Visual Modulation of Neurons in Auditory Cortex

Our brain integrates the information provided by the different sensory modalities into a coherent percept, and recent studies suggest that this process is not restricted to higher association areas. Here we evaluate the hypothesis that auditory cortical fields are involved in cross-modal processing by probing individual neurons for audiovisual interactions. We find that visual stimuli modulate auditory processing both at the level of field potentials and single-unit activity and already in primary and secondary auditory fields. These interactions strongly depend on a stimulus’ efficacy in driving the neurons but occur independently of stimulus category and for naturalistic as well as artificial stimuli. In addition, interactions are sensitive to the relative timing of audiovisual stimuli and are strongest when visual stimuli lead by 20–80 msec. Exploring the underlying mechanisms, we find that enhancement correlates with the resetting of slow (~10 Hz) oscillations to a phase angle of optimal excitability. These results demonstrate that visual stimuli can modulate the firing of neurons in auditory cortex in a manner that depends on stimulus efficacy and timing. These neurons thus meet the criteria for sensory integration and provide the auditory modality with multisensory contextual information about co-occurring environmental events.

Keywords: audiovisual, cross-modal, electrophysiology, local field potential, macaque monkey, multisensory

Introduction

Sensory systems evolve in environments composed of events that rarely influence only a single sensory modality. Thus for animals to survive and properly interact with their environment, they must not only monitor each individual sense but also merge information across modalities. Behavioral advantages of sensory integration are known where combined sensory input lowers reaction times (Hershenson 1962) or improves detection of faint stimuli (Stein et al. 1989; Driver and Spence 1998; McDonald et al. 2000). In addition, sensory combinations can alter the quality of the sensory percept or result in illusions (McGurk and MacDonald 1976; Jousmäki and Hari 1998; Shams et al. 2000). Commonly, it is assumed that sensory integration occurs mostly in higher association cortices and specialized subcortical structures (Jones and Powell 1970; Mesulam 1998). Support for this view comes from earlier anatomical and electrophysiological studies finding zones of sensory convergence and cross-modal responses only in higher level “associative” cortical areas; prominently, neurons responding to stimulation of several modalities were discovered in association areas of the temporal, parietal, and frontal lobes (Benevento et al. 1977; Hyvarinen and Shelepina 1979; Leimonen et al. 1980; Bruce et al. 1981; Rizzolatti et al. 1981; Hikosaka et al. 1988).

Accumulating evidence with recent techniques, however, has been updating this view and suggests that areas hitherto regarded as unisensory can indeed be activated, or at least modulated, by stimulation of several senses (Schroeder and Foxe 2005; Ghazanfar and Schroeder 2006; Kayser and Logothetis 2007). Prominently, functional imaging studies revealed cross-modal interactions that were either localized to areas classically considered unisensory (Calvert 2001) or found interactions that occurred early enough in time to imply a source in lower sensory areas (Giard and Peronnet 1999; Foxe et al. 2000; Molholm et al. 2002; Murray et al. 2005). The auditory system appears to be an exquisite model system for these effects, and a host of studies reported visual activations and visual modulation of auditory responses in a variety of paradigms (Calvert et al. 1997; Laurienti et al. 2002; van Atteveldt et al. 2004; Pekkola, Ojanen, Autti, Jääskeläinen, Motonen, and Sams 2005; van Wassenhove et al. 2005; Lehmann et al. 2006). Additionally, somatosensory stimuli were found to have a similar impact on activity in early auditory areas (Foxe et al. 2002; Lutkenhoner et al. 2002; Kayser et al. 2005; Murray et al. 2005; Schurmann et al. 2006).

Despite the number of imaging studies localizing cross-modal influences to auditory cortical areas, there is little insight into how these manifest at the level of neuronal activity. Especially, imaging studies cannot reveal whether these cross-modal effects have a strictly modulatory influence that is visible only in subthreshold activity or whether individual neurons indeed integrate information from the different modalities. In fact, only few electrophysiological studies directly quantified cross-modal influences in auditory cortex. Whereas 1 study found strong visual responses in the auditory cortex of anesthetized ferrets (Bizley et al. 2006), studies on primates suggest that visual modulation is primarily a subthreshold phenomenon, mostly occurring at the level of “local field potentials” (LFP) and multi-unit activity (MUA) (Schroeder and Foxe 2002; Ghazanfar et al. 2005), or resulting from learned associations between visual and auditory events in highly trained animals (Brosch et al. 2005).

In the present study, we scrutinized the impact of visual stimuli on auditory responses by recording LFP and single units from the auditory cortex of macaque monkeys. The recordings were focused on the posterior auditory cortex, which exhibited strongest cross-modal influences in previous imaging studies (Kayser et al. 2005; Kayser, Petkov, et al. 2007). Our results unequivocally demonstrate an impact of visual stimuli on auditory spiking responses of individual neurons, and we find that cross-modal interactions depend on the efficacy and relative timing of the stimuli, are independent of the type of stimulus used, and occur both in primary and secondary
auditory cortex. As a result, we conclude that the activity communicated from these auditory areas to subsequent sensory processing stages not only contains information about acoustical events but also reflects aspects of the visual environment.

Materials and Methods

Electrophysiological Recording Procedures
Three adult rhesus monkeys (Macaca mulatto) participated in these experiments. All procedures were approved by the local authorities (Regierungspräsidium) and were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals. Prior to the experiments, form-fitting head posts and recording chambers were implanted during an aseptic and sterile surgical procedure (Logothetis et al. 1999). The chambers were positioned based on preoperative magnetic resonance images and stereotaxic coordinates (antero-posterior [AP] = 3 mm, medio-lateral [ML] = 21 mm for animals M1 and M2 and AP = 8 mm, ML = 24 for animal M3) allowing vertical access to the auditory regions on the superior temporal plane. A custom-made multielectrode system was used to lower up to 8 microelectrodes (FHC Inc., Bowdoinham, ME, 0.8-8 MOhm impedance) through a grid placed on the recording chamber down to auditory cortex on the superior temporal plane. The coordinates of each electrode were noted along the AP and ML axes. Signals were amplified using a custom-modified Alpha Omega amplifier system (Alpha Omega GmbH, Ubstadt-Weiher, Germany), filtered between 4 Hz and 10 kHz (4-point Butterworth filter), and digitized at 20.83 kHz. It should be that with such a setting, the power around 10 kHz is about 15 dB below the signal power at 500 Hz and 45 dB below the power at 30 Hz. Hence, it is extremely unlikely that residual signals above 10 kHz, which might be aliased into the low-frequency components, could affect either the LFP or the spiking data. Recordings were performed in a darkened and anechoic booth (Ulbrick acoustic GmbH, Germany).

During each recording session, we advanced the electrodes until the activity of 1 or more neurons was isolated and could be driven by any of our search sounds. These included pure tones, frequency modulated sweeps, noise bursts, clicks, and animal vocalizations. We made no attempt to select neurons based on their response preferences.

The animals were trained on a visual fixation task and received juice rewards for correct performance. Each trial started once the animal engaged in a central fixation dot and consisted of a 500-msec baseline period, a 1200-msec stimulation period, and a 300-msec poststimulus period. Two animals (M2 and M3) fixated in a 4-degree diameter window, whereas the 3rd animal (M1) required a larger window (8 degrees). We found similar visual modulation in auditory cortex in all animals, and importantly, small eye movements during fixation did not correlate with the visual modulation effects (see Supplementary Fig. 2 and Discussion).

