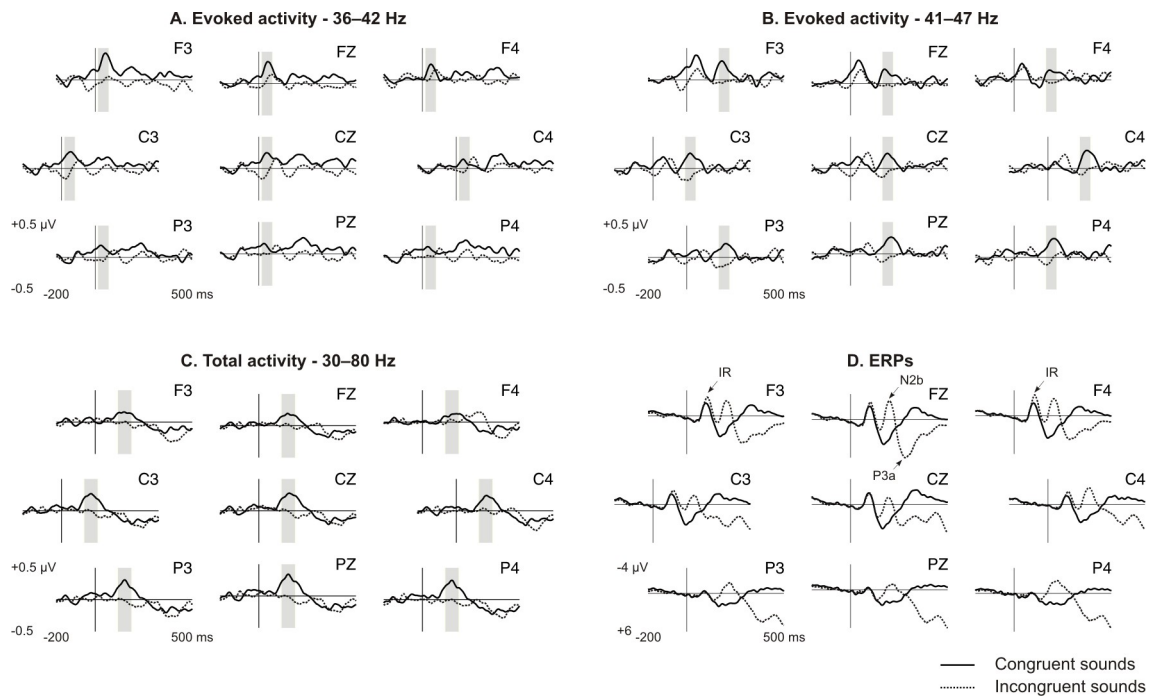


**Figure 1.**  
Sample screenshot of a visual display presented to the subjects.



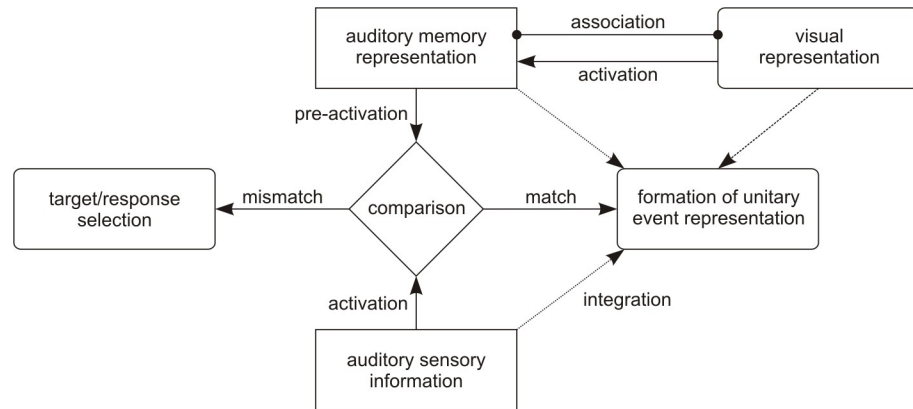
**Figure 2.**

(A, B, D, E) Grand average time-frequency plots of evoked and total oscillatory gamma-band activity for sounds congruent and incongruent to the corresponding visual symbol (averaged across electrode locations F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4; boxes represent the time-frequency windows used for statistical analysis), and phase-locking factor plots (in standard deviations from baseline) within the relevant frequency windows (evoked activity: 36–42 Hz [solid line] and 41–47 Hz [dotted line], total activity: 30–80 Hz). (C, F) Grand average scalp topographies of the incongruent minus congruent difference amplitude for evoked (14–70 ms; 36–42 Hz) and total oscillatory gamma-band activity (116–188 ms; 30–80 Hz).



**Figure 3.**

(A, B) Grand average time-amplitude plots for evoked gamma-band activity (A: 36–42 Hz, B: 41–47 Hz), (C) total gamma-band activity (30–80 Hz) and (D) grand average ERPs for congruent (solid lines) and incongruent stimuli (dotted lines; [Widmann et al. 2004](#)). Bars represent the time windows used for statistical analysis.

**Figure 4.**

A functional model of the processes involved in the symbol-to-sound-matching task. Visual symbolic information activates an auditory representation about the to-be-expected sound that is compared with the current sound (indexed by early evoked GBR). Processing of a mismatch is reflected by IR, N2b, and P3a ERP deflections. Formation of a unitary audiovisual event representation in case of a match is indexed by later induced GBR.