Recording sites were assigned to the auditory core (primary auditory cortex) and auditory belt regions based on stereotaxic coordinates, frequency maps constructed for each animal, and the responsiveness for tone versus band-passed stimuli. Most of our recording sites were located in the caudal portions of primary auditory cortex (mainly field A1) and in the caudal belt (fields caudo-medial [CM] and caudo-lateral [CL], see Supplementary Fig. 1A). Core and belt were distinguished using a suprathreshold procedure (Schroeder et al. 2001; Fu et al. 2004), which probes frequency selectivity using tones and band-passed noise stimuli of different frequency and intensities well above the neurons threshold. Previous work demonstrated the equivalence of this method to the classical, and threshold-based, determination of center frequency (Merzenich and Brugge 1973; Kosaki et al. 1997). The Supplementary Figure 1A shows the different response properties in core and belt, using analogous criteria as used in a recent study comparing different auditory fields in monkey auditory cortex (Lakatos, Pincze, et al. 2005). Sites in the auditory cortex were distinguished from deeper recording sites from the superior temporal sulcus (STS), using the depth of the electrodes, the occurrence of several millimeters of white matter between auditory cortex and STS the longer response latency in the STS, and the prominence of visual responses in the STS.

Acoustical and Visual Stimuli

Sounds were stored as WAV files, amplified using a Yamaha amplifier (AX-96G), and delivered from 2 free field speakers (JBL Professional, Northridge, CA), which were positioned at ear level 70 cm from the head and from 50 degrees to the left and right. Sound presentation was calibrated using a condenser microphone (Bruel & Kjær 4188 and a 2238 Mediator sound level meter, Bruel & Kjær GmbH, Bremen, Germany) to ensure a linear (±1 dB) transfer function (between 88 Hz and 20 kHz). Sounds were presented at an average intensity of 65 dB SPL, unless stated otherwise. Visual stimuli were presented on a 21-inch gamma-corrected monitor, at a distance of 97 cm from the animal and covered a visual field of 24 × 18 degrees. The naturalistic stimuli had a mean luminance of 20 cd/m².

We employed several types of stimuli. First, to characterize auditory tuning and response properties, we used 2 paradigms, 1 consisting of band-passed noise and 1 of pure tones of different frequencies (Rauschecker et al. 1995; Recanzone, Guard, and Phan 2000; Rauschecker and Tian 2004). Both types of stimuli were presented as pseudorandom sequences of 8 repeats, with individual stimulus duration of 50 msec and pauses of 80 msec. In this way, these stimuli are reminiscent of amplitude-modulated tones or noise, which are known to well drive auditory neurons (Creutzfeldt et al. 1980; Bieser and Muller-Peuss 1996; Liang et al. 2002). Pure tones (15 frequencies) ranged from 125 Hz to 16 kHz in half-octave steps, and the band-passed noise (7 bands, 1 octave width) covered the same frequency range. All stimuli were cosine ramped (on/off, 8 msec). See also Kayser, Petkov, and Logothetis (2007) for details.

The audiovisual set consisted of complex, naturalistic scenes of 1200 msec duration. A total of 16 different audiovisual scenes were chosen that consisted of 4 categories (with 4 examples from each category): 1) close-ups of conspecific vocalizing animals (2 cues, 1 grunt, and 1 lip smack); 2) other monkeys making noises in a natural habitat but not vocalizing (baboons and gorillas); 3) different animals in natural settings (lion, birds, and elephant); and 4) artificial stimuli (2 cartoon movies, moving colored squares accompanied by synchronized beeps, and audiovisual pink noise). Each stimulus consisted of the movie and the matching sound. Auditory only, visual only, or combined audiovisual stimuli were presented in random order.

It should be noted that these naturalistic auditory stimuli do not possess sharp onsets as do artificially generated tones or band-passed noise. Because this effective onset latency of the acoustical stimuli is important to interpret the response latencies of the neuronal activity, we computed the onset latency of each sound as the time point where it exceeded 10% of its maximal amplitude for at least 5 msec. The resulting latencies were in a range from 6 to 70 msec (median value 29 msec).

A set of simplistic stimuli was used to systematically assess the impact of varying the stimulus onset asynchrony (SOA). The visual stimulus was a white screen (luminance = 35 cd/m², duration = 20 msec), and the auditory stimulus was a “click-like” short noise burst (20 msec duration, 72 dB SPL, power concentrated between 2 and 8 kHz). The SOA was varied from +320 to −320 msec, as indicated in Figure 6.

A reduced audiovisual set was used to test the influence of the relative timing (SOA) during stimulation with naturalistic stimuli (Fig. 7). Three scenes were chosen from the above set (2 from group 1 and 1 from group 4) and were presented as synchronous audiovisual pair (SOA = 0) as in the paradigm above—and with the visual stimulus starting (SOA = −160 msec before the auditory stimulus). The sound of 160 msec was chosen as it reflects an offset of 4 video frames and produces a clearly discernable difference in the audiovisual alignment. It should be noted that this offset is beyond the SOA range that produced interactions in the paradigm with the simplistic stimuli (Fig. 6). The main source for this discrepancy is in the sluggishness of the naturalistic visual stimuli. Clearly discernable changes occur only across several movie frames (i.e., 80 msec or more). Hence, an SOA of 160 msec will only differ little from an SOA of 80 msec, suggesting that the results from both paradigms reveal similar mechanisms.

Data Analysis—Extraction of Neuronal Responses
The data were analyzed in Matlab (Mathworks Inc., Natick, MA). The LFP was obtained by low-pass filtering the raw data at 300 Hz (3rd-
order Butterworth filter) and convolving individual trials with Gabor wavelets of different center frequencies (5, 10, 20, 40, 60, 80, 120 Hz, and a bandwidth of 0.83). The amplitude of this convolution yields an estimate of the signal power in each frequency band (Tallon-Baudry et al. 1998), whereas the angle of the convolution estimates the phase of the respective oscillation. The LFP response was obtained by normalizing the power of individual frequency bands to units of standard deviations (SDs) from baseline (z-score, a 300-msec window before stimulus onset) (Logothetis et al. 2001). It should be noted that at the lowest frequencies of the LFP, for example around 10 Hz, the filtering procedure blurs the amplitude signal, which can result in low-frequency responses that apparently commence already before stimulus onset (cf., Fig. 1). The “spike-sorted activity” (single-unit activity [SUA]/MUA) and multi-unit clusters (MUA) was extracted using commercial spike-sorting software (Plexon Offline Sorter, Plexon Inc, Dallas, TX) after high-pass filtering the raw signal at 50 Hz. Spike times were saved at a resolution of 1 msec, and peri-stimulus time-histograms were obtained using bins of 5 msec and Gaussian smoothing (10 msec full width at half height). For many recording sites, spike sorting could extract SUA (signal to noise ratio > 7, spike valley [peak] divided by signal SD). For other sites, however, the spike sorting did not yield well-separated clusters and the activity was deemed as multi-unit; for most figures and analysis, we grouped single- and multi-unit sites together (SUA/MUA), but results for single units are reported where appropriate. For all response measures, that is, LFP power or MUA/SUA peri-stimulus time histograms, we 1st averaged the data across all trials before computing the response amplitude or latency.

“Significant responses” to sensory stimulation were determined by comparing the response amplitude of the average response to the response variability during the baseline period. Arithmetically this was done by normalizing the average response to SDs with respect to baseline, and a response was regarded as significant if this z-score breached 2 SDs during a continuous period of at least 50 msec. This test was applied separately to each condition (auditory, visual, audiovisual) and to the LFP and each unit recorded on an electrode. “Response amplitudes” were computed from the trial averaged response using different time windows during the stimulus presentation (50–250, 250–450, 450–650, and 650–850 msec). The “response latency” was computed as the 1st time bin at which the averaged response exceeded 2 SD of its baseline for at least 10 consecutive milliseconds. To confirm that any difference in auditory and audiovisual latencies was not a result of the specific criterion used, we also computed latencies by searching for time bins with spike numbers greater than the 95% percentile of the Poisson distribution estimated from the baseline period (Pouget et al. 2005).

“Phase consistency” of different LFP frequency bands across trials was estimated as follows: the wavelet transformed single trial data was normalized (at each time point) to an absolute value of 1, trials were averaged, and the modulus of the resulting vector was computed. Perfect phase alignment across trials would result in a consistency value of 1, whereas random phases result in a small consistency value. The resulting time course quantifies the phase consistency as a function of time (Fig. 8D), and the average phase consistency was obtained as the average during the response time window (50–250 msec).

From all recording sites, we analyzed only those exhibiting significant responses to at least 1 of the 3 categories (auditory, visual, or audiovisual) in both, the low-frequency LFP (10 Hz) and the spiking activity. For the main paradigm, this was the case for 146 recording sites (97, 75, and 35 sites from monkeys 1, 2, and 3, respectively) and a total of 207 units (of these 99 SUA).

**Data Analysis—Multisensory Interactions**

Recording sites were termed “multisensory” when they either exhibited a significant response to both auditory only and visual only stimuli or when the response to the auditory stimulus was significantly modulated by the visual stimulus. A significant modulation was detected using 2 different tests, 1 quantifying the response enhancement and 1

![Figure 1](https://academic.oup.com/cercor/article-abstract/18/7/1560/420723)
The “time course” and onset latency of multisensory enhancement and suppression were estimated as follows. For each unit, we computed the difference between the audiovisual and the auditory response. This difference time course was averaged separately for units that showed enhanced or suppressed responses in a particular time window (cf., Fig. 4B). The onset latency of the effect was defined as the 1st time point during which the time course breached 2 SDs of its baseline level for at least 10 msec; repeating the same analysis for other thresholds yielded different values for the latency but similar differences between enhancement and suppression as reported.

The paradigm with simplistic stimuli and different SOAs (Fig. 6) was analyzed as follows. The response amplitude was computed using a 150-msec window after the auditory stimulus. To determine the impact of the visual stimulus, individual SOA values were compared with a paradigm in which the visual stimulus could not have any influence; this was the case in the SOA = ~320 msec condition, which hence served as “auditory-only” baseline.

Results

Visual Modulation in Auditory Cortex: Examples and Latency

Complex auditory scenes evoked robust responses throughout posterior auditory cortex (n = 146 LFP sites, 207 SUA/MUA units, of these 99 were characterized as SUA). This is illustrated by the examples in Figure 1A, which show strong responses to both plain auditory and audiovisual stimuli in the low-frequency LFP (10 Hz) and firing rates (SUA/MUA). In addition, these examples exhibit the different facets of visual modulation that can be observed: auditory responses can be enhanced, reduced, or be unaffected by a simultaneous visual stimulus. Enhancement manifests as an audiovisual response that is stronger than the auditory response (e.g., the upper LFP),

quantifying the response additivity (Stein and Meredith 1993; Perrault et al. 2005; Stanford et al. 2005; Avillac et al. 2007). First, a significant enhancement or suppression of the auditory response by a visual stimulus was detected by comparing the auditory and audiovisual conditions across trials (paired t-test); sites that reached a significance level of $P < 0.01$ were regarded as either significantly enhanced or suppressed. The strength of this effect was quantified using the following index: enhancement = $(AV - A)/(A + AV) \times 100$, where $A$ and $AV$ reflect the auditory and audiovisual responses after subtraction of the baseline activity (baseline normalized response). It should be noted that in the original definition of this index, the multisensory response is compared with the maximal unisensory response; however, in the auditory data, the maximal unisensory response was always elicited by the auditory condition.

And 2nd, to determine whether the modulation of the response was compatible with an additive summation, or showed supra- or subadditive effects, a recently proposed bootstrapping procedure was implemented (Stanford et al. 2005): the audiovisual response was compared with all possible summations of auditory and visual responses from individual trials and sites for which the audiovisual response was sufficiently far from the bootstrapped distribution (at least $P < 0.01$ based on a $t$-test) were regarded as nonadditive. The deviation from additivity was quantified using the following index: additivity = $(AV - (A + V))/((AV + (A + V)) \times 100$, where $A$, $V$, and $AV$ reflect the auditory, visual, and audiovisual responses after subtraction of the baseline activity. Please note that some studies refer to additive interactions also as being “linear” and sub- or supra-additive interactions as sub- or supralinear.

“Multisensory units” (Fig. 2C) were classified according to the following rules. An additive multisensory effect was defined as having a significant response to the auditory only stimulus plus either a significant response to the visual only stimulus or a significant enhancement (or suppression); however, units were regarded as additive multisensory only if the additivity index did not reach a significant effect. Nonadditive multisensory units were defined as having a significant response to at least 1 condition plus significant effect in the additivity index.

The distribution of the enhancement and the additivity index across sites for LFP and SUA/MUA. Arrows indicate the mean values. The enhancement index characterizes whether the auditory response was enhanced or suppressed by the visual stimulus, and the additivity index characterizes how much the audiovisual response deviates from an additive summation of the auditory and visual responses. (C) Distribution of sites with significant ($P < 0.01$) multisensory effects. Gray: sites with only auditory responses. Orange: sites with additive multisensory effects. Red: sites with nonadditive multisensory effects. Additive multisensory sites have (in addition to an auditory response $A$) either a significant visual response ($V$) or a significant enhancement or suppression effect (Enh) or both. Nonadditive multisensory sites have a significant response in at least 1 condition and have either a significant supra-additive (Sup) or a significant subadditive response (Sub). (D) Scatter plots of the enhancement index in the SUA/MUA and low-frequency LFP at the same electrode, shown for all stimuli (upper panel) and restricted to responses to the worst and the best stimulus category for each site (lower panels). Numbers and color coding indicate the percentage of units in each quadrant. On average across all stimuli, the firing rates and LFP at individual electrodes often show a differential modulation, enhancement of the LFP, and suppression of the firing rates. This imbalance is shifted toward common enhancement or suppression when considering only the best or worst category.

Figure 2. Response amplitudes and multisensory indices. (A) Response amplitudes for the different LFP frequency bands and firing rates during the initial response window (50-250 msec). The inset indicates the firing rates averaged for single units (SUA) only. Error bars indicate the mean and standard error of the mean, black stars indicate a significant difference (paired $t$-test) between the auditory and audiovisual conditions, and red stars indicate a significant visual response ($t$-test vs. zero); *$P < 0.05$, **$P < 0.01$. (B) Distribution of the enhancement index and the additivity index across sites for LFP and SUA/MUA. Arrows indicate the mean values. The enhancement index characterizes whether the auditory response was enhanced or suppressed by the visual stimulus, and the additivity index characterizes how much the audiovisual response deviates from an additive summation of the auditory and visual responses.

Downloaded from https://academic.oup.com/cercor/article-abstract/18/7/1560/420723 by guest on 21 August 2018
whereas suppression manifests as a weaker audiovisual response (e.g., the middle LFP and all SUA/MUA examples). In addition, some sites exhibit equal auditory and audiovisual responses with considerable responses elicited by plain visual stimulation (e.g., the lower LFP). Noteworthy, audiovisual interactions are not restricted to population activity but are also visible in the responses of isolated single units (middle SUA). In the following, we will 1st compare the response amplitudes for the different conditions and then compute response indices to quantify the strength and nature of the interaction.

Figure 1B displays the grand mean responses, which allow a more systematic assessment of response strength and timing. Comparing the amplitudes across conditions reveals a clear impact of the visual stimulus: For the low-frequency LFP, the peak response is strongest for the audiovisual stimulus and even plain visual stimulation elicits a clearly discernible response. The firing rates, in contrast, show a weaker response to the audiovisual compared with the auditory stimulus and no clear visual response. Yet, given the variability of individual neurons in their baseline activity, only a more sophisticated analysis of response amplitudes can fully reveal the different patterns of cross-modal interactions (see below). Please note that the low-frequency LFP response seems to commence before stimulus onset as a result of the long frequency filter required for quantifying the slow-wave activity (cf., Materials and Methods).

Enlarging the response onset (Fig. 1C) suggests that the latency of the auditory response is unaffected by the visual stimulus. The average SUA/MUA responses in auditory and audiovisual conditions overlap well, and estimates of onset latency return statistically comparable values (auditory: median 36 msec, audiovisual: median 40 msec; Wilcoxon signed-rank test \( P = 0.19 \)). When interpreting these absolute latencies, it should be noted that many of the used naturalistic stimuli do not immediately rise to full intensity but rather have a smooth onset, quite in contrast to laboratory stimuli such as tones or noise bursts. We estimated the latency of these natural sounds (see Materials and Methods) and found a median value around 30 msec. Taking this number into account suggests that the effective onset latency in our data was probably shorter than the numbers reported above and well comparable to those reported in the literature (Recanzone, Guard, et al. 2000; Lakatos, Pincze, et al. 2005).

**Response Amplitudes and Multisensory Indices**

Computing response amplitudes during a 200-msec window (50–250 msec after stimulus onset, see Materials and Methods) and accounting for differences in spontaneous baseline activity allow a systematic assessment of the visual stimulus’ impact on the response (Fig. 2A). The analysis across all stimulus categories reveals a dissociation of how visual stimuli influence the different LFP frequency bands. Whereas low frequencies (up to 10 Hz) show a significant visual response (t-test vs. zero, \( P < 10^{-4} \)), this visual response is absent at higher frequencies. In addition, low frequencies have an audiovisual response that is significantly stronger than the auditory response (paired \( t \)-test, \( P < 0.05 \)), whereas higher frequencies show the opposite result; that is a stronger audiovisual response (\( P < 0.01 \)). At the level of SUA/MUA, in contrast, visual stimuli do not elicit significant responses by themselves (\( P = 0.18 \)). Yet, the response to the audiovisual stimulus is significantly reduced compared with the auditory response (paired \( t \)-test, \( P < 0.05 \)).

Importantly, restricting the analysis to SUA reveals a similar result (see Fig. 2A, inset): no visual response (\( P = 0.19 \)) and a significantly lower response to the audiovisual stimulus (\( P < 0.05 \)). All in all, the present data provide a clear demonstration of visual stimuli altering neuronal firing rates in auditory cortex.

To quantify the strength and nature of this audiovisual interaction, we employed 2 previously established criteria (Stein 1998; Stanford et al. 2005; Avillac et al. 2007). The “enhancement index” compares the bimodal response to the maximal unisensory response and quantifies how much the auditory response is enhanced or suppressed by a visual stimulus. The “additivity index” compares the bimodal response to the sum of the 2 unimodal responses and quantifies how much the audiovisual response deviates from an additive superposition of the unisensory responses. Hence, whereas the 1st index only captures how much the auditory response is altered by a visual stimulus, the 2nd index characterizes the response with respect to a model where the responses to different modalities are additively superimposed.

Figure 2B (left panel) displays the enhancement index for the low-frequency LFP and the firing rates, obtained from the response time window 50–250 msec. For the LFP, the mean enhancement is \( +16\% \) and the majority of recording sites (58%, significantly more than expected by chance: \( \chi^2 = 4.1, P < 0.05 \)) have a positive index. For the SUA/MUA, the mean enhancement is \( -20.5\% \) and the majority of the units have a negative index, that is, show response suppression (63%, \( \chi^2 = 8.2, P < 0.01 \)). Again, restricting the analysis to the SUA results in comparable numbers (mean index: \( -11\% \), 68% of the units, \( \chi^2 = 13, P < 0.001 \)). Together, these results suggest a differential impact of the visual stimulus on field potentials and firing rates, on average across stimuli, the low-frequency LFP is enhanced, while the firing rates are generally reduced.

The right panel in Figure 2B displays the additivity index and illustrates that the audiovisual interaction shows a subadditive behavior, both for LFP and firing rates. The additivity index is negative for 60% of the LFP sites (mean index \( -10.5\% \)) and for 75% of the SUA/MUA (mean index \( -13.8\% \)). When restricted to single units, the subadditivity becomes even more prominent (mean index \( -16\% \) and 82% of the sites). The subadditivity of the firing rates directly reflects the overall response suppression, as for most neurons a visual response is absent. For the LFP, the subadditivity reveals that the enhancement of the auditory response by the visual stimulus is less than the response elicited by the visual stimulus itself.

To determine the percentage of individual recording sites with a significant multisensory effect, we tested individual sites and both indices for a significant difference from zero. These criteria are routinely employed to detect multisensory neurons both in cortical and subcortical structures, and their complementary nature allows discerning an additive combination of auditory and visual responses from nonadditive multisensory effects (Stein 1998; Perrault et al. 2005; Stanford et al. 2005; Sugihara et al. 2006; Avillac et al. 2007). In our data set, a small fraction of sites reaches a significant effect at the \( P < 0.01 \) level and Figure 2C summarizes the different multisensory properties across all recording sites. The vast majority of sites exhibit only a significant auditory response, but neither an additional visual response nor any significant modulation (74% LFP, 88% SUA/MUA). In addition, a fraction of recording sites (19% LFP, 4% SUA/MUA) exhibits an additive multisensory response, defined as either a significant response to both auditory only
and visual only stimulation or as a significant auditory response plus a significant modulation by the visual stimulus. Last, about 8% (both LFP and units) show a nonadditive multisensory response, defined as significant supra- or subadditive response summation (see Fig. 2C or Materials and Methods for details). Analysis of single units revealed similar numbers (SUA: 88% auditory only, 3% additive, and 9% nonadditive). Together, these results reveal a diverse impact of the visual stimulus on the oscillatory slow-wave activity and neuronal firing rates. As we will see next, this visual modulation at a particular site is not fixed but depends on the efficacy of the stimulus in driving the neuronal activity (see Fig. 2D lower panel).

**Stimulus Type and Inverse Effectiveness**

The cross-modal interaction between auditory and visual signals does not depend on the stimulus category (Fig. 3). We quantified the interactions separately for each of the 4 stimulus categories (conspicuous vocalizing macaques, heterospecific primates in natural settings, other animals in natural settings, artificial stimuli), and Figure 3A displays the fraction of LFP and SUA/MUA sites exhibiting response enhancement or suppression separately for each category (during the initial response window, 50–250 msec, averaged across all stimuli within each category). Although most stimuli follow the main trend of the respective signal, enhancement of LFP and suppression of SUA/MUA, only the artificial stimuli reach a fraction significantly different from chance level ($\chi^2$ test, $P < 0.05$). In fact, this stimulus category produces the strongest dissociation between LFP and SUA/MUA. However, the enhancement index revealed no significant effect of the interaction strength across stimulus categories ($F = 1.97$, $P = 0.1$). As a consequence, our results do not provide good evidence that a specific kind of naturalistic stimulus, for example vocalizing conspecifics, is especially prone to cross-modal interactions.

For individual recording sites, though, the enhancement index can vary considerably between different stimuli, with a clear relation between enhancement and efficacy of the stimulus. Figure 3B displays the enhancement index averaged separately across the best and worst stimulus categories. The best/worst categories for a given site are those eliciting the strongest/weakest responses. The figure reveals a positive enhancement index for the worst category but a negative index for the best category—demonstrating overall enhancement for stimuli causing weaker responses and suppression for stimuli eliciting stronger responses. This effect reaches high significance, both for the LFP and the firing rates (see $P$ values indicated in the figure, as well as the fraction of sites having positive and negative indices, respectively). Noteworthy, the best and worst categories were not significantly biased toward any of the stimulus categories, and each stimulus category served as best/worst stimulus for a comparable number of units (macaques 27%, primates 25%, animals 24%, and artificial stimuli 23% as worst category, respectively, not different from chance $\chi^2$ test, $P = 0.3$).

The efficacy of the stimulus in eliciting auditory responses also affects the relationship between how visual modulation affects the LFP and firing rates (Fig. 2D). On average across all stimuli, the low-frequency LFP was enhanced whereas the SUA/MUA was suppressed. This imbalance is shifted toward common enhancement or suppression when considering only the best or worst stimulus (Fig. 2D, lower panel). All in all, this demonstrates that slow-wave activity and firing rates can—in part—be independently modulated by visual stimuli and that the efficacy of the respective auditory stimulus is a good predictor for the kind of interaction to expect.

**Figure 3.** Relevance of stimulus category and stimulus effectiveness. (A) Fraction of sites (units) showing enhancement or suppression for each of the stimulus categories (time window 50–250 msec). For this analysis, the enhancement index was averaged across all stimuli within each category. Stars indicate significant deviations from chance level ($\chi^2$ test, $P < 0.05$). (B) Average enhancement index for the best and worst category. Error bars indicate the mean and standard error of the mean across sites, $P$ values refer to a $t$-test versus zero, and numbers indicate the fraction of sites (or units) with a positive or negative index. (C) Logarithmic scatter plot of the response strength versus the enhancement index. Each dot represents 1 site; for simplicity, only sites with positive enhancement index are shown.
In addition, this finding provides a nice confirmation of the principle of inverse effectiveness, which is often associated with sensory integration (Stein and Meredith 1993; Perrault et al. 2005; Avillac et al. 2007). This principle is based on the reasoning that merging sensory information should be especially helpful when individual senses alone yield only little evidence about the environment. Assuming a direct correlation between neuronal responses and sensory evidence, this translates to stronger interaction for stimuli that elicit weaker responses. As the cross-modal interaction can have 2 different directions, we analyzed sites with response enhancement and suppression separately. For the former, one would expect a decrease of enhancement with increasing efficacy of the stimulus, whereas for the latter one would expect a decrease of the suppression, that is an increase of the negative enhancement index, with increasing efficacy. Figure 3C demonstrates that the principle of inverse effectiveness holds for the present data both for the LFP and SUA/MUA. The scatter plots reveal a significant negative correlation between the enhancement index and the auditory response for sites with response enhancement (Spearman rank correlation, LFP: \( r = -0.74, P < 10^{-8}; \) SUA/MUA: \( r = -0.63, P < 10^{-7} \)). The same analysis on sites with response suppression reveals a significant positive correlation (LFP: \( r = 0.41, P < 0.01; \) SUA/MUA: \( r = 0.39, P < 10^{-5} \)). Together, these findings demonstrate that the nature and strength of audiovisual interaction in auditory cortex depends on the efficacy of the individual stimuli in driving neuronal activity, rather than on the particular stimulus type.

**Time Course of Visual Modulation**

From the above finding that the stimulus efficacy determines the nature of the interaction, one could expect that during stimulation with longer naturalistic stimuli, whose structure changes continuously, a state of enhancement or suppression might persist only for a short interval and might be immediately followed by the opposite effect. To test whether cross-modal interactions indeed persist only over short intervals, we computed the time course of the cross-modal interactions for the SUA/MUA activity (Fig. 4).

Figure 4. Time course of cross-modal interaction. (A) Difference between the audiovisual and auditory responses separately for units showing response enhancement (dark gray) and suppression (light gray). For this panel, units were sorted according to their main effect during the initial response window (see gray box). Solid lines indicate the mean, whereas the shaded area indicates the 1% confidence interval (bootstrap estimate). The number of units in each average is indicated on the side. Arrows at the bottom indicate the mean onset latency of the effect; arrows on top indicate the offset (see text). (B) Difference between the audiovisual and auditory responses separately for units sorted according to their main effect during later response windows (see gray boxes on top).

Panel A displays the difference between the auditory and audiovisual responses, separately for units with response enhancement and suppression during the initial response window (50–250 msec). Although sites were sorted according to their main effect during this window, the interactions persist for a longer time. First, both interactions have similar onset latencies (enhancement: \( 76 \pm 5.7 \) msec, mean \( \pm \) standard error of the mean, suppression \( 79 \pm 6.8 \) msec). Second, both effects persist beyond the response window, with suppression lasting nearly twice as long. To quantify this, we determined the duration of each effect by computing the 1st point at which the confidence interval \( (P < 0.01) \) reaches zero: this was 311 msec for enhancement and 596 msec for suppression. We verified that this result is not restricted to the initial phase of the stimulus. Figure 4B displays the difference between the auditory and audiovisual responses sorted according to units showing enhancement or suppression in later time windows. Although the units are the same as those used in panel A, they are sorted differently into the 2 categories in each panel of Figure 4B. For each time window, suppression persists over longer periods than enhancement. And for later time windows, suppression already occurs before the particular time window chosen to sort the units. These results demonstrate that the pattern of enhancement and suppression is not rapidly changing and suppression dominates large proportions of the SUA/MUA responses.

**Primary versus Nonprimary Auditory Cortex**

Our recording sites covered a larger area in posterior auditory cortex, including both primary (auditory core field A1) and secondary auditory fields (auditory belt fields CM and CL). We identified core and belt regions using the tonotopic frequency gradient, the selectivity of sound frequency tuning, and the responsiveness to band-passed noise and pure tones using a suprathreshold procedure (Lakatos, Pinçze, et al. 2005) (see Supplementary Fig. 1). Previous studies, from functional imaging and single-unit recordings, collectively showed that both regions can be modulated by visual stimulation (Calvert 2001; Brosch et al. 2005; Ghazanfar et al. 2005; Bizley et al.
In agreement with this, we do not find prominent differences between recording sites in core and belt (Fig. 5). The SUA/MUA firing rates for the different conditions do not differ between regions (2-sample t-tests, all P > 0.05) and in the low-frequency LFP only the visual condition shows significantly stronger activity in the belt (P < 0.05), suggesting a slightly enhanced visual input at the secondary auditory stage (Fig. 5B). In addition, in both regions is the LFP enhancement index biased toward the positive side, whereas the same index is biased toward the negative for the MUA/SUA; no comparison of the enhancement index between core and belt reached significance (2-sample t-test, P > 0.05 for LFP and SUA/MUA, Fig. 5C). In addition, the number of recording sites with significant multisensory effects is comparable across regions, with slightly more multisensory neurons in the belt (core: 5% additive and 6% nonadditive multisensory; belt: 4% additive and 9% nonadditive multisensory). And finally, the fraction of sites showing enhancement or suppression for the different stimulus categories is also similar in core and belt (χ² = 6.1, P = 0.1, Fig. 5D). The absence of gross differences between core and belt demonstrates that audiovisual interactions prevail throughout the posterior auditory cortex and are not restricted to secondary auditory cortices.

### Influence of Stimulus Timing

The above experiments used naturalistic audiovisual scenes for which the auditory and visual stimuli are presented in synchrony. In general, the temporal changes in these movies are rather slow (prominent changers occur only over 2 or more frames, i.e., 80 msec and more) posing the question how sensitive the observed audiovisual interactions are with respect to the relative timing of both stimuli. To address this issue, we employed a paradigm using short and well-timed stimuli (auditory noise bursts and visual flashes, each of 20 msec duration) with varying SOA.

Figure 6A displays the average SUA/MUA firing rates (n = 97 units) for auditory stimuli with different onset asynchrony to the visual flash. A careful inspection of the peak responses reveals a lower response in conditions where the visual stimulus just precedes the auditory stimulus (SOA = 20 and 40 msec). When the visual stimulus precedes the auditory stimulus by a longer interval (160 msec), however, the peak response is comparable to auditory only stimulation (SOA = 320 msec). This observation is confirmed by an analysis of the average response amplitude (Fig. 6C), which reveals a clear decrease of the response for SOAs between 20 and 80 msec. To statistically confirm this, we used the SOA = 320 msec condition as auditory-only baseline, as in this condition the visual stimulus was presented well after the time window used to quantify the auditory response. This normalized response (Fig. 6D) reveals clear and significant response suppression of the firing rates when the visual stimulus preceded the auditory by 20–80 msec (individual t-tests for each SOA, P < 0.05).

We verified that the sensitivity of the audiovisual interaction to stimulus timing also occurs during stimulation with naturalistic stimuli (see Fig. 7 and Materials and Methods). Units that showed response enhancement for the synchronous condition (n = 32, P < 0.001 for the enhancement effect) revealed a significant reduction of the enhancement in the temporally offset condition (paired t-test, P < 0.01, synchronous vs. offset condition, Fig. 7) and units with response suppression in the synchronous condition (n = 25, P < 0.01 for the suppression) revealed a significant alleviation of the suppression in the temporally offset condition (P < 0.05). This demonstrates that audiovisual interactions under more naturalistic conditions are similarly dependent on the relative temporal alignment of the stimuli as the interactions studied with the simplistic stimuli above.

### Oscillatory Phase and Cross-modal Interactions

A previous study suggested that nonauditory information modulates auditory responses by affecting the phase of ongoing slow-frequency oscillatory activity (Lakatos et al. 2007). In detail, Lakatos and colleagues found that a somatosensory stimulus by itself causes little change to the amplitude of LFP but systematically aligns the phase of individual trials with a specific phase angle known to elicit maximal stimulus driven responses. In this way, the systematic phase alignment could boost the response to the cross-modal stimulus. We tested whether our audiovisual interactions are compatible with such a mechanism.

First, we verified that the amplitude of auditory spiking responses depends on the phase of the slow-wave activity at stimulus onset. On a subset of recording sites (n = 47), we...
Figure 6. Impact of stimulus timing—simplistic stimuli. (A) Average firing rates for different SOA values ($n = 97$ units). The auditory stimulus was always presented at zero (small gray box at bottom), and arrows indicate the time point of the visual stimulus. The vertical gray box indicates the 150-msec response window for which the amplitudes were used in panels (C), (G). The paradigm. For positive SOA values, the visual stimulus was occurring before, for negative SOA values after, the auditory stimulus. (C) Average responses for the different SOA values (mean and standard error of the mean [SEM], baseline corrected). (D) Response difference for each condition to the auditory only baseline, defined by the SOA = -320 ms condition (mean and SEM). Stars indicate significant differences from zero ($t$-test, *$P < 0.05$, **$P < 0.01$).

Figure 7. Impact of stimulus timing—naturalistic stimuli. In a separate paradigm, we presented audiovisual scenes either as the synchronous audiovisual pair (SOA = 0) or with the visual stimulus starting (SOA = 160 msec before the auditory stimulus. The error bars indicate the mean (and standard error of the mean) responses for each condition, for a set of units showing response enhancement ($n = 97$ units). $p$-values refer to paired $t$-tests.

repeated 1 auditory stimulus (that was chosen to drive the respective neuron) at least 30 times and systematically compared the elicited SUA/MUA response with the phase angle of each frequency band. Across sites, only the theta and alpha bands (5–10 Hz) showed a systematic effect: the average normalized response for the optimal and the null phase was significantly different (2-sample $t$-test, 5 Hz: $P < 10^{-3}$; 10 Hz: $P < 10^{-3}$, see Fig. 8A). These findings are in good agreement with previous studies and show that the excitability of auditory cortex is modulated by the phase of low-frequency oscillations (Lakatos, Shah, et al. 2005).

In a 2nd step, we determined the phase consistency across trials for the individual conditions (Fig. 8B). At frequencies below 40 Hz, all 3 conditions caused an increase in phase concentration across trials, with the effect being stronger for auditory and audiovisual conditions. Hence, the increase in power of the oscillatory activity (cf., Fig. 2A) is accompanied by an alignment of the oscillatory phase across trials, as is frequently observed for evoked responses (Makeig et al. 2002). To see whether differences in the degree of phase resetting could be responsible for cross-modal interactions, we compared the phase consistency between auditory and audiovisual conditions for those recording sites that showed cross-modal enhancement and suppression in the spiking activity (Fig. 8C). The only significant difference between enhanced and suppressed sites was a stronger phase consistency at 10 Hz in the audiovisual condition for the enhanced sites (2-sample $t$-test, $P < 10^{-5}$). This result was confirmed by a significant correlation of the alpha phase consistency with the SUA/MUA enhancement index across all sites (Spearman rank correlation, $r = 0.25$, $P < 0.001$, correlations for other frequency bands were not significant).Inspecting the 10-Hz phase across conditions (Fig. 8C inset) confirms that sites with enhanced response exhibit significantly higher phase consistency in the audiovisual compared with the auditory condition (paired $t$-test, $P < 10^{-4}$ for sites with enhancement, $P = 0.22$ for sites with suppression). Hence we can conclude that the degree of phase consistency correlates with the type of cross-modal interaction.

As a 3rd step, we quantified whether the visual stimulus resets the alpha phase to the optimal excitatory phase angle shown in Figure 8A. To this end, we compared the fraction of trials in which the SUA/MUA responses occurred at the optimal or null phase across conditions: Figure 7D displays the difference in the relative frequency of both phases between auditory and audiovisual conditions separately for sites with enhancement and suppression. Clearly, sites with enhancement show an increased fraction of trials with optimal phase and a reduced fraction of trials with null phase in the audiovisual condition (paired $t$-test comparing both phases, $P < 0.05$). In addition, the excess number of trials with the optimal phase in the audiovisual condition correlated significantly with the MUA/SUA enhancement index (Spearman rank correlation, $r = 0.23$, $P < 0.05$). Sites with response suppression, in contrast,
showed no significant difference between conditions and optimal and null phases. As a result, we can say that our data conform to a model where the visual stimulus enhances auditory SUA/MUA responses by systematically setting the phase of ongoing slow-wave oscillations to an optimal phase angle. In the same framework, response suppression results from a reduced phase consistency across trials.

Discussion

To summarize, our results clearly demonstrate that visual stimuli modulate sound-induced activity in primate auditory cortex, both at the level of field potentials and single-unit responses. These cross-modal interactions occur both in primary and secondary auditory cortices, are independent of the kind of stimulus, but depend on the relative audiovisual timing. In addition, the nature of the interaction (enhancement vs. suppression) is determined by the efficacy of the stimulus in driving responses. And finally, our analysis suggests that the degree of phase resetting of slow-wave oscillatory activity might contribute to the mechanisms behind this cross-modal interaction.

Nonauditory Influences in Auditory Cortex

Cross-modal influences on auditory cortex have been reported by a host of imaging studies, which provided strong evidence that auditory activity can be modulated by visual and somatosensory stimulation (Calvert et al. 1997; Calvert 2001; Bernstein et al. 2002; Foxe et al. 2002; Calvert and Campbell 2003; Pekkola, Ojanen, Autti, Jaaskelainen, Mottonen, Tarkiani, et al. 2005; Martuzzi et al. 2006; Schurmann et al. 2006). For example, auditory evoked potentials were found to modulate with somatosensory stimulation shortly after stimulus onset (Foxe et al. 2000; Murray et al. 2005), and measurements of the functional magnetic resonance imaging (fMRI)—blood oxygen level–dependent response revealed activity in auditory cortex as a consequence of just visual stimulation (Calvert et al. 1997; Bernstein et al. 2002). Often these cross-modal interactions were localized to auditory cortex in general, but some studies localized these effects particularly to the caudal portion of auditory cortex (Calvert et al. 1997; van Atteveldt et al. 2004; Kayser et al. 2005; Lehmann et al. 2006; Kayser, Petkov, et al. 2007), the region in which we directly assessed neuronal responses in the present study.

The cross-modal enhancement reported by functional imaging studies could naively be interpreted as reflecting an increase in neuronal firing rates. Until now, however, there has been little electrophysiological evidence to foster this hypothesis. Several studies found auditory population responses (LFP, current source densities, or MUA) to be enhanced by visual...
Audiovisual Interactions in Auditory Cortex

A functional Role for Cross-modal Inputs to Auditory Cortex

Both from an evolutionary and an engineering point of view, one could propose that cross-modal influences on early auditory processing should aid auditory scene analysis and perception and, as a result, should make us hear better (Lakatos et al. 2007). A classical example is the well-known cocktail party where visual information helps to segregate the voice of a speaker from background noise (Sumby and Polack 1954). Such audiovisual enhancement relies on the relative timing of auditory and visual information, which fits with our results. The optimal time interval for audiovisual interaction was in the range of 20–80 msec, and taking differential processing times for auditory and visual stimuli into account (Schroeder and Foxe 2002), this latency difference corresponds to potential sources within a range of less than about 15 m, that is, events with immediate consequences. In addition, very similar temporal offsets are necessary for human subjects to perceive auditory and visual events as synchronous. Perceptual studies showed that—depending on the exact stimulus and task—visual stimuli need to occur between 20 and 90 msec before the auditory event in order for subjects to report them as synchronous (Zampini et al. 2003; Arrighi et al. 2006; Vatakis and Spence 2006). All in all, this shows that the same kind of event that results in behavioral benefits of integration also elicits cross-modal interactions in early auditory areas.

This notion is also supported by several behavioral studies proposing that visual and tactile stimuli improve detection and perception of auditory stimuli in a manner that is consistent with integration at the sensory-perceptual level. As an example, Gillmeister and Eimer recently reported that irrelevant tactile stimuli enhance the detection of faint tones and increase perceived auditory intensity in a manner that respects the principles of temporal coincidence and inverse effectiveness (Gillmeister and Eimer 2007) and see Odgaard et al. (2004) and Schurmann et al. (2004) for related results. It seems reasonable to speculate that the early cross-modal influences in auditory cortex observed here and in previous studies reflect the cortical basis for such cross-modal benefits at the behavioral level.

In addition to the temporal alignment of stimuli, also their spatial position might be of importance. Evidence from the ferret auditory cortex suggests that auditory neurons can be sensitive to restricted areas of the visual space (Bizley et al. 2006), and the results of Lakatos et al. (2007) clearly demonstrate a spatial selectivity in the auditory–somatosensory interaction. In the primate, the caudal auditory cortex is supposedly part of a spatial where stream and many neurons are sensitive to the spatial location of a sound (Recanzone 2000; Schroeder and Foxe 2002; Ghazanfar et al. 2005) and somatosensory stimuli (Lakatos et al. 2007). Yet, comparable effects on individual neurons were only observed in highly trained monkeys presented with a visual cue during an auditory task (Brosch et al. 2005) and in the auditory cortex of anesthetized ferrets (Bizley et al. 2006). In the ferret, visual stimuli activated or modulated many neurons in the primary auditory field, but contradicting the expectation from the imaging studies, most neurons showed signs of response suppression.

Our results reconcile and extend these previous findings in several ways. Prominently, we show that the observations of response enhancement in local population responses and of response suppression in single-unit firing rates are not mutually exclusive, but both can indeed occur at the same time and at the same cortical site. In addition, our data reveal that the audiovisual interaction does not depend on the kind of visual stimulus but depends on the efficacy of the auditory stimulus in eliciting responses. In fact, a stimulus’ efficacy predicts how it will be modulated by an appropriately timed visual input. As a result, our data show that cross-modal interactions can only be interpreted when considering the efficacy of the individual stimuli.

This efficacy of course depends on the measure of neuronal activity under consideration and might well differ between selected single units and population indices such as field potentials or functional imaging methods. For example, a liberally chosen (naturalistic) stimulus will be optimal only for a few neurons but will suboptimally activate many others. When combined with the matching visual stimulus, this will result in response suppression for the few well-driven sites and response enhancement for many other sites. Hence, when quantifying these responses using a population index (LFP or fMRI), one might be biased toward the response enhancement but miss the response suppression. When recording single units, on the other hand, one might be biased toward the well responding neurons and find more prominent suppression. This differential bias could explain why on average across all stimuli we found response enhancement for the LFP and suppression for the firing rates. Importantly, these considerations highlight the difficulties in comparing the results of studies on cross-modal interactions that use a wide range of stimuli and methods.

Together with a related study on the impact of somatosensory stimulation on auditory responses by Lakatos et al. (2007) (and see Bizley et al. 2006 for related results), the present results confirm that cross-modal influences in early auditory cortices share the attributes that are frequently used to detect and characterize sensory integration (Stein and Meredith 1993; Calvert et al. 2001; Avillac et al. 2007): The interactions are sensitive to the relative timing of auditory and nonauditory stimuli, the qualitative nature of the interaction (enhancement vs. suppression) depends on the efficacy of the unisensory stimuli in driving the neurons, and—as shown in the Lakatos study—the interactions depend on the spatial alignment of the stimuli.

Noteworthy, the results of Lakatos and colleagues also show that the qualitative nature of the interaction can also depend on the relative timing, as they found alternating patterns of enhancement and suppression for different relative SOA of the auditory and somatosensory stimuli. This last result contrasts somewhat with the present findings, as we did not see alternating modes of enhancement and suppression over a large range of SOA. One reason might be that our sampling of SOA was not dense enough. Another reason might be that audiovisual and auditory–somatosensory interactions follow slightly different patterns or might rely on partly different mechanisms.

All in all, these findings demonstrate that the information communicated from early auditory cortices to higher processing stages not only reflects the acoustical environment but also depends on the multisensory context of a sound. Although the main feed-forward drive along the auditory processing stream is for sure is acoustically dominated, a thorough understanding of how a sound-wave impinging on the ear finally results in an auditory percept can only be obtained by considering the brain within its multisensory context.
Rauschecker and Tian 2000; Recanzone et al. 2000; Zatorre et al. 2002). In this respect, sensory interactions in the caudal auditory cortex could help to bring auditory and visual information about spatial attributes into register—a hypothesis that yet remains to be tested explicitly.

In contrast to the relevance of spatial-temporal attributes of a stimulus, our data suggest that the particular type of audiovisual stimulus does not matter; in fact, we found comparable audiovisual interactions during stimulation with naturalistic scenes and with simple noise bursts and flash stimuli. As a result, we suggest that the audiovisual interactions are not optimized for the processing of a specific kind of stimulus, for example communication signals, but are rather part of a general purpose mechanism providing the auditory processing with information about the occurrence, and possible intensity, of a visual stimulus. This reasoning is supported by several previous studies that also found cross-modal interactions using very simplistic visual (Brosch et al. 2005; Bizley et al. 2006) or somatosensory stimuli (Fu et al. 2003; Łakatos et al. 2007).

**Attention, Eye Movements, and Cross-modal Effects**

One could argue that the observed cross-modal interactions reflect enhanced sensory processing due to focused attention. Especially, both sensory integration and attention serve to enhance perception by increasing the sensitivity to particular sensory events, and functional imaging studies have demonstrated that focused attention to 1 modality can enhance processing and activity of colocalized stimuli in another modality or suppress activity in the unattended system (Driver and Spence 1998; Macaluso et al. 2000; Weissman et al. 2004; Johnson and Zatorre 2006). In addition, attentional modulation can interact with and enhance cross-modal effects, highlighting that both mechanisms can simultaneously affect neuronal processing (Talsma et al. 2006). As a result, it can be difficult to disentangle response modulations due to cross-modal and attentional effects.

Yet, there are several arguments convincing us that the observed visual modulation of auditory cortex is not due to attentional modulation alone. First, cross-modal interactions occur in anesthetized animals, as demonstrated by recordings of single neurons in anesthetized ferrets (Bizley et al. 2006) and functional imaging of anesthetized monkeys (Kayser et al. 2005; Kayser, Petkov, et al. 2007). Anesthesia reduces the activity in association areas more than in sensory regions and prevents cognitive and attentive mechanisms (Disbrow et al. 2000; Heinke and Schwarzbauer 2001; Kayser et al. 2005), ruling our attentive feedback mechanisms as major source for the cross-modal interactions. Second, similar cross-modal interactions were found during stimulation with long and naturalistic stimuli and with very short and randomly paired noises and flashes. Although attentive mechanism might affect the presentation of meaningful stimuli, such as conspecific vocalizations, it seems unlikely that attention would affect short and artificial stimuli. Especially the short SOA of 20 ms, which was sufficient to induce audiovisual response suppression, is below the typical SOA values necessary for audiovisual links of exogenous attention (Spence and Driver 1997; Spence et al. 1998; Mazza et al. 2007). And 3rd, the cross-modal effects occurred with short latencies, around 70 msec, and this time is before the attentional modulation of response components that are typically observed (Hillyard and Anllo-Vento 1998; Reynolds et al. 1999; Fritz et al. 2007). Altogether, this suggests that attentional mechanisms are not the main source contributing to the observed visual modulation of auditory activity.

Related to attentional mechanisms are movements of the eyes. Two previous studies reported an effect of eye position on the auditory responses of neurons in the core and belt (Werner-Reiss et al. 2003; Fu et al. 2004). Although this effect occurred with a similar latency as the visual modulation observed in the present study, we are confident that eye movement-related parameters are not the source of visual modulation. First, again findings from experiments in anesthetized animals rule out contributions from eye movements (Kayser et al. 2005; Bizley et al. 2006; Kayser, Petkov, et al. 2007). Second, for the present data, we systematically quantified the relation of eye movements and cross-modal effects (Supplementary Fig. 2). Neither the number of fast saccadic eye movements nor the SD of eye position during a trial differed significantly between conditions. In addition, neither of these eye movement characteristics correlated with the enhancement index, and visual modulation occurs also for recordings with tight fixation. If one could see any trend in our data, it would rather suggest that residual eye movements abolish the audiovisual interaction, as the animal that was fixating in the larger window showed on average smaller differences between auditory and audiovisual responses. In addition, alternations of auditory responses by eye position occur over larger spatial scales, usually 10 degrees or more (Werner-Reiss et al. 2003; Fu et al. 2004), which are larger than the fixation windows used in the present study. And 3rd, our finding that audiovisual interactions depend on the relative temporal alignment of the stimuli make it unlikely that eye movements play a prominent role. On the contrary, a previous study on the superior colliculus has shown that fixations can reduce sensory-related responses and shift the audiovisual integration mode toward more subadditive effects (Bell et al. 2003). Although the responses in the present paradigm were mostly subadditive, we did not see any correlation between the degree of subadditivity and eye movement characteristics. Hence we can only speculate whether a specific pattern of eye movements might bias audiovisual interactions in either direction, but for the data presented here, it seems unlikely that residual eye movements within the fixation window would cause or enhance the cross-modal interactions.

**The Mechanisms behind Cross-modal Interactions**

An important question regarding the nonauditory modulation of auditory activity pertains to origin of the anatomical projections mediating these effects and the underlying mechanisms. Although we can only speculate on which structures provide the cross-modal input to auditory cortex (see discussions in Kayser et al. 2005; Kayser and Logothetis 2007; Kayser, Petkov, et al. 2007), the present results reveal aspects of the mechanisms by which visual stimuli modulate auditory responses. Following the results of previous work by the Schroeder group (Lakatos, Shah, et al. 2005; Lakatos et al. 2007), we tested the hypothesis that nonauditory input alters the auditory activity by manipulating the oscillatory phase of ongoing LFP. In agreement with their report, we found that the strength of a plain auditory evoked response systematically depends on the phase of the slow-wave oscillatory activity at the time point of stimulus presentation. This phase-dependent response modulation was strongest for the frequencies...
between 5 and 10 Hz, which is in good agreement with the theta phase reported by Lakatos et al. Although these authors found similar effects pertaining to the delta band (1–2 Hz), we could not investigate such low frequencies in the present study due to the 4-Hz high-pass filter used during recording.

Continuing along their lines, we verified that the cross-modal enhancement correlates with the degree of phase consistency across trials, as well as with the prevalence of trials with the phase of optimal excitability. These findings are fully compatible with the idea that the phase of low-frequency oscillations is a determining factor that controls the response strength for auditory stimuli and hence provides a mechanism through which input from other modalities could alter auditory responses (Ghazanfar and Chandrasekaran 2007). Along these lines, one can speculate about further relations between the observed cross-modal influences and neuronal oscillations. For example, enhancement and suppression were found to persist over different time scales; whereas enhancement lasted less than 300 msec on average, suppression prevailed for nearly half a second. Noteworthy these numbers match well with the 2 prominent modes of oscillations observed in auditory cortex, the delta (1–2 Hz) and theta (5–10 Hz) bands. If cross-modal interactions are specifically mediated by oscillatory activity, then it might be conceivable that interactions persist for or occur at the time scales of the respective oscillatory frequencies—as for example observed in the auditory-somatosensory interactions (Lakatos et al. 2007). Despite this exciting perspective, one should keep in mind that a causal relation between cross-modal interactions and oscillatory activity has yet to be confirmed and will provide a challenging topic for future research.

Supplementary Material
Supplementary figures can be found at: http://www.cercor.oxfordjournals.org/.

Funding
The Max Planck Society; the German Research Foundation (KA 2661/1); the Alexander von Humboldt Foundation.

Notes
Conflict of Interest None declared.

Address correspondence to email: christoph.kayser@tuebingen.mpg.de.

References


